Л.Т. Кузин

ОСНОВЫ КИБЕРНЕТИКИ

В двух томах
Том 1

Математические основы кибернетики

2-е издание, переработанное
и дополненное

Рекомендовано Комитетом по высшей школе Министерства науки, высшей школы и технической политики Российской Федерации в качестве учебного пособия для студентов вузов, обучающихся по специальностям: «Прикладная математика» и «Автоматизированные системы обработки информации и управления»

МОСКВА · ЭНЕРГОАТОМИЗДАТ 1994

ББК 32.81
К89
УДК 51:007(075.8)

Рецензенты: кафедра математической кибернетики МАИ; кафедра проектирования и организации систем МФТИ

Кузин Л. Т.
К89 Основы кибернетики. В 2 т. Т. 1. Математические основы кибернетики: Учеб. пособие для вузов.— 2-е изд., перераб. и доп.— М.: Энергоатомиздат, 1994.— 576 с.: ил.
ISBN 5-283-02533-0

Представлены специальные разделы дискретной математики, необходимые для подготовки инженеров, специализирующихся по информатике и математическому обеспечению кибернетических систем. Изложены основы современных неклассических логических исчислений, теории категорий, реляционной и тензорной алгебры и исчисления. Материал книги полностью переработан и доведен до современного состояния в данной области знаний.

Для студентов вузов, обучающихся по специальностям АСУ и «Прикладная математика».

$К\frac{1402010000\text{-}059}{051(01)\text{-}94}$ 1-91 ББК 32.81

Учебное издание

Кузин Лев Тимофеевич

ОСНОВЫ КИБЕРНЕТИКИ

Т. 1. Математические основы кибернетики

Редактор **А. Д. Начинкин**
Редактор издательства **З. И. Михеева**
Художественный редактор **Т. А. Дворецкова**
Технические редакторы **Г. В. Преображенская, Г. С. Соловьева**
Корректор **Л. А. Гладкова**

ИБ № 2499

Сдано в набор 03.12.91. Подписано в печать 12.02.93. Формат 60×90/16. Бумага офсетная № 2. Гарнитура литературная. Печать офсетная. Усл. печ. л. 36,28. Усл. кр.-отт. 36,28. Уч.-изд. л. 35,87.
Заказ 6891

Энергоатомиздат, 113114, Москва, М-114, Шлюзовая наб., 10

Типография ВИ

ISBN 5-283-02533-0(Т.1)
ISBN 5-283-01532-7

ПРЕДИСЛОВИЕ

Интенсивное внедрение электронных цифровых вычислительных машин во все сферы человеческой деятельности, разработка и внедрение автоматизированных систем управления и интеллектуальных систем очень остро поставили проблему подготовки и переподготовки кадров, специалистов по кибернетике, системному программированию и интеллектуальным системам. Особенность научно-технической революции, заключающаяся в одновременности решения научных и технических проблем, теоретических и практических вопросов, обусловила появление специалиста нового типа — инженера-математика, или математика — специалиста по прикладной математике, сочетающего глубокие теоретические знания с инженерными навыками и эффективно использующего достижения математики в конкретной практической деятельности. Специалист такого профиля, который еще называется математиком-прикладником, может стимулировать или сам разрабатывать прикладные разделы математики, предназначенные для решения инженерных задач, которые возникают в его практической деятельности, и внедрять новые методы исследования. Рациональное сочетание математических и инженерных знаний позволит инженеру-математику разрабатывать и внедрять новые технические системы даже при отсутствии их теоретических основ. Но необходимо отметить, что подготовка таких специалистов осложняется тем, что учебные руководства в ряде случаев не успевают за развитием новых разделов кибернетики. В качестве примера можно указать раздел кибернетики, связанный с интеллектуальными системами и индустрией искусственного интеллекта.

Данное пособие сформировалось в результате чтения курсов лекций в Московском инженерно-физическом институте в течение 1963—1990 гг. студентам старших курсов, а также работникам промышленных предприятий.

Содержание пособия удовлетворяет квалификации инженер-математик, или математик-прикладник, т. е. в нем изложены основные математические методы кибернетики без ненужного для инженера углубления в математические тонкости и без снижения до примитивного уровня изложения. Найти во всех разделах кибернетики такую оптимальную грань между «чистой» и «инженерной» математикой оказалось нелегкой задачей, и автор признает, что ему не везде это удалось.

При подборе материала автор в основном руководствовался центральной идеей обеспечения читателя необходимыми сведениями в области электронной обработки информации, в том числе интеллектуальных информации или знаний. Данная книга представляет собой первый том общего руководства по «Основам кибернетики», который по сравнению с первым изданием, по существу, написал заново. В него включены разделы дискретной математики, обычно не включаемые в стандартный курс, которые названы специальными разделами дискретной математики.

Подробное изложение принципов и концепций, положенных в основу для формирования и подбора разделов математики для этого курса, сделано в гл. 1 данного тома, здесь же остановимся только на некоторых вопросах становления курса, который иногда называется новой математикой.

Разделы, составляющие 50 % т. 1 второго издания, как бы прошли испытание временем. Заново написаны гл. 3 и 4, посвященные общей алгебре и теории категорий, гл. 12, посвященная тензорам, и гл. 6, посвященная неклассической логике и теории доказательств, гл. 7 «Теория типов» и вся геометрическая ч. 5, состоящая из гл. 13 «Элементы алгебраической топологии» и гл. 14 «Алгебраическая и дифференциальная геометрия».

Данное учебное пособие предназначено для студентов специальности «Прикладная математика», оно также может быть использовано при подготовке инженеров по специальностям «Автоматизированные системы обработки информации и управления» и «Программное обеспечение ВТ и АСУ». Большую помощь оно может оказать аспирантам и инженерам, желающим повысить свое образование в области кибернетики. Автор предполагает, что читатель знаком с курсом высшей математики и в том числе дискретной математики в объеме, предусмотренной программой, утвержденной для вузов. Большую помощь в написании данного пособия оказали сотрудники кафедры кибернетики Московского ордена Трудового Красного Знамени инженерно-физического института, ученики и коллеги. Автор выражает им благодарность, особенно В. Э. Вольфенгагену, а также Н. Н. Гавриловой, С. И. Днепровскому, М. Н. Давыдовой, Д. Ю. Абрамову, Л. Ю. Исмаиловой, А. И. Черемисину, А. Г. Зенину, А. А. Нефедову, В. Г. Толстову, В. В. Тимофеевой, А. Е. Петрову, а также всем, кто способствовал выходу этой книги, и надеемся, что она может оказать помощь студентам, аспирантам и инженерам, изучающим вечно молодую науку об управлении — кибернетику.

Автор

Часть 1

ВВОДНЫЕ РАЗДЕЛЫ. ВВЕДЕНИЕ В КИБЕРНЕТИКУ

Начало второй промышленной революции в значительной степени связано с появлением вычислительных машин или компьютеров. Следует заметить, что чаще используется термин «научно-техническая революция», который подчеркивает одновременность коренных изменений в технике и науке, чего не наблюдалось в первой технической революции, начало которой обычно связывают с изобретением паровой машины Уатта (1884 г.) [1—22].

Первая техническая революция была направлена на механизацию деятельности человека, на замену мускульной силы машинами. Можно утверждать, что первая техническая революция касалась энергетической стороны явлений, тогда как вторая затрагивает в основном информационные аспекты процессов. Одной из задач второй технической революции, начало которой можно отнести к 1945 г., является автоматизация производственных и других процессов — замена человека вычислительными и кабернетическими машинами в процессе управления. В настоящее время поставлена цель создания машин, способных частично или полностью заменить человека в сфере его умственной, интеллектуальной деятельности, которые были бы своего рода усилителями человеческого, естественного интеллекта, так же как механические, электрические, гидравлические машины являются усилителями мускульной силы человека. Таким образом, прослеживается исторически предопределенный эволюционный процесс механизации, автоматизации и интеллектуализации человеческой деятельности в любой сфере. Причем процесс развития идет не всегда последовательно, так как в одной области не закончен процесс механизации, а в другой осуществляется уже процесс интеллектуализации. По данным технического комитета ООН за последние 15 лет доля неавтоматизированного производства (ручного труда) сократилась с 75 до 15 %, при этом доля полностью автоматического производства возросла с 12 до 32 %, а доля автоматизированного — с 12 до 60 % [22]. Наряду с электронно-вычислительными машинами (ЭВМ), предназначенными для обработки данных, появились нового типа компьютеры для обработки знаний, которые стали называться интеллектуальными компьютерами. Вместе с персональными компьютерами (ПК), вычислительными системами 5-го поколения (ВС-5), экспертными системами (ЭС) и новой информационной технологией (НИТ) интеллектуальные компьютеры должны обеспечить построение индустрии знаний. Таким образом, естественный процесс развития привел к проблеме построения систем искусственного интеллекта как основного, центрального раздела кибернетики. Характерной особенностью нового этапа использования вычислительной техники является вовлечение в процесс проектирования программирования пользователя-некибернетика благодаря знаниям, закладываемым в ЭВМ, т. е. речь идет о всеобщей компьютеризации трудящегося населения благодаря созданию дружелюбных к пользователю, «народных» и дешевых компьютеров. Другой очень важной особенностью ЭВМ широкого использования является ее дешевизна, точнее, свойство, заключающееся в уменьшении стоимости ЭВМ и составляющих ее частей в 2 раза каждый год. Это усилило тенденцию увеличения использования локальных персональных компьютеров за счет ослабления доли использования ЭВМ, включенных в сеть передачи данных [22—24].

В 1983 г. в США было продано около 10 млн. персональных компьютеров, из них примерно 6 млн.— в учреждения, а остальные — в личное пользование. В нашей стране и США в 1980 г. больше половины работающего населения (примерно 50 млн.) было занято в сфере управленческого труда, т. е. не было связано непосредственно с производством и обрабатывало бумажную информацию. В СССР — это служащие, которых за рубежом называют «белыми воротничками». В США в 1980 г. эта армия служащих получала в год зарплату около 790 млрд. долларов [22]. Поэтому механизация, автоматизация и интеллектуализация процессов обработки бумажной информации — вечная проблема, которая требует своего эффективного решения в настоящее время, а также в ближайшем и далеком будущем. С помощью интеллектуальных систем, которые помогают пользователю принимать решение, делается еще одна попытка создать безбумажные технологии управления и уменьшить объем бумажной информации. В связи с этим большое значение для создания дружелюбных компьютеров и кибернетических систем имеют системы общения, в том числе на естественном языке.

Автоматизированные методы обработки информации, особенно обладающие свойством интеллектуальности и разработанные в кибернетике, имеют отношение к специалистам любой профессии, будь то физик, химик или инженер-технолог, так как все они имеют дело с переработкой деловой, технической, научной и любой другой информации. Эта особенность кибернетики как науки о переработке информации в целях управления придает ей универсальный характер и возводит ее в категорию фундаментальных наук.

В.1. ИСТОКИ КИБЕРНЕТИКИ

Для того чтобы управлять, необходимы технические средства управления: измерительные датчики для получения информации, исполнительные устройства, усилители, вычислительные машины. Но прежде всего необходима наука об управлении, которая предоставила бы человеку эффективные методы управления и проектирования систем управления.

Такой наукой является кибернетика, название которой происходит от соединения двух греческих слов: «кибер» (в переводе «над») и «наутис» («моряк»), т. е. «кибернаутис» — старший над моряками, главный моряк, кормчий. Греческий философ Платон впервые использовал термин «кибернетика» в смысле искусства управления обществом. В XVIII в. французский ученый Ампер, составляя классификацию наук, также назвал кибернетикой науку об управлении обществом. Но в то время эта наука не имела математической и технической основы, и этот термин был на 200 лет забыт. В 1948 г. Винер в своей книге «Кибернетика или управление и связь в животном и машине» возродил этот термин в более широком, современном смысле и наметил, по существу, программу развития кибернетики. В этой книге о кибернетике, как науке об управлении в живом и неживом мире утверждается, что она основывается на математике и вычислительных машинах [1—5]. После этого было несколько уточнений понятия кибернетики. В частности, кибернетика определялась как фундаментальная наука о переработке информации в живой и неживой природе в целях управления, т. е. на первое место в этом определении была поставлена переработка информации, что имело свои положительные и отрицательные последствия.

С философской методологической точки зрения место кибернетики среди других фундаментальных наук, таких, как физика и химия, можно определить схемой, представленной на рис. В.1. Если физика рассматривает свойства материи с точки зрения движения и преобразования энергии, а химия — с точки зрения преобразования вещества с одной молекулярной структурой в другую, то кибернетика исследует третью группу свойств материи — информацию.

Нам представляется, что кибернетика много взяла от разных наук. Исторически она создавалась как бы по двум направлениям. С одной стороны, многие науки — физика, химия, экономика, радиотехника, медицина, биология и др. — на определенном этапе своего развития в стремлении активно воздействовать на изучаемые явления и управлять ими стали созда-

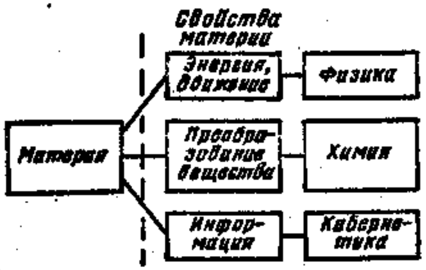


Рис. В.1. Взаимосвязь различных наук

вать математические и кибернетические модели соответствующих явлений. Случилось так, что различные по природе явления, изучаемые разными науками, стали описывать одинаковыми математическими моделями. Например, движение ракеты, работа промышленного предприятия, процесс размножения клеток могут быть описаны дифференциальными уравнениями. Для управления этими различными по природе явлениями потребовались одни и те же математические методы автоматизированной обработки информации, а также технические средства, начиная от ЭВМ с дисковой и оперативной памятью и кончая устройствами ввода и вывода информации.

С другой стороны, математики, понимая все значение задач управления и переработки информации, стали разрабатывать методы, которые позволили бы наиболее полно и адекватно описывать различные процессы с информационной точки зрения, находить подходящие режимы управления, а также автоматически обрабатывать информацию с помощью ЭВМ. Благодаря кибернетике, главным образом ее разделу по системам искусственного интеллекта, получили интенсивное развитие такие новые разделы математики, как теория категорий λ-исчисление, неклассические логики, теория типов и др. Развитие вычислительных машин происходило не только в направлении повышения их быстродействия, но и в части создания аппаратных и программных средств хранения, обработки и поддержания больших емкостей информации. Благодаря деятельности кибернетиков, системных программистов, математиков, главным образом при разработке и внедрении АСУ и интеллектуальных систем, появились совершенно новые методы переработки информации для управления, новые методы ее представления и выдачи человеку. С помощью этих методов и специальных технических средств человек может обращаться с большими массивами информации, записанными на магнитных носителях (магнитных лентах, барабанах и дисках). Вместо бумажных деловых документов и книг появились документы на магнитных носителях. Появился человек (пользователь), который в процессе управления должен научиться «читать» информацию, записанную на магнитных носителях, так же свободно, как бумажные документы и книги, при этом благодаря высокому уровню автоматизации и быстродействию ЭВМ производительность (эффективность) процесса получения

научной информации должна повышаться на несколько порядков. Появились инженерные методы хранения и обработки информации в виде банков данных, баз данных (БД), систем управления базами данных (СУБД), базы знаний (БЗ), систем управления базами знаний (СУБЗ). Интеллектуальные банки данных (ИБД) как минимум состоят из БД и БЗ. Из машин-арифмометров ЭВМ превратились в компьютеры обработки данных, в компьютеры обработки знаний. Эти новые человеко-машинные системы обработки информации образовали новую теорию — теорию информации для кибернетики. Появились теоретические работы, новые разделы математики для описания БД и БЗ. При изучении процессов принятия решений человеком естественным интеллектом были установлены ряд новых свойств, главными из которых являются внутренняя противоречивость (отсутствие строгой логики), нечеткость и человеко-машинный аспект. Классическая схема философской логики Аристотеля «посылка-следствие», которая полностью формализована в математической логике, состоящей из исчисления высказываний и исчисления предикатов 1-го порядка, оказалась недостаточной и несостоятельной. Поэтому кибернетики обратились к противоречивым философским логикам — диалектическим логикам Гегеля и Канта, где следствие помимо посылки зависит еще от действующего фактора ситуации. При одной и той же посылке, но разных ситуациях могут быть разные следствия, в том числе и полностью противоположные.

В связи с этим большое внимание стали привлекать конструктивные, неполные, внутренне противоречивые, нечеткие математические построения. Развитием этих разделов математики применительно к потребностям должны заниматься специалисты по прикладной математике, которая с точки зрения классификации наук должна заниматься математическими основами кибернетики, так же как она занималась математическими основами физики, химии, квантовой механики и т. д. Однако у кибернетики, так же как у любой другой самостоятельной науки, имеется свой, специфический объект исследования — процессы управления, основанные на преобразовании информации. Поэтому и появилась концепция инженера-математика, который, имея минимальную инженерную подготовку (в основном в области технологии программирования и использования ЭВМ) и обладая повышенной математической подготовкой, может найти и развить подходящий математический аппарат. Различают кибернетику общую, изучающую принципы и методы управления, справедливые для любых систем, и специальные, в которых изучаются методы управления определенными классами систем, методы составления кибернетических моделей этих систем. Поэтому с этой точки зрения правомерно рассматривать специальные кибернетики: физическую, химическую, техническую, экономическую, военную и т. д. Кроме того, появились понятия теоретической и прикладной кибернетики. Последняя в какой-то мере объединяет специальные кибернетики. В последнее время стал популярен термин «информатика», который в определенной области пересекается с кибернетикой. В свое время у нас в стране была издана книга с таким названием [25], в которой рассматривались информационно-поисковые системы, что внесло определенную путаницу в трактовку этого термина. В современной трактовке информатика — это наука об использовании ЭВМ [26] т. е., то, что за рубежом и в нашей стране понимается под термином Computer Science.

В последнее время этот термин был расширен до задач обработки данных в ЭВМ. В этом смысле имеется перекрытие информатики с кибернетикой. Но в информатике отсутствуют концепция и методы управления, кроме того, кибернетика существует как наука независимо от ЭВМ, которые используются в ней так же, как физические приборы в физике.

В научно-технической литературе часто встречается термин «исследование операций», который в большинстве случаев совпадает с кибернетикой при нашей трактовке ее содержания. Вся разница заключается в точке зрения на одни и те же задачи управления. В кибернетике явления рассматриваются с точки зрения управления с помощью преобразования информации, а в исследовании операций — с точки зрения операции как составного понятия любого процесса управления.

Характерная особенность методов исследования операций заключается в расчленении (представлении) исходного процесса на совокупность мелких операций, что позволяет лучше понять поведение исходного процесса. Так, в случае применения сетевого планирования исходный комплексный проект представляется в виде совокупности отдельных работ, которых может насчитываться несколько тысяч. Однако кибернетика содержит в своем арсенале также метод сетевого планирования, и различие в подходе заключается в методологии изучения явлений. Кибернетика на первое место выдвигает преобразование информации, а теория исследования операций — представление исходного процесса в виде совокупности мелких операций. Иногда указывают на то, что при исследовании операций рассматриваются математические методы управления массовыми явлениями, вопросы организации, тогда как кибернетика наряду с этим исследует процессы управления отдельными объектами. Однако процесс управления отдельным объектом можно расчленить на совокупность операций и, применив методы исследования операций, свести его, по существу, к массовому явлению. Такая путаница в терминах возникла, по-видимому, из-за того, что в современном понимании исследование операций появилось раньше, чем кибернетика; в 1942—1943 гг. английские ученые и инженеры ввели этот термин для обозначения науки, занимающейся математическими методами организации сложных систем, а термин «кибернетика» появился позднее в США (1945—1946 гг.) для обозначения науки об управлении [1—7].

В результате быстрых темпов научно-технической революции появляется много новых научных направлений и теорий. Некоторые из них имеют под собой серьезную основу и право на длительное существование. Так, благодаря необходимости системного подхода при разработке АСУ и других кибернетических систем появилась теория систем, также противопоставляемая кибернетике как теория, в которой на первое место выделяются не информационные принципы, а вопросы сложности. Однако следует заметить, что кибернетика немыслима без теории систем, так как она занимается системами управления, правда, с информационной точки зрения. Но, как будет видно из дальнейшего, мера информации существенным образом зависит от размеров и сложности системы. Наконец, в литературе появился новый термин — «информатика», который объединяет методы обработки информации, не предусматривающие процесс управления или передачу информации по каналам связи (например, перевод с одного естественного языка на другой, поиск заданного типа информации и т. д.). Однако практически все приемы обработки информации, разработанные в информатике, применяются в системах управления.

В.2. ОСНОВНЫЕ ЧЕРТЫ КИБЕРНЕТИКИ

Первой особенностью кибернетики является информационный подход к процессам управления. Можно сказать, что там, где имеет место переработка информации в целях управления, присутствует кибернетика. Если при механизации имеют дело с энергией (ее усилением или преобразованием из одного вида в другой), то при автоматизации управления на первый план выдвигаются информационные свойства сигналов. Так, в системах автоматического регулирования (или следящих систем) должна быть минимальной разница между входным и выходным сигналами (сигнал ошибки), а в системах автоматического управления — разность между заданной функцией входного сигнала и выходным сигналом. Наиболее отчетливо информационный подход в кибернетике проявляется при проектировании и эксплуатации АСУ организационого типа, которая предназначена для автоматизации управленческой деятельности служащих и построения индустрии электронной обработки данных (ИЭОД), при этом создается информационное обеспечение, основными составляющими которого являются информационная модель объекта (предприятие, цех, отрасль промышленности и т. д.) и банк данных. Информационная модель предусматривает создание схемы документооборота объекта, в частности предприятия, в котором отражаются структура и динамика образования всех документов, циркулирующих на предприятии. Замоделированная на ЭВМ такая схема при наличии эффективных средств ввода-вывода информации, а также средств общения человек-машина со временем приведет к полному уничтожению деловых бумажных документов и замене их документацией на магнитных носителях.

Однако после определенного опыта проектирования и эксплуатации АСУ стало очевидным, что для эффективности их применения требуется прежде всего решать две проблемы:

создание интеллектуальной АСУ (ИАСУ), которая помогла бы пользователю принимать решения, а не просто моделировала документооборот;

создание эволюционной, постоянно развивающейся АСУ за счет совмещения системы автоматизированного проектирования (САПР) АСУ с эксплуатацией АСУ. САПР должна быть встроена в АСУ.

Но проектирование или развитие АСУ должно осуществляться пользователем (некибернетиком, непрограммистом), работником функциональных подразделений в АСУ предприятия (бухгалтером, плановиком, начальником цеха и т. д.) с помощью интеллектуальной САПР, которая содержит все недостающие, но нужные для проектирования сведения и в которой реализована так называемая новая информационная технология (НИТ). Эта технология [27—29] является безбумажной САПР, в которой не требуется техническое задание на разработку эскизного, технического и рабочего проектов (так называемых спецификаций). С помощью интеллектуального постановщика задачи пожелания заказчика, сформулированные неточным и неполным образом, возможно, на естественном языке доводятся до строгих в техническом смысле формулировок. Это техническое задание, так же как другие документальные сведения (эскизный и другие проекты), хранятся в памяти ЭВМ в так называемом банке разработки (БР). Весь ход разработки фиксируется в БР, любой участник разработки может запросить и получить информацию в любом виде — на экране дисплея или в виде бумажной документации на выходе печатающего устройства ЭВМ. Помимо БР в технически высокоразвитой системе электронной обработки данных (СЭОД) или любой другой кибернетически развитой системе, использующей ЭВМ, должен быть свой банк данных (а также СУБД), в котором содержалась бы вся необходимая для управления информация. Концепция БД появилась в значительной степени независимым способом, когда от машин-арифмометров (первое и второе поколения ВС) был осуществлен переход к ВС третьего и четвертого поколений (ВС-3, ВС-4), которые предназначены для обработки данных и в которых сильно развиты средства хранения информации во внутренней и внешней памяти, а также системы ввода и вывода информации (рис. В.2). Что касается программного обеспечения, то были разработаны надстройки над операционными системами в виде СУБД. Первая СУБД или БД, которая была в 1972 г. изготовлена фирмой IBM, имела следующую идеологическую основу. Так как ранее при решении задач на ВС-1, ВС-2 программа содержала операторы и данные, причем в программе указывалось, какие данные, в каком виде и где их найти, то при составлении новых программ раздел данных надо было составлять заново. С этим еще можно было мириться, когда ЭВМ решала вычислительные, арифметические задачи. Когда ЭВМ стала применяться для электронной обработки данных, оказалось, что без специальных средств автоматизации составления раздела данных принципиально невозможно осуществлять электронную обработку данных (ЭОД), так как скорость нарастания ошибок значительно опережала сложность программ с данными. Были сделаны попытки создания языков программирования, ориентированных на обработку данных (COBOL), которые повысили эффективность составления программ для электронной обработки данных. Но творческая инициатива инженеров и ученых осуществляла движение дальше — были предложены СУБД. Эти системы представляют программисту универсальные, для широкого класса задач инструментальные средства и методику составления программ СЭОД. В СУБД наблюдается даже ситуация, когда данные управляют программой, останавливают ее и вызывают ее продолжение. В настоящее время в нашей стране существуют типовые СУБД различных классов — ОКА, ИНЭС, СПЕКТР и т. д. — на современном уровне проектирования СЭОД, начиная с выбора СУБД. Появились даже системы управления проектированием, основанные на данных (COBOL), где дан-

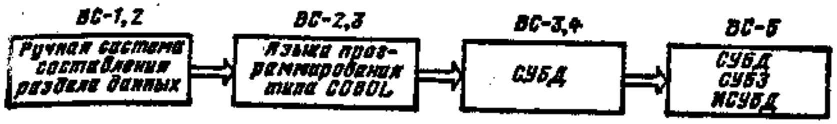


Рис. В.2. Эволюция систем электронной обработки данных и вычислительных систем:

ВС-1, ВС-2 — для вычислений; *ВС-3, ВС-4* — для обработки данных; *ВС-5* — для обработки знаний ные вызывают процедуры, которые сами могут иметь вид данных, и т. д. Кроме того, при объеме данных в БД более 20 Мбайт она становится неуправляемой «вручную» и разрушается. Однако практика эксплуатации, поддержания БД большого объема (более 50 Мбайт) показала, что для их нормального функционирования необходима БЗ предметной области (ПО), в которой содержались бы закономерности, семантика, смысловое содержание ПО. Такая надстройка защищает БД от вредных, разрушающих ее запросов, помогает отыскивать ошибки, неисправности БД, которые были допущены при проектировании и появились в результате эволюции БД как постоянно развивающейся системы. При эксплуатации таких больших БД кибернетики столкнулись с характерными свойствами больших систем, в частности при увеличении объема информации на порядок изменяются закономерности этой системы. Поэтому после решения теоретических и инженерных проблем в области БД и СУБД потребовались научно-технические разработки БЗ и СУБЗ для поддержания БД большого объема или ИБД, содержащих как минимум две базы: БД и БЗ.

Теоретические исследования в области БД, БЗ и ИБД, по существу, создали для кибернетики теорию информации человеко-машинного типа, которая необходима при проектировании кибернетических систем. Появились модели, средства машинного представления семантической, смысловой информации в виде фреймов, семантических сетей и т. д. Все они графически могут быть изображены в виде графа с корневой вершиной, соответствующей основному понятию, смысл которого требуется раскрыть (к примеру, полупроводник), а другие вершины, соединенные между собой ребрами, соответствуют подчиненным, вторичным понятиям. Ребра несут также семантическую нагрузку, задавая ролевые отношения между понятиями (какую роль играют отдельные понятия в общей схеме основного понятия). Эта структура соответствует тезаурусу в информационно-поисковых системах (см. т. 1 «Основ кибернетики» 1-го издания). Но в кибернетике, теории искусственного интеллекта этот термин «не прижился», а вместо термина «тезаурус» используют базу знаний. На сегодняшний день технология проектирования с помощью ЭВМ любых систем является технологией проектирования программных комплексов, реализация которых на ЭВМ дает на выходе, в частности, документацию на серийное производство. Начиная с концепции первой в мире технологии программирования IPT (Improving Programming Technology), предложенной фирмой в 1972 г., эта технология состоит из последовательности описаний (рис В.3) целевого, функционального, логического, алгоритмического, информационного, программного, документального.

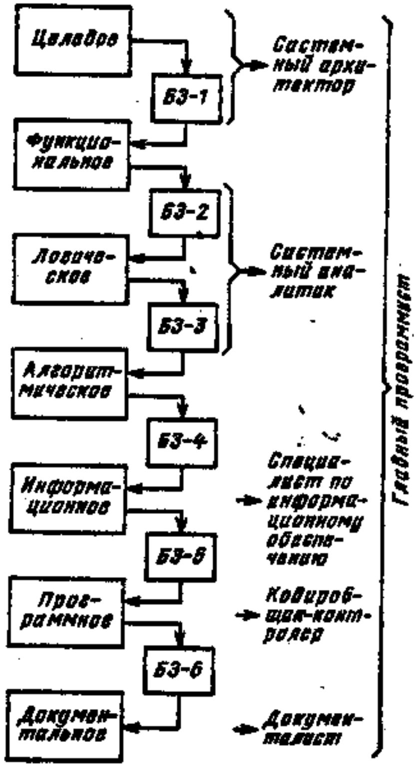


Рис. В.3. Технология проектирования

В создании программного комплекса (или системы) участвуют следующие специалисты: главный программист, системный архитектор, системный аналитик, алгоритмист, специалист по информационному обеспечению (информатик), программист (кодировщик), контролер, документалист.

Сам процесс проектирования представляется как последовательная смена одного описания другим. Поэтому любой процесс проектирования может рассматриваться как система переформулирования.

На самом деле процесс проектирования осуществляется последовательно-параллельным способом, когда обработка (разработка) одного описания происходит одновременно с формированием других, при этом возможен возврат к предыдущим описаниям. Как правило, выбирается не оптимальный вариант проектирования, а подходящий, допустимый, который на каждом этапе (описании) удовлетворяет ограничениям (спецификациям). Может иметь место такая ситуация, что на данном этапе (при наличии ограниченных ресурсов, имеющихся на данном этапе) могут не удовлетворяться ограничения за счет неудачного выбора на предыдущем (предыдущих) этапе. В этом случае осуществляется возврат на предыдущий (предыдущие) этап и повторение выбора. Может получиться, что при выделенных (малых) ресурсах на каждом этапе поставленная задача проектирования не может быть выполнена. Тогда ставится задача увеличения ресурсов. Таким образом, любая САПР является информационной (кибернетической) системой, в которой формируются последовательные описания. Если формирование описания и переход от одного описания к другому осуществляются с помощью БЗ, то такая технология будет интеллектуальной.

Второй особенностью кибернетики является моделирование — сведение процессов управления к кибернетическим моделям. Одна и та же модель может применяться для разработки оптимального управления парикмахерскими, счетчиками космических частиц, системой противовоздушной обороны, процессами наследственности и пр. Как только процесс управления сведен к одной из кибернетических моделей, далее уже безразлично, какой конкретно это процесс: физический, экономический и т. д., так как кибернетические методы исследования всех моделей одни и те же. В самом этом смысле методы кибернетики аналогичны методам теории колебаний, которую в настоящее время можно, по-видимому, рассматривать как раздел ккибернетики, занимающийся управлением процессами возникновения и поддержания колебаний с заданными параметрами [9].

Приведем несколько примеров кибернетических моделей, которые широко используются при исследовании систем управления, прежде всего модель в виде автоматической системы регулирования (АСР). Вся система регулирования представлена в виде структурной схемы (рис. В.4), состоящей из отдельных звеньев $Y_1(S)$, $Y_2(S)$. Как правило, во всех системах имеется звено сравнения управляющего (входного) сигнала $x_{вх}$ с сигналом обратной связи (в данном случае $x_{вых}$). Принцип обратной связи, позволяющий получить более высокое качество управления, широко используется в кибернетике. Каждое звено модели исследуемой системы описывается дифференциальным уравнением или передаточной функцией (в случае линейных систем с постоянными параметрами). Эта кибернетическая модель преобладает при исследовании систем управления ракетой, электродвигателем и т. д. Структурный подход к исследованию позволил создать частотный графоаналитический метод расчета и проектирования [19].

Инженера-кибернетика, который проектирует АСР с электродвигателем, мало интересуют число обмоток этого двигателя, его электрические параметры и т. д. Он прежде всего хочет знать частотные свойства электродвигателя и других элементов системы, их амплитудно- и фазочастотные характеристики. Ему важно определить, какие частоты пропускаются, какие подавляются, какое запаздывание по фазе вносится на определенных частотах, критических для системы (частоте среза и пр.). По существу, в течение многолетней практики проектирования выработался специфический частотный язык, с помощью которого инженер как бы общается с системой. Этот пример служит хорошей иллюстрацией информационного подхода к исследованию кибернетических систем.

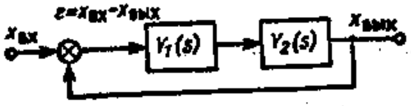


Рис. В.4. Структурная схема АСР

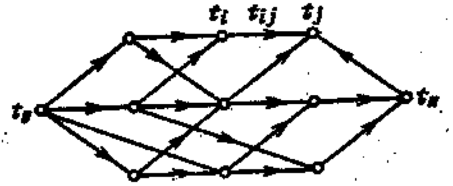


Рис. В.5. Сетевой график

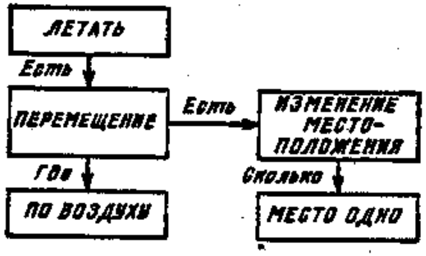


Рис. В.6. Граф-структура фрейма ЛЕТАТЬ

К сожалению, для расчета других кибернетических моделей (сетевых, лингвистических) нет такого же эффективного метода. Есть попытки использовать модель автоматического регулирования при исследовании экономических систем типа промышленного предприятия. Однако для таких систем, которые в кибернетике называются большими системами, модель, использующая аппарат дифференциальных уравнений, малопригодна. В этом случае наиболее применима сетевая модель в виде графа, называемого транспортной сетью или сетевым графом (рис. В.5). Каждая отдельная работа (обточка детали, монтаж электронной схемы и т. д.) изображается в виде дуги графа, причем вершина, из которой она исходит, соответствует началу работы, а вершина, в которую она входит,— концу работы. Сетевой график работ может насчитывать несколько тысяч ребер. Важно заметить, что поскольку время выполнения работы считается случайной величиной, то сетевой график — это вероятностная модель системы планирования работ. На рис. В.6 представлена семантическая модель понятия ЛЕТАТЬ в виде граф-структуры фрейма. У компьютера при хранении информации в таком виде появляется способность понимать, различать, что такое летать (это перемещение по воздуху). Подробно различные концепции кибернетических моделей будут описаны во втором томе «Основ кибернетики», здесь же рассмотрим этот вопрос кратко. Прежде всего, так как кибернетика стала заниматься семантическими, смысловыми, понятийными аспектами управления и информации, то это оказало существенное влияние на кибернетическое моделирование. Как известно из математической логики (исчисления высказываний и предикатов), семантика, семантическая интерпретация моделей может быть реализована только с помощью конкретной предметной области (к примеру, аэропорт «Внуково», Одесский морской порт). Поэтому имеет место класс кибернетических моделей, ориентированных на ПО, при этом общность и эффективность теории составления этих моделей невысоки. Можно рассчитывать только на методику составления таких моделей.

Модели, независимые от ПО

Модели, настраиваемые на ПО

Модели, ориентированные на ПО

Рис. В.7. Разновидности кибернетических моделей

Но в последнее время усилия ученых и инженеров сосредоточились на развитии теории и практики технологий создания семантических моделей (в том числе интеллектуальных систем), которые позволяют с помощью комплекса средств настраивать модель-основу на данную предметную область. Поэтому наряду с автономными моделями в кибернетике существуют настраиваемые на ПО и ориентированные на ПО модели (рис. В.7). Под влиянием лингвистики и семиотики (науки о знаковых системах) в кибернетике появились синтаксические, семантические и прагматические модели (рис. В.8). К примеру, фразы на естественном языке

«комната сына убирается матерью»,

«мать убирает комнату сына»

имеют разные синтаксические структуры (модели), а семантическая (глубинная) модель у них одна в виде семантической сети, представленной на рис. В.9. Второй пример, поясняющий эту классификацию,— сборка подъемного крана из детского конструктора. Синтаксические знания обеспечивают правильность соединений планок болтами, но они не гарантируют от сборки ошибочных с точки зрения семантики конструкций подъемного крана со стрелой, прикрепленной снизу платформы. Прагматика в этом примере предусматривает создание не только семантически правильных подъемных кранов, но и оптимальных, к примеру из минимального числа деталей, минимальной массы и т. д. В терминологии теории искусственного интеллекта синтаксическая часть информации связана с базой данных, семантическая — с базой знаний, а прагматическая — с базой целей.

Человеко-машинная особенность моделей и методов кибернетики заключается, в частности, в создании смешанных систем, в которых функционируют три части: техническая (машинная), человеческая (биологическая) и смешанная (человеко-машинная). Особенно это относится к системам искусственного интеллекта, которые тем более совершенны, чем более развита человеко-машинная часть. В этом проявляется основная задача кибернетики, сформулированная в книге Н. Винера [1].

Синтаксические

Семантические

Прагматические

Рис. В.8. Классификация кибернетических моделей

Современные АСУ являются человеко-машинными системами, в которых совместно функционируют коллективы людей и технических устройств. Семантику и прагматику информации на сегодняшний день можно реализовать только человеко-машинными методами. В кибернетике, особенно в разделе систем искусственного интеллекта, моделируется деятельность человека, создаются усилители интеллектуальных способностей человека и более совершенные по сравнению с человеческими и машинными человеко-машинные системы, в которых гармонично сочетаются и дополняются способности и появляется проблема изучения новой, реально-существующей проблемно-предметной области или среды — человеко-машинной. Здесь было подмечено явление деинтеллектуализации задач искусственного интеллекта. Когда для исходной задачи принятия решений составляются математические модели, алгоритмы, информационно-программное обеспечение и это обеспечение загружается в память ЭВМ, то эта задача перестает быть предметом искусственного интеллекта (ИИ), она приобретает синтаксическую форму данных, а ИИ имеет место там, где есть семантика, прагматика и т. д. Тесно связан с информационной и человеко-машинной концепцией лингвистический, или естественно-языковый, подход.

Во-первых, естественный язык развивался вместе с естественным интеллектом и всегда был основным средством выражения мыслей. В теории ИИ считается, что лучшую среду для моделирования, чем естественный язык, трудно придумать, так как в нем в неформальном виде содержатся все основные свойства и особенности естественного интеллекта (правда, рассматриваемого только на вербальном уровне), т. е. таких явлений мыслительных процессов, которые выражаются фразами естественного языка. Поэтому прогресс в управлении и кибернетике тесно связан с открытыми, формальными конструкциями, новыми машинными реализациями естественного языка. Кроме того, в человеко-машинной среде, которая создается в современных системах электронной обработки данных и АСУ, общение человека с ЭВМ осуществляется на искусственном и естественном, или близком к естественному, языках, в которых используются все достижения лингвистики и пр. И наконец, языки программирования имеют тенденцию создавать комфорт, удобство пользователю и (как вариант) приближаться к естественному языку. Первая математическая модель естественного языка Хомского использует процедуры, продукции, переписывания, подстановки, которые стали использоваться во всех языках программирования, в том числе в самом популярном в последнее время языке PROLOG [32].

Рис. В.9. Пример семантической модели в виде семантической сети

Лингвистическая модель, обладающая большими логическими возможностями, наилучшим образом описывает поведение систем, в которых существенную роль играют люди, так как представляется возможным (через семантику) учесть элементы их разумного, творческого поведения. Термин «лингвистический» означает использование методов, развиваемых в математической логике, которая исследует алгебраическими методами естественные языки.

Применительно к большим системам чаще используется термин «семиотическая система» [12, 15].

Семиотические, или знаковые, системы или модели в широком смысле определяются как системы, состоящие из знаков, между которыми установлены определенные соотношения. В качестве знаков могут фигурировать мимика, жесты человека или животных, рукописные знаки и т. д.

В широком смысле под лингвистическими моделями понимаются системы, состоящие из слов, подчиняющихся определенной взаимосвязи. В качестве слов могут быть использованы слова и фразы естественного языка, жесты, счетные палочки и т. д. При таком понимании лингвистические методы мало чем отличаются от семиотических методов исследования. В семиотических методах в отличие от лингвистических часто подчеркивается физическое существование семиотических моделей аналогично тому, как специалисты по АСР всегда имеют в виду конкретную АСР (например, АСР давления пара, уровня жидкостей и т. д.). Однако будем считать, что любая модель системы (сетевая, массового обслуживания и т. д.) становится семиотической, когда ее записывают в виде программы в кодах машины или на каком-нибудь алгоритмическом языке (СЛЕНГ или СИМУЛА) для моделирования на ЭВМ.

Третьей, не менее важной, особенностью кибернетики, ставшей наиболее четкой в связи с разработкой АСУ, является системный подход к исследованию процессов управления. Системность предполагает единое комплексное рассмотрение и проектирование системы управления в целом и представляет достаточно сложное переплетение организационных, информационных, математических, программных и технических средств и приемов. Первый этап разработки АСУ состоит в системном обследовании объекта, который предполагается автоматизировать, в результате чего уточняются его функциональная структура, объем массивов информации и документов, циркулирующих в нем. В современных интеллектуальных системах с базами знаний системное обследование приобрело первостепенное

значение, так как вручную создать базу знаний реальной предметной области невозможно, и поэтому строятся автоматизированные системы обследования, которые иногда называются системами формирования знаний.

Системность достигается единым информационным, математическим и техническим обеспечением в отдельных функциональных подсистемах (например, в финансово-бухгалтерской, планово-производственной, материально-технического снабжения и т. д.), при этом во всех подсистемах используются технические устройства одной из вычислительных систем (например, ЕС ЭВМ), одна и та же система автоматизированной обработки данных с СУБД типов ОКА, СПЕКТР, единое представление информации на магнитных носителях, одна и та же система программирования с единой библиотекой стандартных программ и операционной системой. С другой точки зрения системность предусматривает оптимальное распределение ресурсов (финансовых, весовых, объемных, временных и т. д.) между отдельными частями технических средств системы, такими, как терминальные устройства, линии связи, внешние устройства ЭВМ, оперативная память ЭВМ, процессор ЭВМ, а также оптимальное распределение функционального вклада в систему между математическими моделями, алгоритмическим, программным и техническим обеспечениями. Системность не допускает, например, простаивания терминалов с абонентами из-за ограниченных возможностей линии связи или процессора ЭВМ из-за неподготовленности информационных массивов. В этом случае следует вложить больше средств в систему связи за счет уменьшения средств на собственно вычислительную машину. Нередки случаи, когда в результате некачественно выполненного системного обследования требования по точности в одних подсистемах нерационально высоки по сравнению с предельно реализуемым значением ошибок в других подсистемах. При этом получается, что дополнительные ресурсы, вложенные для получения этой высокой точности, неоправданны, так как нет положительного эффекта для всей системы.

Четвертая особенность кибернетики — это вероятностный, статистический подход к процессам управления. Эта концепция взята из статистической физики. Известно, что поведение газа в сосуде определяется случайным движением отдельных молекул. Однако молекулярная физика, статистически усредняя движения молекул, получила детерминированные законы поведения газа (законы Бойля—Мариотта, Гей-Люссака).

Аналогично этому можно считать вызовы на телефонной станции случайными событиями во времени, так как каждый вызов определяется большим количеством факторов, учесть которые невозможно. Однако зная статистические характеристики случайных вызовов, с помощью кибернетической модели массового обслуживания можно определить оптимальные законы управления телефонной связью.

Кибернетика считает, что любой процесс управления подвержен случайным, мешающим воздействиям. Это в одинаковой мере относится к любой системе управления. Например, на производственный процесс влияет так много факторов (состояние станка, качество материала, своевременность доставки комплектующих изделий и пр.), что учесть их детерминированным образом нельзя. Поэтому считается, что на производственный процесс воздействуют случайные сигналы, и при управлении производством следует исходить из того, что планирование может быть только вероятностным и выполнение плана к определенному сроку можно ожидать с какой-то вероятностью (например $p=0{,}9$). Так, радиолокационная антенна в своей работе подвержена случайным флюктуациям, вызванным отражением сигнала от цели и собственными шумами приемника. Поэтому и здесь можно обсуждать только вероятность отклонения антенны от направления на цель.

Иногда для уменьшения объема детерминированной информации с сохранением ее основных черт кибернетика искусственно вводит случайность. Например, производственный процесс состоит в выпуске нескольких сотен типов электронных блоков, каждый из которых изготовляется в разных количествах (от сотни до тысячи). Вместо всего множества этих блоков можно ввести некоторые абстрактные эталонные блоки, появляющиеся (материализующиеся) случайным образом в виде реальных блоков с частотой появления, определяемой относительным количеством этих блоков. В результате такой процедуры сложная детерминированная исходная информационная система заменяется более простой эквивалентной вероятностной информационной системой.

Аналогичная процедура часто применяется при решении задач оптимизации и синтеза алгоритма управления. Такие задачи (например, распределение ресурсов или синтез конечных автоматов) в детерминированной постановке требуют нереально большого объема вычислений. При переводе этих задач из детерминированных в вероятностные существенно сокращается объем вычислений, хотя это и не позволяет получить точного решения.

Другой моделью, близкой к вероятностной, является модель нечетких множеств, которая иногда связывается с термином «теория возможностей». Как частный случай она включает теорию вероятностей. Однако при привлечении вероятностных моделей к описанию неизвестной ситуации в управлении искажается, сужается реальная ситуация, которая, к примеру, имеет место в системах электронной обработки данных, в АСУ.

Здесь характеристики просто неизвестны, вероятностная модель неправильно моделирует, отражает это свойство неизвестности, неопределенности. Модель, использующая нечеткие множества с функциями принадлежности, более близка к действительности.

Следует заметить, что гегемония вероятностных методов (так же как методов оптимизации), которая проявлялась в первые два десятилетия становления кибернетики, существенно поколебалась и уменьшилась главным образом за счет появления большой памяти ЭВМ и методов искусственного интеллекта [33]. Прежде всего потребовалась индивидуальность управления, в настоящее время необходимо управлять хорошо не в среднем, а в каждом индивидуальном случае. Поэтому статистические методы обладают определенными недостатками из-за требования усреднения по большому количеству реализаций (закон больших чисел). Как уже указывалось, процесс принятия решений человеком внутренне противоречив, т. е. в зависимости от ситуации одна и та же посылка может вызывать разные, противоположные иногда следствия, и при усреднении получается нулевой результат. Поэтому стали развиваться кибернетические модели, рассчитанные на индивидуальное, разовое управление с помощью больших хранилищ и средств запоминания в структурированном виде отдельных реализаций.

В.3. МЕТОДЫ КИБЕРНЕТИКИ

Вряд ли целесообразно излагать все методы кибернетики; их достаточно много, и они подробно будут рассмотрены в последующих разделах. Поэтому здесь будут изложены только основные инженерные методы, применяемые почти во всех разделах кибернетики для исследования большинства кибернетических моделей.

Начиная с АСР и кончая системами искусственного интеллекта, в кибернетике при инженерных расчетах и проектировании широко применяют структурные методы, использующие геометрическое представление системы, т. е. методы или процедуры в виде графа.

Структурные методы исследования АСР вытеснили все другие методы. Они используют геометрическое представление системы в виде направленного графа и при структурных преобразованиях применяют правило Мэзона. Их применяют при расчете электрических цепей, электрических машин, электронных схем высокочастотных электромагнитных линий и устройств. Изображение в виде графа применяется в системах массового обслуживания, теории игр, теории детерминированных и вероятностных автоматов, при сетевом планировании, матричных расчетах и, наконец, в лингвистических методах анализа и синтеза. В теории искусственного интеллекта применяется пространство состояний, которое представляет собой семантический граф, аналогичный фазовому пространству состояний в физике и теории колебаний. Во всех случаях геометрическое изображение в виде графа включает дополнительный участок мышления инженера, помогая ему найти оптимальное решение. Появляется даже термин «компьютерная геометрия», которая в идеале должна помочь инженеру оперировать (с помощью компьютера) переходами с помощью тензоров из трехмерного в четырехмерное пространство и т. д., делать необходимые конструкции в n-мерном пространстве и возвращаться в трехмерное пространство с готовой оптимальной конструкцией.

До последнего времени преобладающим инженерным методом расчета систем управления был метод дифференциальных уравнений, который применительно к АСР был трансформирован в частотный структурный метод. Применение этого метода обусловливалось тем, что в центре внимания научно-технической мысли были системы управления, поведение которых достаточно хорошо описывалось аппаратом дифференциальных уравнений. Это системы типа наведения ракеты, управления антенной. При управлении большими системами (промышленным предприятием) также применялся аппарат дифференциальных уравнений, однако здесь он, как правило, плохо описывал явления и мало помогал при расчете и проектировании. Поэтому для описания таких систем метод дифференциальных уравнений был заменен топологическим методом с использованием алгебры, логики, теории графов, комбинаторного и тензорного исчислений. Это дало, в частности, в сетевых методах планирования возможность учитывать отдельные дуги, изображающие этапы работы, и в то же время получать обобщенные характеристики всей системы, такие, как минимальный разрез, максимальный поток, минимальное дерево и др.

Такое положение способствовало развитию и популярности в кибернетике методов так называемой дискретной математики, которая не использует таких понятий непрерывной математики, как непрерывность, дифференцируемость. Вопрос о диалектическом взаимоотношении непрерывной и дискретной математики, об отдельных исторических этапах интенсивного развития одних и замедленной эволюции других ее методов неоднократно обсуждался. Интенсивное внедрение кибернетики и ЭВМ в различные области науки и техники оказало влияние и на такую фундаментальную науку, как математика, инициировав развитие дискретных методов исследования. Очевидно, что современные высокие темпы научно-технической революции требуют включения в общематематическую подготовку инженеров разделов дискретной математики.

Уже в первых АСР применялся принцип обратной связи, который в дальнейшем перерос в методы самонастройки, самоорганизации и самоусовершенствования. Так, при распознавании образов (печатных букв) с помощью специализированных вычислительных машин или вычислительных сред используется режим обучения с учителем и без учителя, в обученной машине предусматривается возможность дообучения, т. е. режим самоусовершенствования, самонастройки. Проблемы обучения (настройки) лингвистической математической среды семантике (правилам построения правильных решений) аналогичны обучению человека в учебных заведениях.

Особое значение эти методы приобрели в системах искусственного интеллекта в связи с требованием индивидуальности. Комплекс средств позволяет создать интеллектуального помощника пользователя — базу знаний, адаптируемую под его стиль и навыки. В таких системах работают специальные механизмы усвоения знаний, которые автоматически «арестовывают» поступающую информацию, если она представляет интерес.

В традиционных, классических методах кибернетики при выборе оптимального варианта при проектировании или управлении требуется решать задачу оптимизации. Поэтому на первом этапе развития кибернетики большое место занимали классическое и расширенное вариационное исчисление, линейное и нелинейное программирование и т. д., которые достаточно полно рассмотрены в первом издании т. 1 «Основ кибернетики». Однако с широким распространением этих методов значение их существенно уменьшилось и появились новые человеко-машинные методы поиска решения. Системы электронной обработки данных, АСУ, интеллектуальные системы являются большими системами из-за больших емкостей информации (БД и БЗ), хранимых в ЭВМ. Если первоначально имеется «небольшая» система типа САР, то в результате превращения ее в интеллектуальную она становится большой, так как в ней создаются БД и БЗ. В последнее время все чаще стали использовать «крамольные» для ортодоксальных кибернетиков высказывания, что в больших системах отсутствует «оптимум» и бессмысленно решать задачу оптимизации (искать несуществующее состояние!). Методы сведения большой системы к малой, упоминавшейся выше, за счет введения вероятностных моделей и моделей в виде дифференциальных уравнений сильно упрощают реальную систему. Кроме того, появилось требование *индивидуальности* управления. Поэтому возникли новые концепция и человеко-машинные методы поиска подходящего допустимого решения, как это делается естественным интеллектом человека. С помощью компьютера создается система поиска допустимого решения путем создания (параллельно) последовательного процесса, состоящего из отдельных этапов. Переход от предыдущего этапа к последующему разрешается, если удовлетворяются условия (или фильтры), причем используется любое, первое попавшееся условие, которое добавляет информацию (деформирует) к пропускаемым данным и знаниям. Если на данном этапе фильтр не пропускает информацию к следующему этапу, то осуществляется возврат (back tracking) к предыдущему этапу, на котором ищется новое допустимое решение, а старое ликвидируется (это вторая попытка), и т. д. Возможен отказ от поиска, когда все возможности исчерпаны. Тогда конструктор вносит изменения в систему, расширяя фильтры на каждом этапе и дополняя БД и БЗ человеко-машинной системы оптимизации. Это является дальнейшим развитием и обобщением понятия обратной связи в САР. В интеллектуальную систему управления (самонастраивающуюся систему) встроена человеко-машинная система оптимизации с обратной стратегией (back tracking) поиска, которая постоянно находится в работе (так же как обратная связь в дискретных САР). Появилось большое количество модификаций этого метода, но изложенные выше идеи присутствуют во всех модификациях.

Но никакая математическая или кибернетическая модель реальной системы не может полностью отразить все свойства и особенности, неизбежны искажения и упрощения. Кроме того, как уже указывалось, математическое описание многих реальных систем настолько сложно, что не поддается аналитическим расчетам, и приходится прибегать к моделированию на ЭВМ. Поэтому одним из основных методов кибернетики стал метод моделирования процессов управления на ЭВМ. Для этого разработаны специальные проблемно-ориентированные языки моделирования, такие, как СИМУЛА, САМСКРИПТ, СЛЭНГ и другие, а также информационно-программные средства общения человек—машина, включающие языки-диалоги, и технические средства общения типа дисплея, использующие электронно-лучевые трубки.

Очень эффективным средством моделирования больших систем является метод создания прототипа, в котором есть большое количество модификаций. К примеру, для создания эффективной системы общения на естественном языке (или системы электронной обработки данных, или системы искусственного интеллекта) нужны мощные ЭВМ с большими памятью и быстродействием. Но их еще нет или они дороги (аренда их), поэтому конструктор создает прототип системы на плохой ЭВМ, но в переносимом (мобильном) на другую ЭВМ варианте. Он работает очень медленно, т. е. вы сегодня задаете системе вопрос, а она отвечает через несколько дней. Но если в остальном система удовлетворяет заказчика, то можно продолжать работы по устранению замечаний или переносить на другую ЭВМ, закупать дорогую аппаратуру и т. д. Этот метод, естественно, возможен в коллективах, не обладающих достаточно большими финансовыми возможностями.

В.4. КИБЕРНЕТИКА И ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ

Вопрос о месте вычислительных машин в кибернетике трактуется по-разному. Кибернетика является самостоятельной наукой со своими определениями и методами, так же как физика или химия, и имеет право на существование независимо от вычислительных машин. Однако в отличие от всех других наук она теснее связана с вычислительными машинами и зависит от них более сильно, чем, например, физика от физических приборов. Это не означает, что кибернетика — это вычислительные машины. Инженер-кибернетик должен знать принцип работы вычислительных машин, но не обязан уметь их создавать, настраивать или эксплуатировать. Для кибернетика обязательно уметь доводить свои задачи до составления программы на каком-либо алгоритмическом языке типа АЛГОЛ, уметь общаться на искусственном или естественном языке. Следует заметить, что кибернетиков в последнее время все чаще и чаще стали привлекать к созданию начальной архитектуры, конфигурации идеологической основы компьютеров, особенно в связи с проектом вычислительных систем пятого поколения, так как считается, что успеха в создании новых кибернетических систем, типа систем икусственного интеллекта, можно достичь только при гармоническом сочетании и одновременном проектировании информационно-программного (software) и аппаратного (hardware) обеспечений. Вместе с развитием и становлением кибернетики и, главным образом, систем электронной обработки данных ЭВМ прошли несколько стадий развития. Первое и второе поколения характеризовались вычислительными функциями и представляли собой большие арифмометры.

По-видимому, основное различие между ними заключалось в элементной базе: первое поколение ЭВМ было на лампах, а основной парк ЭВМ второго поколения реализовался на полупроводниках. ЭВМ третьего и четвертого поколений в основном предназначались для электронной обработки данных. В них существенно развиты все виды памятей — оперативная, дисковая и ленточная. Наиболее важное значение приобретают дисковая и оперативная памяти ЭВМ или ВС. Если в ВС-1 и ВС-2 основной целью было большое быстродействие, то в ВС-3, ВС-4 основной целью становится большой объем оперативной и, главным образом, дисковой памяти. Требование быстродействия в ВС-3 и ВС-4 как бы отодвинулось на второй план, в АСУ организационного типа можно было обрабатывать данные достаточно медленно. Переход от ВС-1 и ВС-2 к ВС-3 и ВС-4 происходил постепенно. В частности, примером может быть ЭВМ «Минск-32» (главная «рабочая лошадка АСУ» в нашей стране). Когда было принято решение о массовом использовании этой машины в АСУ, ее оперативная и внешняя память, устройство ввода и вывода существенно изменились и расширились, что привело, в частности, к увеличению стоимости этой ЭВМ примерно в 3 раза. Но в этой серии ЭВМ не было самого главного — дисковой памяти, без которой электронная обработка данных невозможна. Поэтому полноценным инструментом для обработки данных стали ВС серии ЕС ЭВМ с дисковой памятью, развитой операционной системой и другим информационно-программным обеспечением. Высшие модели ВС-4 этой серии (ЕС-55, ЕС-60, ЕС-61) приспособлены для размещения банков данных с СУБД, системой коллективного пользования ЭВМ, обеспечением функционирования сетей передачи данных с терминала и ЭВМ и для распределенной обработки данных. В ВС-4 особое внимание уделяется удобству общения с пользователем, что позже будет называться свойством «дружелюбия». В процессе развития ВС-3 и ВС-4 появились развитые терминалы типа автоматизированных рабочих мест (АРМ) с алфавитно-цифровым и графическим дисплеями, графопостроителями, автономными устройствами подготовки данных на магнитных носителях (лентах, дисках).

Системы ВС-3 и ВС-4 послужили средством построения индустрии электронной обработки данных, или просто индустрии данных. И наконец, основное назначение ВС-5 — электронная обработка знаний, построение индустрии знаний. В них предполагается гармоническое сочетание информационно-программного и аппаратного обеспечений. Появилось много направлений разработки ВС-5 с новой архитектурой, отличной от классической, фон неймановской, которая использовалась в ВС-1—ВС-4. Основная их особенность заключается в создании процессоров различного функционального назначения (ЛИСП-процессор или процессор символьной обработки, логический процессор или ПРОЛОГ-компьютер, компьютер реляционной базы данных и т. д.) или компьютера «параллельных вычислений». Появились потоковые, конвейерные, рекурсивные компьютеры. Таким образом, в последние годы наблюдается существенное влияние кибернетики на вычислительную технику. В частности, появились интеллектуальные и персональные компьютеры.

Конфигурация ВС-5 (рис. В.10) напоминает функциональную схему системы искусственного интеллекта [22]. Следует заметить, что при реализации систем искусственного интеллекта, которые должны работать в реальном масштабе времени, потребовалось помимо большой памяти высокое быстродействие, только вместо оценки быстроты арифметических действий «бит в секунду» была введена новая единица быстродействия — «число логических операций вывода в секунду». Интеллектуальные компьютеры должны помогать пользователю уточнять задачу, доводить ее до формальной постановки, формировать математическую модель, составлять алгоритм, програму и т. д.

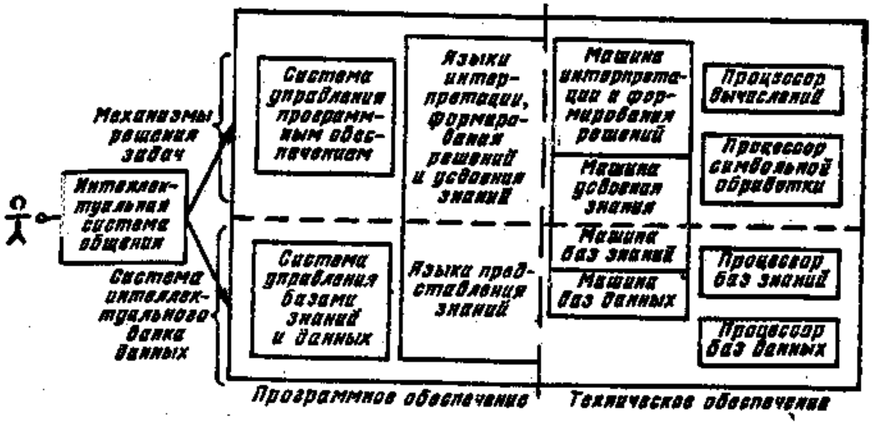


Рис. В.10. Конфигурация ВС-5

Появилось еще одно новое направление в ВС, связанное с микропроцессорами, когда информационно-программное обеспечение как бы реализуется за счет создания системы микропроцессоров [22, 24].

В.5. СПЕЦИАЛЬНЫЕ И ПРИКЛАДНЫЕ ВОПРОСЫ КИБЕРНЕТИКИ

В процессе формирования и развития кибернетики как науки об управлении появились понятия и разделы, специфичные только для кибернетики, в частности понятие большой системы, под которой понимается такая кибернетическая система, которая насчитывает несколько тысяч или десятков тысяч компонентов (степеней свободы), например промышленное предприятие, насчитывающее 10—15 тыс. рабочих. Такое предприятие может быть описано в виде сетевого графика, насчитывающего несколько десятков тысяч ребер, где каждое ребро — это компонент (или степень свободы). Оно также может быть представлено в качестве сети массового обслуживания логических, лингвистических операторов; в этом случае каждый рабочий или группа рабочих описывается локальной системой массового обслуживания или логическим оператором, которых в общей сложности может насчитываться несколько тысяч.

Другими примерами большой системы могут служить телефонная сеть или система парикмахерских большого города. Вызовы на телефонную станцию, так же как клиенты в парикмахерских, приходят случайно и в большом количестве — несколько тысяч или десятков тысяч.

Примером большой системы является также многослойная распознающая машина, состоящая из нескольких тысяч нейронов или пороговых элементов и используемая в качестве модели мозга человека.

Большую систему часто сравнивают с газом, заключенным в сосуд. Большое количество молекул совершают хаотическое движение. Методами молекулярной физики (микроподход) учитывают движения отдельных молекул, выводят усредненные характеристики поведения всей системы. Если же газ в сосуде рассматривают как большую систему, то движениями отдельных молекул не занимаются, а сразу выводят законы для всего объема газа (макроподход). Частным подклассом больших систем являются автоматизированные системы управления, для которых характерны присутствие человека или группы людей и человеко-машинные методы управления.

Но наиболее характерным представителем больших систем в настоящее время является банк данных, АСУ или информационно-поисковая система с банком данных и т. д. Самая главная особенность больших банков данных, емкость которых превышает 10 Мбайт, заключается в том, что свойства их, закономерности в корне изменяются при увеличении объема информации на порядок (100 Мбайт вместо 10, 1000 Мбайт вместо 100 и т. д.). Это положение напоминает классическую и неклассическую (релятивистскую) механику (физику). Только в физике параметром, от которого зависит закономерность явлений, является скорость. Большие банки данных позволяют моделировать на ЭВМ неформальные, вероятностные системы, хотя само устройство ЭВМ детерминированное и формальное. Дело в том, что, как это имеет место с псевдослучайными последовательностями чисел, полученных с помощью ЭВМ, детерминизм и формальность можно обнаружить при длительном наблюдении (более 100 лет), при большом переборе вариантов, что в конце концов тоже сводится к большому времени наблюдения. Крупным достижением кибернетики является эволюционный принцип разработки вечной кибернетической системы (в том числе системы электронной обработки данных). Благодаря встроенной системе проектирования современная кибернетическая система постоянно подстраивается под новые требования пользователя и заказчика. Независимость от технических средств системы (ЭВМ), которые обновляются примерно каждые пять лет, обеспечивается так называемым свойством мобильности. Информационно-программное обеспечение проектируется на языках программирования высокого уровня: ПАСКАЛЬ, ЛИСП и др., трансляторы для которых будут созданы на всех компьютерах.

С течением времени в системе создаются базы данных и знаний, программные средства, приспособленные для коллектива пользователей для функционального назначения системы, ориентированной на предметно-проблемную область. Создается техническая система, очень похожая на человеческий организм, биологическое существо, которое развивается, размножается, умирает и живет после смерти в следующих поколениях, передавая им информацию. Здесь снова реализуется один из основных принципов кибернетики, сформулированный Винером,— создание общих для живой и неживой природы методов управления.

В последние десятилетия центральной проблемой кибернетики стало построение индустрии искусственного интеллекта [22], в которой нашли свое воплощение основные идеи кибернетики:

переработка информации в целях управления;

формирование решений для управления;

разработка общих для живой (человека) и неживой природы моделей принятия решений.

Поставлена задача разработки универсальной технологии проектирования и внедрения системы искусственного интеллекта (это основное содержание проектов ВС-5). Это направление стало основным стратегическим направлением научно-технического прогресса, которое определяет темпы развития, темпы повышения производительности труда, качество изготовляемой продукции и т. д. Внедрение этой технологии должно качественно изменить характер деятельности любого «труженика», особенно в управленческой деятельности, где будут интенсивно внедряться безбумажные интеллектуальные технологии. В период 1982—1986 гг. наблюдался подъем в области разработки технологий проектирования систем с базами знаний (экспертных систем), которые появляются ежемесячно в США. Это оказало существенное влияние на развитие и становление теории искусственного интеллекта, так как потребовалось создание математических, теоретических методов проектирования инструментальных средств, пригодных для любых предметных областей. Поэтому в силу ограниченности возможностей ЭВМ по быстродействию и памяти разрабатывались узко направленные системы представления знаний, механизмы и т. д., которые должны были обеспечить эффективную работу конкретной системы искусственного интеллекта. Теоретический уровень разработок этого этапа был занижен. Появились две тенденции развития научно-технического направления: разработка новых систем искусственного интеллекта и внедрение интеллектуальности в существующие системы, главным образом, за счет введения базы знаний. Искусственный интеллект оказал большое влияние на кибернетику, вдохнув в нее второе дыхание, сформировав новые основы теории семантической (и прагматической) информации, стимулировав новые разделы математики, названные специальными главами дискретной математики, сформулировав новые требования к компьютерам и дав новый импульс к разработке ЭВМ нового типа.

Большое значение эта проблема искусственного интеллекта имеет в разработке автоматизированных систем проектирования новых систем. Современным высоким темпам научно-технической революции совершенно не удовлетворяют медленные ручные методы проектирования. Дело в том, что идеи, заложенные в разработку новой системы, начинают устаревать спустя один-два года после начала работ, а к моменту запуска системы в серию (примерно через три—пять лет после начала работ) совершенно идеологически устаревают. Выход из такого положения один — сокращение сроков разработки и запуска в серию систем за счет внедрения интеллектуальных автоматизированных систем проектирования на всех этапах разработки — от эскизного проекта до изготовления серийной документации, а также внедрение автоматизированной системы управления самим серийным производством, при этом перед началом работ принимается стратегия разработки не конкретной системы, а поколения систем, определяется их облик и готовится информационное, программное и техническое обеспечения для автоматизированной системы проектирования.

В.6. СТРУКТУРА КИБЕРНЕТИКИ

Известно несколько схем, представляющих структуру кибернетики. Наибольший интерес представляет схема Ляпунова—Яблонского [17]. Приводимая ниже структура кибернетики разработана в результате чтения курса кибернетики в Московском инженерно-физическом институте на протяжении нескольких десятков лет [18, 29]. Основным направлением этого курса служила трактовка положений кибернетики на инженерном уровне, предназначенном для подготовки инженеров-кибернетиков широкого профиля.

Структура кибернетики в нашем понимании, в соответствии с которой излагается курс «Основы кибернетики», может быть разбита на четыре раздела:

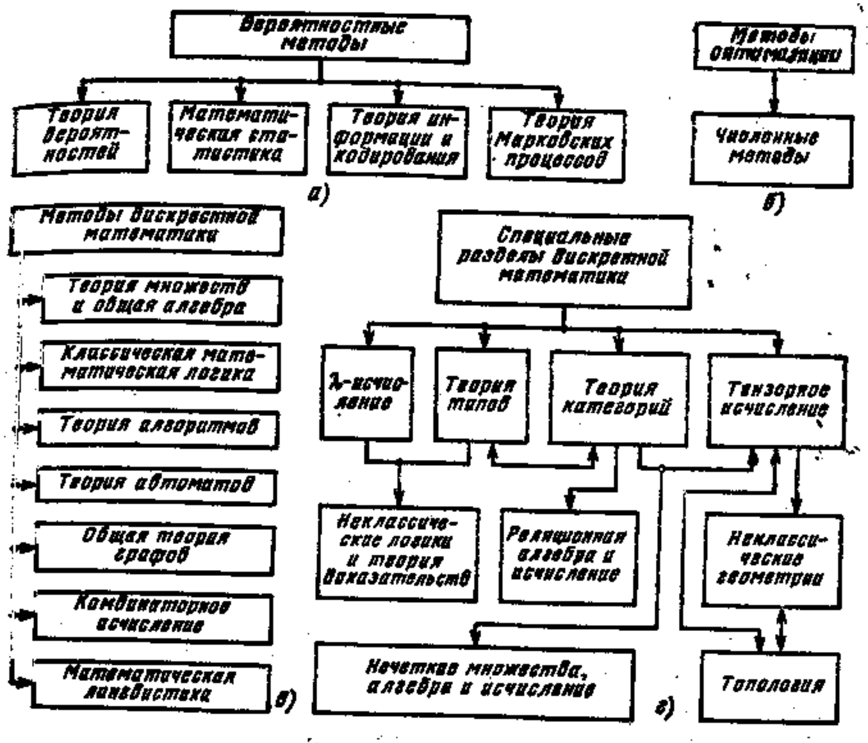


Рис. В.11. Структура математических основ кибернетики

1. Общие сведения о кибернетике.
2. Математические основы кибернетики.
3. Основы кибернетических моделей.
4. Специальные и прикладные вопросы кибернетики.

Математические основы кибернетики (рис. В.11) разделены на четыре цикла: вероятностные методы (рис. В.11, *а*), методы оптимизации и численные (рис. В.11, *б*), методы дискретной математики (рис. В.11, *в*), специальные разделы дискретной математики (рис. В.11, *г*). Первые три раздела вероятностных методов посвящаются теории вероятностей, математической статистике и теории марковских процессов. В разделе математической статистики в основном излагаются вопросы определения числовых значений характеристик случайных величин и законов распределения. Уже на примере этого раздела можно убедиться в том, что различные разделы кибернетики взаимно проникают друг в друга, и резкого разделения провести не удается. Часто изложение теории вероятности производится с привлечением общей алгебры и математической логики, которые рассматриваются в разделе дискретной математики. Изложение математической статистики заканчивается теорией статистических решений, предлагающей математические методы для принятия решений в условиях неопределенности. Эти методы в значительной степени перекрываются методами теории игр и теории искусственного интеллекта, где для принятия решений применяются принципиально другие методы.

Четвертый раздел вероятностных методов посвящен теории информации и кодирования.

В значительной степени этот вопрос изложен в первом издании «Основ кибернетики» (т. 1, «Математические основы кибернетики»). Однако со времени выхода первого издания т. 1 сильно развилось направление теории информации, связанное с человеко-машинными методами представления и переработки информации. В настоящем издании этот вопрос частично освещен в т. 2, ч. 3 «Модели банков данных», «поближе» к вопросу об искусственном интеллекте, с которым он тесно связан. Первый том содержит новый раздел «Категориальная модель семантической квантовой информации по В. Г. Толстову».

Оптимальные методы обработки, передачи, преобразования и защиты информации существенным образом зависят от способов ее кодирования, которым уделялось большое место в т. 1 первого издания. Для построения оптимальных систем управления необходимо иметь в наличии математический аппарат для отыскания оптимальных законов управления, который составляет основное содержание второго цикла математических основ кибернетики. В зависимости от специфики системы управления могут применяться различные методы оптимизации — от классических методов Эйлера—Лагранжа, динамического программирования и принципа максимума Понтрягина, методов математического программирования до человеко-машинных диалоговых методов и методов искусственного интеллекта, при этом рассматриваются непрерывные и дискретные, детерминированные и вероятностные варианты этих методов.

Особого внимания заслуживает третий цикл математических основ кибернетики — дискретная математика. Как уже указывалось, большинство процессов управления, особенно в АСУ, дискретные. Массивы информации и программы, записанные на машинных носителях, дискретны по своей структуре. В редких случаях их можно описать с помощью аппарата непрерывной математики (дифференциальных уравнений). Поэтому для инженера-специалиста по управлению требуется другая математическая подготовка по сравнению с той, которая дается в настоящее время в технических вузах. Ему необходимо знать дискретную математику, которая называется так потому, что в ней нет понятия непрерывности, дифференцируемости. Дискретная математика включает следующие разделы: теория множеств и общая алгебра, математическая логика, теория алгоритмов, теория автоматов, теория графов, комбинаторное исчисление, математическая лингвистика.

Эти разделы составили так называемую классическую дискретную математику, которая в современном смысле стала внедряться в инженерные круги в нашей стране в 60-е годы. Однако за два десятилетия развития кибернетики, особенно разделов банков данных и систем искусственного интеллекта, для инженера потребовались знания дополнительных разделов дискретной математики, объединенных названием «специальные разделы дискретной математики». Этот раздел включает теорию категорий, теорию типов, λ-исчисление, алгебру и исчисление нечетких множеств, реляционную алгебру и исчисление, неклассические логики, тензорное исчисление, топологию и неклассические геометрии. Многие из этих разделов кратко представлены в первом издании т. 2 «Основ кибернетики» [30].

Естественно, что в зависимости от специфики учебного процесса отдельные разделы могут излагаться или в дискретной математике, или в специальных разделах дискретной математики. В частности, для кибернетика, инженера знаний требуется специфическое изложение теории множеств и общей алгебры. Поэтому эти два раздела размещаются в ч. 2 под названием «Основополагающие разделы» и включены в курс «Специальные разделы дискретной математики». Аналогичным образом комбинаторное исчисление в смысле Карри дано в ч. 4 «Алгебры и исчисления» в гл. 11 под названием «Комбинаторная логика, алгебра и исчисление».

В ч. 3 дискретная математика (классическая) в разделе математической логики излагается только классическая логика (логика высказываний и предикатов), а неклассические логики вместе с теорией доказательств (включая логику предикатов выше 1-го порядка) рассматриваются в ч. 4 «Алгебры и исчисления».

Более детально структура специальных разделов дискретной математики рассматривается в гл. 1 настоящего тома.

Все кибернетические модели, рассматриваемые в третьей структурной части кибернетики — основах кибернетических моделей (рис. В.12), разделены на три группы. Первую группу составляют модели, в основе которых лежит вероятностная природа. Это модели теории массового обслуживания, теории игр, распознавания образов. Вторая группа объединяет кибернетические модели, поведение которых описывается дифференциальными или разностными уравнениями. Большинство методов исследования таких систем излагается в работах по системам автоматического регулирования. Для этих методов

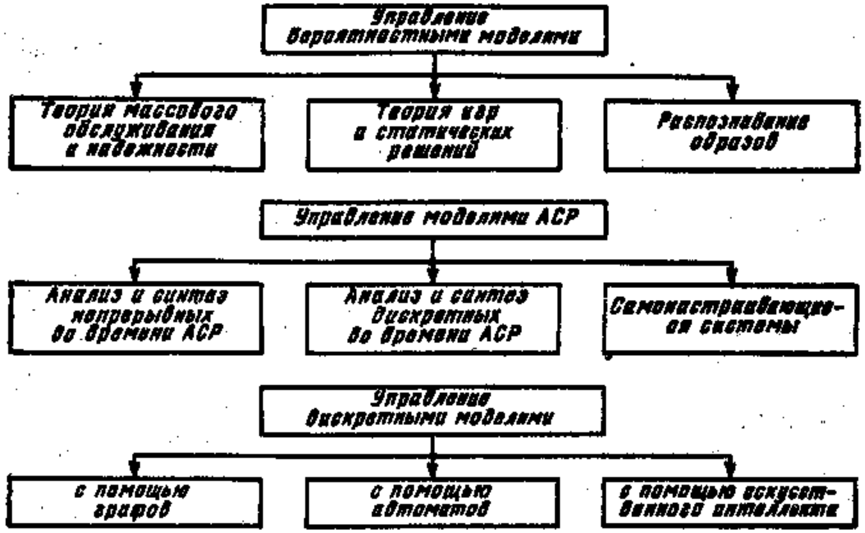


Рис. В.12. Классификация кибернетических моделей

характерно рассмотрение процессов во времени, поэтому такие модели могут быть названы динамическими системами. Третью группу кибернетических моделей составляют дискретные кибернетические модели. Эти модели применяются и для исследования процессов управления, протекающих во времени, но в основном в них время не используется. Например, требуется с помощью вычислительных машин раскроить листовое железо для обшивки корабля наилучшим образом с точки зрения расхода материала, причем время, в течение которого производится раскрой, не имеет значения. Здесь с успехом применяются как детерминированные, так и вероятностные методы расчета. Рассматриваемые в этой группе методы наиболее слабо освещены в литературе и представляют собой наибольшую ценность. В этой же группе имеется раздел «Управление с помощью искусственного интеллекта», который мог быть назван иначе — «Интеллектуальное управление», или «Интеллектуальные системы управления», или «Системы искусственного интеллекта». Он включен вместо раздела «Лингвистическое управление» в т. 1 первого издания и является логическим его развитием, так как лингвистические модели как составная часть вошли в модели искусственного интеллекта.

В общей структурной схеме кибернетики предусматривается еще четвертая часть, посвященная специальным и прикладным вопросам кибернетики (рис. В.13). Содержание этой части выглядит наиболее неопределенным, так как ее название

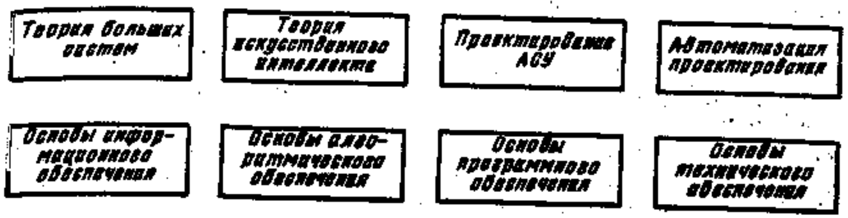


Рис. В.13. Специальные и прикладные вопросы кибернетики

допускает включение самых разнообразных разделов. Вопросы, связанные с проектированием АСУ, выделены в самостоятельный раздел. Хотя они и представляют модификацию больших систем, для них характерен человеко-машинный способ управления. В этом разделе в основном рассматриваются общие вопросы проектирования АСУ и особенности проектирования функциональных подсистем.

Раздел, названный «Теория искусственного интеллекта», включает различные аспекты теории принятия решений в больших системах, а также вопросы создания информационно-программных комплексов, моделирующих профессиональный искусственный интеллект. Необходимость введения самостоятельного раздела по автоматизации проектирования диктуется тем обстоятельством, что в ближайшие годы любое проектирование, в том числе АСУ и ЭВМ, немыслимо без автоматизации. Кроме того, как уже указывалось, в современных кибернетических системах совмещаются процессы управления и проектирования.

Вторая группа разделов этой части посвящена так называемым обеспечивающим подсистемам систем управления. Хотя концепция обеспечивающих подсистем возникла в процессе разработки АСУ, она полностью применима к любым системам управления. Для любого управления нужны данные об объекте управления, которые могут обновляться в процессе управления, а также различные модели преобразования информации в системе, что и составляет основное содержание информационного обеспечения. Начало проектирования начинается с системного обследования и составления технического задания, после чего составляются математические модели управления (кибернетические модели). Если модель использует дифференциальные уравнения, то необходимо, применяя численные методы, их «арифметизировать», чтобы получить процедуру вычисления решений для конкретных числовых начальных условий, использующую четыре арифметических действия и другие элементарные машинные операции. Эта процедура и рассматривается как алгоритм решения задачи планирования с помощью, например, линейного программирования; в качестве математической модели выступает модель задачи оптимизации в виде линейного программирования, а алгоритм представляет конкретную процедуру отыскания решения этой задачи. При наличии алгоритма решение в принципе может быть получено ручным способом, однако в большинстве случаев требуются вычислительные машины, и необходимо по алгоритмам составить программы вычислений на ЭВМ, которые являются частью программного обеспечения системы. Три раздела: математические модели, алгоритмическое и программное обеспечение — составляют содержание математического обеспечения.

Основным техническим средством является ЭВМ. При проектировании технического обеспечения в этой части поступают двояким образом: или разрабатывают специализированную управляющую машину, как часто бывает при управлении непрерывными объектами, или останавливаются на определенной системе ЭВМ, допускающей модульный синтез конкретной ЭВМ, в частности, в виде сети микропроцессоров. В последнем случае из машинно-ориентированного программного обеспечения (операционной системы и т. д.), широко используя режимы генерации, создают программное обеспечение (главным образом, операционную систему), ориентированное на конкретную систему, для которой создается АСУ. Помимо ЭВМ при проектировании технического обеспечения требуется определить состав АСУ при наличии модульного исполнения или спроектировать терминальные устройства и линии связи. При разработке программного обеспечения кибернетик должен знать с позиций пользователя имеющиеся в его распоряжении машинно-ориентированные системы программирования, уметь создавать из них «дочерние» проблемно-ориентированные системы программирования и дорабатывать их в процессе эволюции системы управления, что особенно важно для АСУ.

В связи с этим можно было бы добавить еще пятую структурную часть, посвященную методам программирования на ЭВМ, начиная от методов программирования в кодах машины, списковых структурах и кончая алгоритмическими языками и метаязыками. Хотя в строгом смысле этот раздел носит самостоятельный характер и не включается в кибернетику, большинство деловых методов кибернетики так тесно переплетается с вопросами программирования и интерпретации на ЭВМ, что очень трудно бывает разделить, где «чистое программирование», а где «чистая кибернетика». Знание методов программирования для инженера-кибернетика обязательно, однако по установившейся традиции в высшей школе они изучаются независимо и вне кибернетики.

Глава 1

ОБЩИЕ СВЕДЕНИЯ О МАТЕМАТИЧЕСКИХ ОСНОВАХ КИБЕРНЕТИКИ

1.1. АКТУАЛЬНОСТЬ СОЗДАНИЯ КУРСА «СПЕЦИАЛЬНЫЕ РАЗДЕЛЫ ДИСКРЕТНОЙ МАТЕМАТИКИ» И ОБЩИЕ ТРЕБОВАНИЯ К НЕМУ

Как сказано выше (см. § В.5), математические основы кибернетики состоят из трех частей (рис. 1.1): вероятностных основ кибернетики, дискретной математики и специальных разделов дискретной математики. Данная книга посвящена в основном специальным разделам дискретной математики. Комплектация или формирование этого раздела для автора происходило следующим образом. Инженерная практика (точнее, инженерно-математическая или практика специалиста по прикладной математике) в области АСУ, САПР, банков данных, систем искусственного интеллекта более или менее определила интерес специалиста-кибернетика к различным, как правило, достаточно молодым разделам математики. Так, «индуктивным путем» стал постепенно складываться курс «Специальные разделы дискретной математики» (другое название — «Новая математика»), структура которого, представленная на рис. 1.2, в значительной степени отражает содержание данного тома [18, 22, 38]. Вначале постараемся изложить индуктивный путь становления этой части математических основ кибернетики. Заметим, что искусственный интеллект, по нашему мнению, в последнее время стал основным, ведущим, определяющим разделом кибернетики, без которого ни одна кибернетическая система не может обойтись. Если еще раз учесть, что путь к искусственному интеллекту лежит через банки информации как хранилища знаний, то можно утверждать, что специальные разделы дискретной математики предназначены в основном для обслуживания этих двух частей (банков информации и искусственного интеллекта). Еще раз подчеркнем необходимость в дополнительном математическом образовании инженера-кибернетика, инженера знаний и т. д. и необходимость создания курса дополнительных разделов математики (специальных разделов дискретной математики). Дело в том, что курс «Дискретной математики», созданный за последние 20 лет, уже не удовлетворяет инженера-кибернетика (или инженера-математика), особенно если он занимается банками информации и искусственным интеллектом. Остановимся на трех аспектах или причинах возникновения новой математики.

Рис. 1.1. Укрупненный состав математических основ кибернетики

С одной стороны, работы по перспективному направлению «искусственный интеллект» последних лет убедили специалистов, что без создания специальных технических средств, сокращенно именуемых интеллектуальными компьютерами (ИК), невозможно решать удовлетворительно проблему создания коммерческих (для продажи на рынке) интеллектуальных систем (ИС). Наиболее рельефно этот тезис сформулирован в японском проекте вычислительных систем пятого поколения (ВС-5) и повторен в аналогичных проектах в США, Англии и странах Европейского экономического сообщества (ЕЭС).

С другой стороны, разработчики компьютеров под воздействием потребностей рынка пришли к выводу, что пользователю нужны не столько мощные вычислительные машины универсального типа, сколько интеллектуальные компьютеры, которые, обладая интеллектуальными способностями, были бы дружелюбны и помогали бы в решении интеллектуальных задач.

Таким образом, образовался разрыв между разработчиками вычислительных машин и специалистами по интеллектуальным системам. Первые потребовали объяснения непонятного для них объяснения термина «интеллектуальный компьютер».

Кроме того следует заметить, что определенный теоретический уровень работ в области ИС не соответствует уровню кибернетических концепций и моделей (фреймы, семантические сети, механизмы сопоставления, унификации, формирования решений и т. д.), которые нуждаются в математической поддержке, описании.

И наконец, необходимо определить состав и содержание дополнительной математической подготовки инженеров знаний, которых по прогнозу в 1995 г. в США потребуется около 500 тысяч [24].

Рассмотрим кратко особенности интеллектуальных систем, которые должны описывать математический аппарат, условно названный «Математические основы интеллектуальных систем» или «Специальные разделы дискретной математики».

Математическое обеспечение интеллектуальных систем должно содержать средства описания следующих свойств.

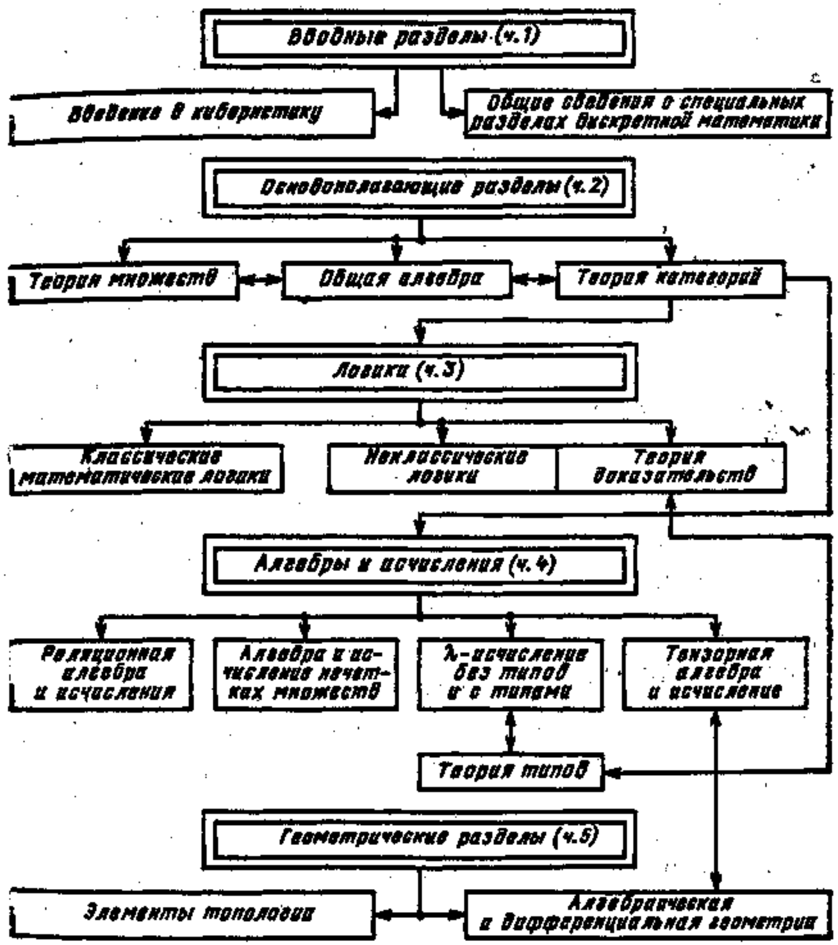


Рис. 1.2. Общая структура специальных разделов дискретной математики

1. *Дуализм* процедурально-декларативного представления знаний, который заключается в особенностях представления одних и тех же знаний «динамически» в виде процедур (процессов, программ) или статически в виде данных (записей, файлов).

2. *Противоречивость* принятия решений, которая заключается, в частности, в том, что в зависимости от ситуации одна и та же «посылка», одно и то же условие дают различные следствия, иногда даже противоположные друг другу, что исключает вероятностные методы, где имеет место усреднение. В ка-

честве других вариантов модели принятия решений можно указать модальные и многозначные логики, когда одно и то же условие приводит к разным следствиям (логики, в которых не соблюдается закон исключенного третьего), и, наконец, модель немонотонной логики, когда разрывается последовательная цепочка рассуждений, имеется возврат к предыдущему (в частности, Backtracking).

3. *Отсутствие оптимума* и отказ от его поиска в больших системах, к которым относятся интеллектуальные системы благодаря наличию большого объема информации в базе данных (БД) и базе знаний (БЗ). Вместо этого вводится поиск пригодного допустимого и наилучшего из подходящих решений. В результате — практически отсутствие методов и алгоритмов оптимизации в теории искусственого интеллекта.

4. *Нечеткость* (размытость) процесса принятия решений, связанная с нечеткостью смысла, семантики естественного языка, что способствовало появлению алгебры и исчисления нечетких множеств, нечетких алгоритмов принятия решений.

5. *Человеко-машинный аспект* принятия решений, связанный с п. 3 и 4, который способствовал появлению различных логик типа вопросно-ответной и др.

1.2. СТРУКТУРА КУРСА «СПЕЦИАЛЬНЫЕ РАЗДЕЛЫ ДИСКРЕТНОЙ МАТЕМАТИКИ»

На рис. 1.2 приведена общая структура специальных разделов дискретной математики. В целом структура содержит пять частей: вводную, основополагающую; логики; алгебры и исчисления; геометрическую. Она отражает информационный (декларативный) подход к кибернетике и математическим основам интеллектуальных систем (МО ИС), т. е. соответствует тезису «путь к искусственному интеллекту лежит через банки данных (и знаний)!». Она является открытой системой, в которую можно добавить другие разделы дискретной математики. В частности, в этой схеме плохо обеспечивается процедуральный подход к интеллектуальным системам, который требует включения разделов, связанных с нормальными алгоритмами Маркова, теоремами Геделя, теорией чисел и т. д.

Остановимся на особенностях отдельных разделов, которые в соответствии с требованиями к МО кибернетики и ИС излагаются не совсем традиционным стилем. Основу МО кибернетики и ИС составляют теория множеств, общая алгебра и теория категорий. В теории множеств особое внимание уделяется разделу частично упорядоченных множеств, так как он непосредственно обслуживает иерархические структуры, ISA — иерархию баз знаний и баз данных в ИС. Общая алгебра, в которой в основном рассматриваются теория групп, полугрупп, колец и т. д. и изоморфизмы, обслуживает модели баз знаний и данных и механизмы интерпретации и усвоения знаний, а также механизмы формирования решений. Кроме того, этот раздел обеспечивает необходимыми знаниями для понимания последующих разделов и прежде всего IV цикла — геометрического (главным образом, алгебраической топологии и алгебраической геометрии).

Новым и дискуссионным разделом является теория категорий — сравнительно молодой раздел дискретной математики. Ее методы стали популярными и эффективными для описания объектов кибернетики, которые обладают свойствами декларативности и процедуральности фреймов, реляционных (и любых) баз данных. Теория категорий с успехом применяется для описания программного продукта, в частности семантики программ (см. рекурсивное программирование). Декартово-замкнутая категория (ДЗК) — определенный раздел теории категорий — совпадает с λ-исчислением с типами (типовым λ-исчислением). Можно вообще показать, что ДЗК является некоторой системой типов.

Среди всех претендентов на построение оснований математики (λ-исчисление, логика, теория доказательств) теория категорий, по-видимому, обладает большими возможностями. Теория категорий, так же как и λ-исчисление, обладает характерным для так называемых аппликативных систем свойством задавать вычислительное предписание (к примеру, в определении функции). Для ИС особенно важно то, что раздел теории категорий, называемых топосами, эффективно описывает неклассические, модальные логики, в частности интуиционистскую логику Крипке [38, 39].

Теория типов, которые используются в λ-исчислении, в результате чего существенно расширяются его возможности, в теории доказательств, в исчислении предикатов порядка выше первого и т. д., заслуживает самостоятельного, отдельного рассмотрения. В какой-то мере этот раздел пересекается с моделью абстрактных типов данных в теории баз данных.

Часть 3, посвященная алгебрам и исчислениям, может быть названа как логико-алгебраическая. Она содержит алгебру и исчисления нечетких множеств — раздел, который обеспечивает математическое описание (моделирование) свойства нечеткости, неопределенности и т. д.

Возможны два стиля изложения специальных разделов дискретной математики: индуктивный и дедуктивный. Во втором случае последовательно излагают части с первой по пятую. При индуктивном изложении, по собственному опыту автора, целесообразно начинать «раскрутку» всех разделов с реляционной алгебры и исчисления, послужившим основой для теории реляционных, иерархических и сетевых банков (баз) данных. Этот путь наиболее приемлем для инженера-математика с инженерной (реляционной) точки зрения. В частности, на базе реляционной алгебры и исчисления Руссопулосом и Вольфенгагеном построена теория реляционной базы знаний фреймового типа, в которой «пользователя» подводят к теории типов, как это сделано в системе CSDL [36, 37].

Одной из разновидностей категорий являются тензоры. Определенные по Крону как специальные геометрические объекты, они обладают эффективными средствами описания семантики информации, т. е. базы знаний. Тензорные инварианты, соединенные процедуральности и декларативности, универсальность описания — эти и другие особенности тензорных методов привлекают внимание инженеров-математиков в течение многих лет.

В 50-х годах японцы (эти возмутители спокойствия) создали с помощью научно-технической группы Кондо теоретические основы единого образования инженера-математика (точнее, специалиста по прикладной математике) на базе тензорной алгебры и исчисления. Кроме того, с методологической точки зрения тензоры устанавливают мост между аналитическими и геометрическими методами математики.

Последний цикл — геометрический — предназначен для ликвидации геометрической безграмотности инженера. Первостепенное значение здесь отводится алгебраической геометрии евклидова, аффинного и проективного пространств. Этот раздел позволяет с помощью ЭВМ проектировать новые системы и строить ИС с использованием четырех- и n-мерных пространств, особенностей аффинного и проективного пространств, что недоступно «ручным» методам без компьютера.

1.3. ЭВОЛЮЦИОННЫЙ АСПЕКТ КУРСА

По мнению автора, интенсивное распространение математических методов началось с реляционных банков данных (РБД) в начале 70-х годов, когда Кодд построил реляционную модель данных на базе реляционной алгебры и исчисления [36]. Этот метод, обладающий минимальной математической строгостью, оказался чрезвычайно эффективным на практике, породил новый тип банков данных — реляционный, к которому сводятся любые банки данных, в частности иерархические и сетевые. Реляционный банк данных явился основой для построения интеллектуальных систем. Это связано с японским проектом вычислительных систем пятого поколения (ВС-5) и новой информационной технологией (НИТ) для создания систем электронной обработки данных [27, 28]. Для поддержания РБД потребовалась база знаний в виде сетей фреймов предикатного типа, которая развита в работах Руссопулоса [37] и известна как система CSDL (рис. 1.3) и которая на первый взгляд имеет скромный вид, но, так же как работы Кодда, она сыграла решающую роль в проникновении новых математических методов в кибернетику. Модель единицы (фрейма) семантической информации (в виде предикат-действия и множеств типов переменных на висящих концах дуг-ролевых падежей) обеспечивает успешное математическое описание с помощью теории типов, категорий и т. д. Уже здесь появляется тяготение к модели вычислительного предписания, к означиванию фрейма, связанному с приданием конкретных значений переменным. Вычислительное предписание широко используется в исчислении предикатов и λ-исчислении, на котором основаны язык ЛИСП и процедура-функция ЛЯМБДА во многих других языках программирования. С другой стороны, в современных системах программирования, в теории вычислений и вычислимости резко повысился интерес к модели вычислений вместо записи функций как связи (отображения, отношения) двух переменных из разных множеств (так называемый аппликативный аспект). Это вызвало интерес и к λ-исчислению, и к теории λ-процедур, и к λ-конверсии без типов и с типами [37, 51]. При этом внутренняя противоречивость математических теорий оказалась не вредной, а полезной для описания современных кибернетических систем, особенно систем искусственного интеллекта, которые при разных условиях дают разные следствия при одной и той же предпосылке. Это касается λ-исчисления, неклассических логик, теории категорий, которые успешно применяются в теории рекурсивного программирования для решения проблемы семантики программы. Популярными у инженеров-математиков стали средства модальной логики, алгебра и исчисления нечетких или размытых множеств, которые оперируют нечеткими понятиями типа «много», «немного», «немного больше» и т. д. и успешно описываются с помощью теории категорий. Неклассические

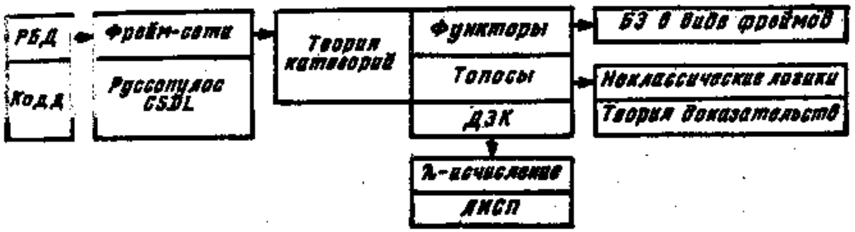


Рис. 1.3. Схема индуктивного формирования специальных разделов дискретной математики

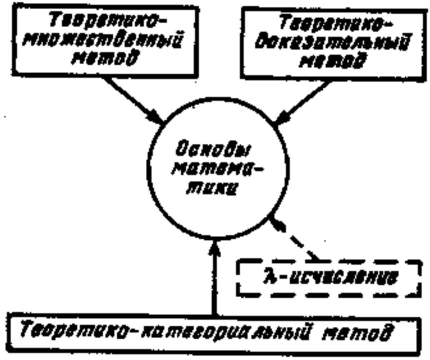


Рис. 1.4. Различные методы создания единых основ математики

логики используются для создания механизмов вывода и формирования решений в интеллектуальных системах и успешного поиска информации банков данных и знаний. Большая часть в этом разделе предоставляется методам теории доказательств, которая используется для определения правильного поведения кибернетических систем, правильной истинной реакции системы, в том числе системы искусственного интеллекта. В дополнение или вместо оптимального поведения и реакции выступает правильное (истинное с позиций теории доказательств) поведение (реакция), которое не противоречит ограничениям, называемым фильтрами. Так или иначе теоретико-доказательный подход явился вторым подходом к созданию основ математики наряду с теоретико-множественным, предложенным Бурбаки (рис. 1.4). Третьим подходом явилась теория категорий, которая, взяв свое начало в теории множеств (так же как теория групп, дальнейшим развитием которой она является), переживает бурное развитие. В качестве самостоятельного подхода к проблеме основания может котироваться λ-исчисление с типами и без типов, точнее, теория аппликативных систем (рис. 1.5), в которую наряду с λ-исчислением входят комбинаторная алгебра и логика. Определение функции в аппликативной системе вычислений как вычислительного предписания проще всего пояснить на примере.

Пример 1.1. Рассмотрим систему $\left.\begin{array}{l} f_1 = (x+2)(x-2); \\ f_2 = x^2 - 4 \end{array}\right\}$. На множестве Q оба этих вычислительных предписания дают одни и те же значения: f_1 в применении, приложении (аппликация f) к $x \in Q$ всегда дает те же значения, что и f_2 в применении к x. Пишем $f * x$ вместо «f в применении к x» или аппликация f к x. Таким образом, $f_1 \neq f_2$, хотя $f_1 * x = f_2 * x$ для произвольного x. Два этих предписания отличаются не только внешним видом, но и сложностью: f_1 требует больше шагов вычислений, чем f_2 (три против двух). Поэтому в классическом реляционном определении функции как отношения двух множеств (подмножеств), из которых одно для аргумента, другое для функции, она представляется как единое целое во всем своем протяжении или экстенсионале (экстенсиональное задание функции, функция как экстенсионал).

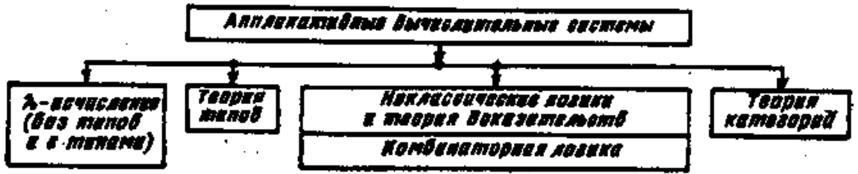


Рис. 1.5. Состав аппликативных вычислительных систем

В соответствии со вторым определением вводится в рассмотрение понятие алгоритмики [39] как теории оперирования с вычислительными предписаниями или теории аппликативных систем [38, 39]. Аксиоматическое обоснование алгоритмики или теории аппликативных систем напоминает аксиоматическую теорию групп. В математике часто встречаются разнообразные предметные области, представляющие собой группы. Аналогичным образом встречаются замкнутые в себе совокупности вычислительных предписаний, такие, как геометрические построения, разнообразные языки программирования, предписания о формальном оперировании с алгебраическими и логическими выражениями и т. д. Теория групп работает в основном с операцией умножения элементов группы, а в аксиоматической алгоритмике основной операцией является применение функции к аргументу, или аппликация $f * a$. Вместо экстенсионала функции, который обеспечивает совокупность значений или определяет область истинности понятия, при аппликативном подходе на первое место выдвигается интенсионал, который определяется совокупностью вычислительных операций, составляющих вычислительное предписание или «как вычисляется». Концепция аппликативного подхода в математике рассматривается в главах, посвященных λ-исчислению, неклассическими логиками, в частности комбинаторной логикой, и в какой-то мере теорией категорий (рис. 1.5). Кроме того, в нашем варианте специальных разделов дискретной математики с разных позиций будет выступать тройка: алгебра, исчисление и логика (рис. 1.6). Алгебра изучает действия над элементами множества, исчисление — операции получения следствия из посылки, причем в отличие от алгебры вводятся кванторы. Логика занимается тем же самым, только она, как правило, испытывает на себе влияние философии (философская и математическая логика). Поэтому нередко будут встречаться такие тезисы: «изложение логики первого порядка с помощью исчисления секвенций (или исчисления предикатов первого порядка)», «изложение логики второго и более высокого порядка с помощью теории типов» и т. д. В связи с тем что кибернетиков заинтересовали противоречивые системы и логики, помимо аристотелевской философской логики их внимание привлекли диалектические

Рис. 1.6. Пояснение к логико-алгебраической концепции

логика Гегеля и логика Канта (рис. 1.7). Важно отметить геометризацию практически всех частей специальных разделов дискретной математики. Прежде всего, геометрия как математическая наука содержит дополнительные средства описания, анализа и синтеза, включая способности естественного интеллекта типа геометрического воображения, мышления и т. д., которые в других математических методах отсутствуют. Переходным разделом от негеометрических методов к геометрическим служат тензорная алгебра и исчисление. Неоднократно делались попытки положить в основу образования инженера-математика (или специалиста по прикладной математике) тензорное исчисление. Наиболее значительный вклад сделан группой специалистов Японии в 1960—1970 гг. в электротехнической лаборатории Токийского университета под руководством проф. Кондо [22]. Кибернетиков привлекают тензоры как универсальные аппараты для описания реальных объектов, процессов их функционирования и проектирования. Имеется возможность описывать семантику кибернетических систем с помощью тензоров, которые рассматриваются не только как математические, но и как физико-геометрические объекты. По Крону тензоры и матрицы аналогичны скульптуре и ее плоской фотографии соответственно. Кроме того, тензоры — это преобразования, свойственные объекту, системе, благодаря чему они выступают как инварианты, описывающие объект, систему или процесс проектирования в САПР. Снова здесь наблюдается, прежде всего, повышенное внимание к процессу, динамике (а не к статическому описанию), так же как в аппликативных системах. Наконец, тензоры являются мощным средством для математических характеристик аффинных пространств, в том числе понятия аффинной связности пространств, в которых может отсутствовать метрика, понятие длины. Тензоры выступают как путеводители к геометрическому разделу книги.

Популярный в последние годы алгебраический подход в геометрии позволяет распространить на геометрию все последние достижения алгебраических разделов дискретной математики, подключая к ним геометрические мышление, воображение. Любой алгебраический метод находит себе применение,

Рис. 1.7. Философские и математические логики

интерпретацию в геометрии и называется геометрическими логиками и исчислениями. Кроме того, современные компьютеры тесно связаны с геометрическими моделями (так появились топологические компьютеры) на однородных вычислительных средах, сотоподобные вычислительные среды и т. п. Графические устройства ввода-вывода типа графопостроителей, графических дисплеев требуют от специалиста знания геометрии, а также языка программирования, рассчитанного на обработку не списков, не символов, а геометрических фигур, образов объектов и ситуаций. Но наибольший интерес представляет манипулирование одними и теми же предметами, фигурами, конструкциями в разных пространствах и системах координат. То, чего нельзя сделать в одной системе координат, к примеру в трехмерной, можно сделать в четырехмерном пространстве, преобразовав исследуемую проектируемую систему из трехмерного пространства в четырехмерное. Таким образом можно «снять рубашку, не снимая пиджака». Хотя реально не существует четырехмерного пространства, но на компьютере можно путем машинного моделирования реализовать пространства размерности более трех и даже исследовать безразмерные (аффинные пространства), которые характеризуются такими параметрами, как аффинная связанность, внутренняя кривизна. Как перейти от трехмерного пространства в четырех- и пятимерные? Как преобразуются при этом объекты и предметы, например параллелепипеды, шары, сферы и т. д.? Какое логико-алгебраическое выражение описывает преобразование параллелепипеда в шар в трехмерном пространстве? Чему соответствует в трехмерном пространстве четырехмерная плоскость? Решение подобного рода задач повышает эффективность проектирования и создает средства исследования, превышающие возможности естественного интеллекта, так как человек может мыслить только в метрическом пространстве измерения не более трехмерного. В связи с этим появилось понятие машинной геометрии. В последней главе данной книги рассмотрены с алгебраических позиций и с помощью методов дифференциальной геометрии аффинное, проективное и евклидово пространства, их связь и преобразования из одного в другое (рис. 1.8).

Рис. 1.8. Состав геометрических разделов

При таком индуктивном методе построения специальных разделов дискретной математики оказалось, что различные разделы объединяют общие свойства, которые будут использованы при построении дедуктивным способом некоторых математических основ кибернетики. Кратко они сводятся к следующему:

логико-алгебраический характер разделов;

аппликативность;

внутренняя противоречивость разделов дискретной математики (неклассических логик, неклассических геометрий);

динамичность методов (вычислимость, алгоритмика, тензоры);

геометризация методов за счет введения алгебраической геометрии.

1.4. СПЕЦИАЛЬНЫЕ РАЗДЕЛЫ ДИСКРЕТНОЙ МАТЕМАТИКИ И МАТЕМАТИЧЕСКИЕ АБСТРАКТНЫЕ МАШИНЫ

Взаимосвязь различных частей специальных разделов дискретной математики лучше всего продемонстрировать с помощью так называемых абстрактных машин, при этом инженер-математик, сразу «приземляя» определенный раздел, думает о его реализации с помощью вычислительных машин. Термины «машина», «механизм», «процессор» в соответствии с определением акад. А. А. Андронова используются для обозначения любого преобразователя (кибернетической системы по переработке информации), включая систему автоматического регулирования, следящие системы, системы искусственного интеллекта. На рис. 1.9 представлена условная категориальная абстрактная машина (КАМ), или машина-ядро. В ней выделены четыре памяти: категорий, функторов, ДЗК, топосов, три процессора-механизма: процессор получения функторов для категорий, процессор ДЗК из категорий, процессор получения подобъектов из категорий, и три выхода: функторы, ДЗК, топосы [40, 41].

При очень смелом подходе можно предлагать инженерам-конструкторам реализовать схему КАМ по аналогии с машиной Дж. фон Неймана, основанной на автоматах и булевой алгебре. Но для этого требуется, чтобы эти инженеры знали теорию категорий хотя бы в объеме данной книги (см. гл. 4). В настоящее время в связи с японским проектом машин пятого поколения (ВС-5) можно предложить направления по созданию компьютеров нового типа, в том числе логических компьютеров на базе секвенциального исчисления, где вместо триггера, с помощью которого в машине фон Неймана реализуются булевы алгебра и логика, используется логический секвенциальный элемент для секвенциального исчисления, который эквивалентен 300—400 триггерам. Следующие абстракт-

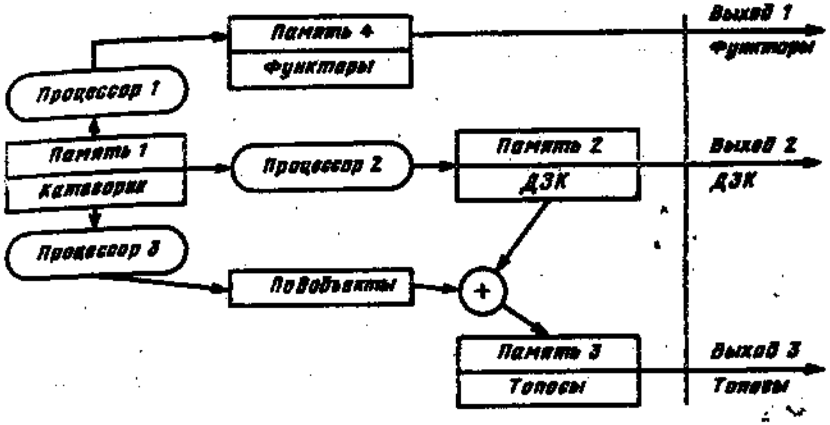


Рис. 1.9. Категориальная абстрактная машина (КАМ) (машина-ядро)

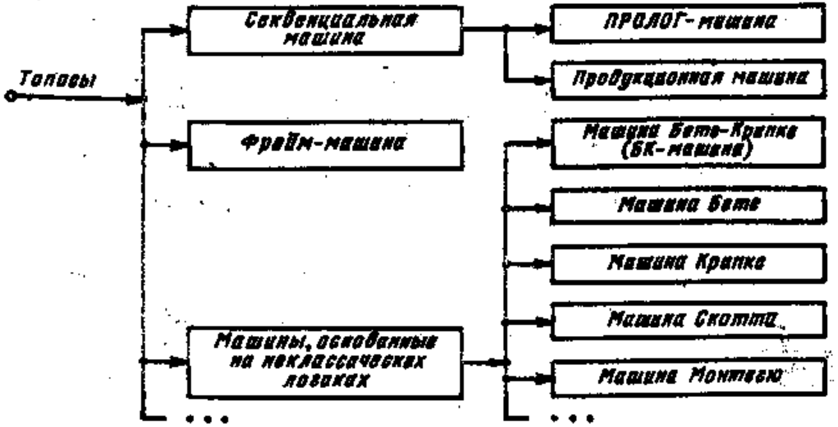


Рис. 1.10. Топос-машины (1-й вариант)

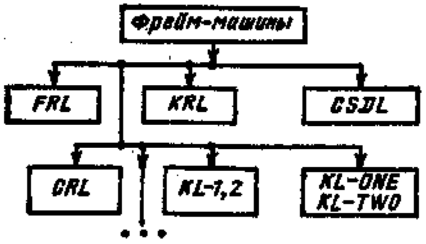


Рис. 1.11. Семейство фрейм-машин

ные машины — это топос-машины (рис. 1.10). Входом машин этой серии служат топосы, как выходом *3* в машине КАМ (рис. 1.9.). В этом семействе имеется первый уровень — абстрактные машины типа секвенциальной машины, машины неклассических логик (Н/К-машина) и второй уровень — машины конкретных Н/К-логик Бете, Крипке, Бете—Крипке, ПРОЛОГ-машин и т. д. На рис. 1.11—1.13 представлены семейства фрейм-машин, ПРОЛОГ-машин и продукционных машин соответственно, которые могут рассматриваться уже как конкретные машины, продаваемые на рынке. Они, как правило, различаются по названию языка представления знаний (ЯПЗ): KRL, FRL и т. д. Некоторые машины встречаются в разных семействах, что указывает на смешанный, гибридный вариант машины, в которой используются несколько ЯПЗ. На рис. 1.14 представлены семейства ДЗК- и функтор-машин. Первое семейство разделяется на семейство ЛЯМБДА-машин (от λ-исчисления), второе распространяется на фрейм-машины, имеющие второй источник — предок ЛИСП-машины.

В заключение этой главы приведем семейство абстрактных геометрических математических машин. Специалисты по машинной реализации искусственного интеллекта «пристегнули» к проекту топологическую геометрическую машину, имеющую кристаллическую, сотоподобную структуру, компьютер FAIM-1, на котором реализуется язык представления знаний OIL. Алгебраическая геометрия аффинного, проективного и евклидова пространств описывает свойства и закономерности этих пространств с помощью алгебраических символов. Но машины-компьютеры символьной обработки давно существуют (ЛИСП-машина). Поэтому вполне разумно обсуждать (рис. 1.15) аффинную геометрическую абстрактную машину (АГАМ), проективную абстрактную геометрическую машину (ПАГМ) и евклидову геометрическую абстрактную машину (ЕГАМ).

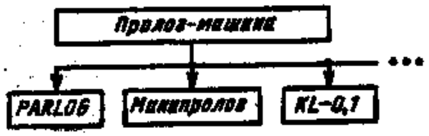


Рис. 1.12. Семейство ПРОЛОГ-машин

Рис. 1.13. Семейство продукционных машин

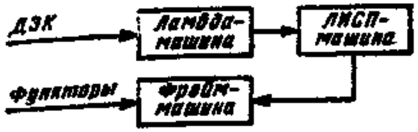


Рис. 1.14. Семейство ДЗК-машин и функтор-машин

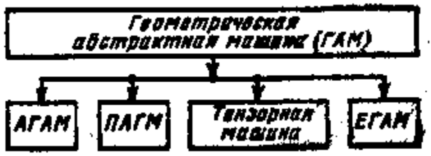


Рис. 1.15. Семейство геометрических абстрактных машин

Можно говорить о геометрической абстрактной машине общего назначения, которая может настраиваться на режимы аффинного, проективного и евклидова пространств. В схеме на рис.1.15 имеется тензорная машина, которая обеспечивает выполнение операций тензорной алгебры и исчисления. Заметим, что тензорные банки данных и их теория в значительной степени созданы. Если бы такие геометрические машины физически были реализованы, то можно было бы исследовать различные пространства и преобразования пространств и объектов в них. До появления таких специализированных компьютеров эти машины будут реализовываться как информационно-программные комплексы на универсальных компьютерах. Но в любом варианте реализации нужны инженеры-математики, которые знают на теоретическом уровне эти сравнительно «молодые» разделы математики, а также архитектуру современных вычислительных систем. В этом томе учебного пособия делается попытка решить первую часть задачи — ознакомить со специальными разделами дискретной математики.

Часть 2

ОСНОВОПОЛАГАЮЩИЕ РАЗДЕЛЫ

Глава 2

ЭЛЕМЕНТЫ ТЕОРИИ МНОЖЕСТВ

2.1. ОСНОВНЫЕ ПОНЯТИЯ

Для дальнейших рассуждений потребуется термин *множество,* который впервые введен Кантором и может быть описан следующим образом: под множеством M понимается любое объединение в одно целое определенных вполне различаемых объектов из нашего восприятия или мысли, которые называются *элементами множества* M.

Теоретико-множественные понятия будут строиться в виде аксиоматической системы подобно тому, как это принято в геометрии, не отвечая на вопросы, что такое точки, прямые, плоскости и другие «первичные термины»; из определенной системы аксиом выводятся все теоремы, при этом смысл значений первичных терминов во внимание не принимается [39, 43, 44].

Введем основные обозначения и терминологию, которые широко используются во всех разделах, занимающихся изучением алгебраических моделей. Множество, не имеющее ни одного элемента, называется пустым и обозначается символом $\varnothing$.

Пример 2.1. Множество всех целых чисел, квадраты которых отрицательные, *пустое.*

Множество, состоящее из конечного числа элементов, называется *конечным*; в противном случае оно называется *бесконечным.* Например, множество жителей Земли — конечное, а множество всех трехмерных кубов пространства — бесконечное.

Пишем, что $a \in A$, если a является элементом множества A, и $a \notin A$ в противном случае. Если каждый элемент множества A принадлежит множеству B, то пишем $A \subset B$ и говорим, что A является подмножеством множества B. Отношение $\subset$ называется *включением.*

Множество, образованное из n элементов $a_1, a_2, \ldots, a_n$, обозначается $\{a_1, a_2, \ldots, a_n\}$, в частности множество из одного элемента a обозначается $\{a\}$ и называется *единичным* множеством.

Чтобы задать множество, надо или перечислить все его элементы (что можно сделать только для конечных множеств), или указать общее свойство всех его и только его элементов. Так, например, множество всех действительных чисел, заключенных между 0 и 1, определяется свойством «быть действительным числом a, $0<a<1$».

Обобщая сказанное ранее, можно отметить, что если для каждого t из непустого множества T a_t является элементом множества A, то функция, ставящая в соответствие каждому $t \in T$ элемент $a_t \in A$, будет обозначаться $\{a_t\}_{t \in T}$, или просто $\{a_t\}$.

Каждому множеству X можно поставить в соответствие особые символы $x, x', x''\ldots$, называемые *предметными переменными* или переменными, принимающими значения в множестве X, вместо которых в выражениях, их содержащих, можно поставить любой элемент этого множества.

Множество X называется *областью возможных значений* переменных $x, x', x'', x'''\ldots$

Два множества X_1 и X_2 называются *равными*, если они состоят из одинаковых элементов. В этом случае пишут $X_1 = X_2$. Ясно, что два множества равны тогда и только тогда, когда каждое из них включается в другое.

2.2. ОПЕРАЦИИ НАД МНОЖЕСТВАМИ

Подмножеством или частью множества X называется любое множество X_1, включающееся в X.

Пример 2.2. Пусть X — множество действительных чисел. Тогда его подмножествами являются:

X_1 — множество всех рациональных чисел;
X_2 — множество всех целых чисел;
X_3 — множество всех натуральных чисел;
X_4 — единичное множество.

Для любого множества само это множество и пустое множество можно рассматривать как его подмножества, которые называются *несобственными* подмножествами; все другие подмножества множества называются *собственными.*

Над подмножествами одного и того же множества производятся операции дополнения, пересечения и объединения; первая из них — всюду определенная унарная операция (от латинского unus — один), т. е. она применяется к любому одному множеству; две другие — всюду определенные бинарные операции (от латинского bini — два), — они применяются к любым двум подмножествам.

Для любых множеств A и B символ $A \cup B$ ($A \cap B$) обозначает объединение (пересечение) множеств A и B, т. е. множество всех элементов, принадлежащих, по крайней мере, одному из множеств A или B (принадлежащих обоим множествам A и B).

Если $A \cap B = \varnothing$, то множества A и B называются *непересекающимися*. Разность множеств A и B, т. е. множество таких элементов из A, которые не принадлежат B, обозначают $A - B$.

При исследовании конкретной предметной области часто рассматривают только подмножества некоторого фиксированного множества X. Множество X называют *пространством*, а разность $X - A$ (где $A \subset X$) — *дополнением множества* A и обозначают $-A$. Поэтому, если A и $B \subset X$, то $A - B = A \cap -B$.

Пример 2.3. а) Пересечение множества всех прямоугольников на плоскости с множеством всех ромбов есть множество всех квадратов.

б) Пусть

$$X_1 = \{1,\ 2,\ \pi,\ \sqrt{10}\};$$
$$X_2 = \{1,\ \pi,\ \sqrt{7}\}.$$

Тогда

$$X_1 \cup X_2 = \{1,\ 2,\ \pi,\ \sqrt{7},\ \sqrt{10}\}.$$

в) Пусть X — множество целых чисел, а A — подмножество четных чисел, тогда $(X - A)$ — множество нечетных чисел.

2.3. ДЕКАРТОВО ПРОИЗВЕДЕНИЕ МНОЖЕСТВ

Прежде всего потребуется ввести определение *упорядоченной пары*. Если a и b — какие-либо элементы, то из них можно образовать двухэлементное множество $\{a, b\}$ (или, что то же самое $\{b, a\}$). Из этих же элементов можно образовать еще два новых объекта:

1) упорядоченную пару с первым компонентом a и вторым компонентом b, которая обозначается (a, b);

2) упорядоченную пару с первым компонентом b и вторым компонентом a, которая обозначается (b, a).

Если $a \neq b$, то упорядоченные пары (a, b) и (b, a) считаются различными. Вообще, две упорядоченные пары (a, b) и (c, d) считаются равными только в том случае, когда соответствующие их компоненты равны: $a = c$ и $b = d$.

Такое определение легко обобщается до случая n-элементных множеств. Так, из элементов $a_1, a_2, \ldots, a_n$ наряду с множеством $\{a_1, a_2, \ldots, a_n\}$ можно образовать:

1) упорядоченную n-систему с первым компонентом a_1, вторым компонентом $a_2 \ldots$, n-м компонентом a_n, которая обозначается $(a_1, a_2, \ldots, a_n)$;

2) упорядоченную n-систему с первым компонентом a_2, вторым компонентом a_1, третьим компонентом $a_3 \ldots$, n-м компонентом a_n, которая обозначается $(a_2, a_1, a_3, a_4, \ldots, a_n)$. и т. д.

Если $a_1 \neq a_2$, то упорядоченная n-система $(a_1, a_2, \ldots, a_n)$ отлична от упорядоченной n-системы $(a_2, a_1, a_3, \ldots, a_n)$. Вообще, две упорядоченные n-системы $(a_1, a_2, \ldots, a)$ и $(b_1, b_2, \ldots, b)$ считаются равными в том и только в том случае, когда $a_1 = b_1$, $a_2 = b_2$, $\ldots$, $a_n = b_n$.

Пусть теперь M и N — два множества.

Декартовым или прямым произведением $M \times N$ называется множество всех упорядоченных пар, первый и второй компоненты которых принадлежат соответственно множествам M и N.

Множество M называется первым, а множество N — вторым сомножителем декартова произведения $M \times N$. Меняя ролями множества M и N, получаем декартово произведение $N \times M$, которое отлично от декартова произведения $M \times N$ (если $M \neq N$), поскольку элементами $N \times M$ будут упорядоченные пары, первый компонент которых принадлежит множеству N, а второй — множеству M.

Пусть множество M — конечное множество, состоящее из n элементов $x_1, x_2, \ldots, x_n$, и N — конечное множество, состоящее из m элементов $y_1, y_2, \ldots, y_m$. В этом случае декартово произведение $M \times N$ также будет конечным множеством, элементами которого будут

$$(x_1, y_1), (x_1, y_2), \ldots, (x_1, y_m), (x_2, y_1), (x_2, y_2), \ldots$$
$$\ldots, (x_2, y_m), \ldots, (x_n, y_1), (x_n, y_2), \ldots, (x_n, y_m).$$

Обобщим это определение на случай n множеств M_1, $M_2, \ldots, M_n$ $(n = 2, 3)$. Декартовым или прямым произведением $M_1, M_2, \ldots, M_n$, обозначаемым $M_1 \times M_2 \times \ldots \times M_n$, называется множество всех упорядоченных n-систем, первый, второй, ..., n-й компоненты которых принадлежат соответственно 1-му, 2-му, ..., n-му множествам.

При большей строгости требуется провести следующее рассуждение. Пусть символ $\underset{t \in T}{P} A_t$ будет обозначать прямое (декартово) произведение $A_t (t \in T)$, т. е. множество всех отображений $a = \{a_t\}_{t \in T}$, таких, что $a_t \in A_t$ для всех $t \in T$. В частности, $\overset{m}{\underset{n=1}{P}} A_n$ будет обозначать множество всех m элементарных последовательностей (a_n), таких, что $a_n \in A_n$ для $n = 1$, $2, \ldots, m$. Вместо $\overset{m}{\underset{n=1}{P}} A_n$ обычно пишут $A_1 \times \ldots \times A_m$.

Если все множества равны, т. е. $A_t = A$ при любом $t \in T$, то вместо $\underset{t \in T}{P} A_t$ пишут A^T. Другими словами, A^T является множеством всех отображений из T в A. Вместо $A \times \ldots \times A$ (m раз) пишут также A^m.

Пример 2.4. 1. Пусть $x_{01}, x_{11} \in M_1$, $x_{02} \in M_2$, $x_{02} \in M_2$, $\ldots$, $x_{0n} \in M_n$. Тогда $(x_{01}, x_{02}, \ldots, x_{0n})$, $(x_{11}, x_{02}, \ldots, x_{0n}) \in M_1 \times M_2 \times \ldots \times M_n$.

2. Пусть $M=\{a, b\}$. Тогда $M^2=\{(a, a), (a, b), (b, a), (b, b)\}$, $M^3=\{(a, a, a), (a, a, b), (a, b, a), (b, a, a), (a, b, b), (b, a, b), (b, b, a), (b, b, b)\}$.

2.4. ОТНОШЕНИЯ И ФУНКЦИИ

Пусть $M_1, M_2, \ldots, M_n$ — некоторые множества ($n=2, 3\ldots$). Отношением порядка n или n-отношением между элементами множества $M_1, M_2, \ldots, M_n$ называется подмножество декартова произведения этих множеств. Множества $M_1, M_2, \ldots, M_n$ называются *базисными множествами* n-отношения, соответственно первым, вторым, ..., n-м; n-отношение, все базисные множества которого совпадают, называется *однородным*. Отношение второго порядка называется также бинарным, отношение третьего порядка — тернарным [43].

В частности, если ρ — бинарное отношение между элементами множеств M и N и упорядоченная пара (x_0, y_0) принадлежит ρ, то говорят, что элемент x_0 находится в отношении ρ с элементом y_0, в противном случае говорят, что элемент x_0 не находится в отношении ρ с элементом y_0.

Пример 2.5. Совокупность всех упорядоченных пар взаимно перпендикулярных прямых и плоскостей пространства является бинарным отношением между прямыми и плоскостями пространства. Одним из важнейших типов n-отношений является функция. Пусть X и Y — некоторые множества. Функцией, определенной на множестве X и принимающей значения в множестве Y, называется бинарное отношение f между элементами множеств X и Y, такое, что для каждого элемента множества X существует единственный элемент множества Y, с которым выбранный элемент множества M находится в отношении f.

Термины «отображение», «функция», «преобразование» всегда будут иметь одинаковый смысл. Пишем:

$$f: X \to Y,$$

чтобы указать, что f — отображение, определенное на множестве X со значениями из множества Y.

Множество X называется областью определения (доменом) функции f. Множество Y называется областью значения (диапазоном) функции f.

Функция f, определенная на множестве X и принимающая значения из множества Y, называется также отображением X в Y.

Домен и диапазон называются соответственно первой и второй проекциями f и обозначаются $\text{пр}_1 f$ и $\text{пр}_2 f$.

Отображение f множества X в множество Y называется отображением X на Y, если $\text{пр}_2 f=Y$. Отображение f множества X в множество Y называется взаимно однозначным, если образами двух любых различных элементов множества X являются различные элементы множества Y.

Пример 2.6. Отображение d множества всех отрезков некоторой прямой на множество действительных чисел, такое, что образом каждого отрезка является его длина (при выбранной единице масштаба)

$$d:(AB) \to |AB|,$$

есть отображение первого множества на второе, но не взаимно однозначное.

Операции пересечения и объединения множеств связаны между собой следующими двойственными друг к другу законами дистрибутивности: для любых трех множеств A, B, C

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C); \tag{2.1}$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{2.2}$$

В отображении φ: $A \to B$ образ элемента $a \in A$ будем обозначать через $a\varphi$ [45].

Если φ: $A \to B$, ψ: $B \to C$, то последовательное выполнение отображений φ и ψ приводит к вполне определенному отображению множества A в множество C, которое будет обозначаться $\varphi\psi$ и называться произведением отображения φ на отображение ψ. Очевидно, что для всех a из A

$$a(\varphi\psi) = (a\varphi)\psi. \tag{2.3}$$

Назовем это умножение отображений частичным: если даны два любых отображения φ: $A \to B$ и ψ: $A' \to B'$, то произведение $\varphi\psi$ существует не всегда; оно существует тогда и только тогда, когда для всех $a \in A$ имеет место $a\varphi \in A'$. Отсюда следует, что для отображений любого множества A в себя произведение всегда существует.

Умножение отображений ассоциативно: если даны отображения

$$\varphi: A \to B; \quad \psi: B \to C; \quad \chi: C \to D, \tag{2.4}$$

то

$$(\varphi\psi)\chi = \varphi(\psi\chi). \tag{2.5}$$

Действительно, если a — любой элемент из A, то в соответствии с (2.3)

$$a[(\varphi\psi)\chi] = [a[\varphi\psi)]\chi = [(a\varphi)\psi]\chi = (a\varphi)(\psi\chi) = a[\varphi(\psi\chi)].$$

В соответствии с (2.5) результат последовательного выполнения отображений (2.4) можно записать через $\varphi\psi\chi$.

Тождественное отображение множества A в себя условимся обозначать через e_A, таким образом,

$$ae_A = a \text{ для всех } a \in A.$$

Тождественное отображение играет при умножении отображений роль единицы, так как для любых отображений φ: $A \to B$ и $\psi: C \to A$

$$\varepsilon_A\varphi = \varphi; \quad \psi\varepsilon_A = \psi.$$

Очевидно, что отображение $\varphi: A \to B$ тогда и только тогда будет взаимно однозначным отображением A на B, если для него существует обратное отображение, т. е. отображение φ^{-1}: $B \to A$, которое удовлетворяет условиям

$$\varphi\varphi^{-1} = \varepsilon_A; \quad \varphi^{-1}\varphi = \varepsilon_B.$$

Если существует взаимно однозначное отображение множества A на множество B, то множества A и B называются равномощными или имеющими одну и ту же мощность, при этом мощность конечного множества совпадает с числом его элементов; множества, равномощные с множеством всех натуральных чисел, называются *счетными,* а о множествах всех действительных чисел говорят, что они имеют *мощность континиума.* При изучении бесконечных множеств очень часто используют аксиому выбора: если дано множество M, то существует функция φ, сопоставляющая каждому непустому подмножеству A из M один определенный элемент $\varphi(A)$ этого подмножества.

Иначе говоря, функция φ как бы отличает по одному элементу в каждом из непустых подмножеств множества M.

В общем случае формулировка логических основ доказательства этого положения представляет большие трудности. Но для счетных множеств она почти очевидна. Действительно, если элементы множества M пронумерованы натуральными числами, то получим требуемую функцию, если в каждом подмножестве A из M отметим тот его элемент, который имеет наименьший номер. В множестве M квадратом $M \times M$ называется множество всех упорядоченных пар (a, b), где $a, b \in M$. Пусть R — любое подмножество из $M \times M$. Оно следующим образом определяет бинарное отношение, которое также будет обозначаться символом R: если $a, b \in M$, то элемент a находится в отношении R к элементу b, и записывается это через aRb в том и только в том случае, если пара (a, b) принадлежит к подмножеству R, т. е. записи aRb и $(a, b) \in R$ равносильны.

Изучение бинарных отношений в множестве M не отличается, следовательно, от изучения подмножеств множества $M \times M$. Можно говорить, в частности, о включении бинарного отношения R в бинарное отношение R', $R \subseteq R'$, а также о пересечении и объединении бинарных отношений. Дополнением к бинарному отношению R является бинарное отношение $\overline{R}$, определяемое подмножеством

$$\overline{R} = (M \times M) \setminus R,$$

или

$$\overline{R} = (M \times M) - R,$$

иначе $a\overline{R}b$ тогда и только тогда, когда $(a, b) \notin R$. То обстоятельство, что бинарные отношения задаются множествами упорядоченных пар элементов из M, делает алгебру бинарных отношений более богатой, чем простая алгебра подмножеств произвольного множества. Так, пусть в множестве M заданы произвольные бинарные отношения R и S. Назовем их произведением RS бинарное отношение, которое определяется следующим образом: $a(RS)b$ тогда и только тогда, когда в M существует хотя бы один такой элемент c, что aRc и cSb. Умножение бинарных отношений ассоциативно

$$(RS)T = R(ST),$$

так как элемент a тогда и только тогда находится в каждом из отношений $(RS)T$ и $R(ST)$ к элементу b, когда существуют элементы c и d, такие, что aRc, cSd, dTb.

Умножение бинарных отношений в общем случае не является коммутативным: бинарные отношения R и S лишь иногда являются перестановочными: $RS = SR$.

Если в множестве M даны бинарные отношения R_i (i пробегает множество индексов I) и S, то

$$(\bigcup R_i)S = \bigcup_{i \in I} R_iS; \quad S\Big(\bigcup_{i \in I} R_i\Big) = \bigcup_{i \in I} SR_i. \tag{2.6}$$

Действительно, $a\Big[\Big(\bigcup\limits_{i \in I} R_i\Big)S\Big]c$ равносильно существованию такого элемента c, что $a\Big(\bigcup\limits_{i \in I} R_i\Big)c$ и cSb. Это равносильно существованию такого индекса i_0, что $aR_{i_0}c$ и cSb, т. е. $a(R_0S)b$, и поэтому $a\Big(\bigcup\limits_{i \in I} R_{i_0}S\Big)b$.

Заметим, что в равенствах (2.6) объединения нельзя заменять пересечениями. Из (2.6) следует, что если даны бинарные отношения R, R' и S, причем $R \subseteq R'$, то

$$RS \subseteq R'S; \quad SR \subseteq SR'. \tag{2.7}$$

Действительно, включение $R \subseteq R'$ равносильно равенству $(R \cup R') = R'$, из которого следует равенство

$$(R \cup R')S = RS \cup R'S = R'S,$$

равносильное включению

$$RS \subseteq R'S.$$

Для всякого бинарного отношения R в множестве M существует обратное отношение R^{-1}, определяемое следующим образом: $aR^{-1}b$ тогда и только тогда, когда bRa. Ясно, что

$$(R^{-1})^{-1} = R$$

и что из $R \subseteq S$ следует $R^{-1} \subseteq S^{-1}$.

Если в множестве M даны бинарные отношения R_i, $i \in I$, S и T, то

$$\left(\prod_{i \in I} R_i\right)^{-1} = \prod_{i \in I} R_i^{-1}; \tag{2.8}$$

$$\left(\bigcup_{i \in I} R_i\right)^{-1} = \bigcup_{i \in I} R_i^{-1}; \tag{2.9}$$

$$(ST)^{-1} = T^{-1}S^{-1}. \tag{2.10}$$

Действительно, $a\left(\prod_{i \in I} R_i\right)^{-1} b$ означает, что $b\left(\bigcap_{i \in I} R_i\right)a$, т. е. bR_ia для всех $i \in I$, откуда $aR_i^{-1}b$ для всех $i \in I$, и поэтому $a\left(\bigcap_{i \in I} R_i^{-1}\right)b$. Аналогично доказывается и равенство (2.9). Далее $a(ST)^{-1}b$ означает, что $b(ST)a$, т. е. существует такой элемент c, что bSc и cTa, а поэтому $aT^{-1}c$ и $cS^{-1}b$, откуда $a(T^{-1}S^{-1})b$.

Единичное отношение E определяется следующим образом: aEb тогда и только тогда, когда $a = b$, иначе отношение задается множеством всех пар вида (a, a), $a \in M$. Очевидно, что $E^{-1} = E$, и для любого бинарного отношения R

$$RE = ER = R.$$

Отметим пустое отношение O, определяемое пустым подмножеством множества $M \times M$. Ясно, что для любого бинарного отношения R в множестве M $O \subseteq R$ и $RO = OR = O$. Представляют интерес такие бинарные отношения R, заданные в множестве M, которые обладают некоторыми из следующих четырех свойств (рис. 2.1):

рефлексивность: aRa для всех $a \in M$, иначе $E \subseteq R$;

транзитивность: если aRb и bRc, то aRc, иначе $RR \subseteq R$;

симметричность: если aRb, то bRa, иначе $R^{-1} = R$;

антисимметричность: если aRb, то $a = b$, иначе $R \cap R^{-1} \subseteq E$.

Эти свойства определяют тип бинарных отношений (рис. 2.2). Если бинарное отношение R обладает любым из указанных четырех свойств, то обратное отношение R^{-1} обладает этим же свойством. Действительно, если $E \subseteq R$, то $E = E^{-1} \subseteq R^{-1}$. Если $RR \subseteq R$, то $R^{-1}R^{-1} = (RR)^{-1} \subseteq R^{-1}$. Если $R^{-1} = R$, то $(R^{-1})^{-1} = R = R^{-1}$. Наконец, если $R \cap R^{-1} \subseteq E$, то $R^{-1} \cap (R^{-1})^{-1} = R^{-1} \cap R \subseteq E$.

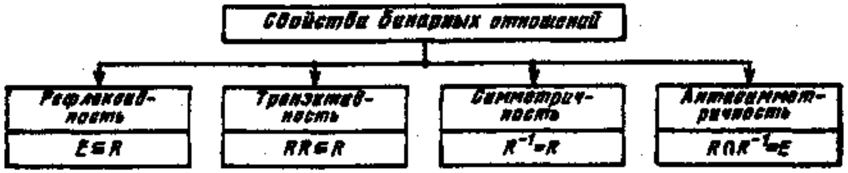


Рис. 2.1. Пояснение к свойствам бинарных отношений в множествах

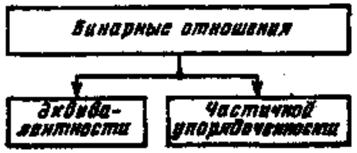


Рис. 2.2. Типы бинарных отношений в множествах

Пусть в множестве M выбрано подмножество N. Бинарное отношение R, заданное в M, естественным образом индуцирует бинарное отношение R^N в множестве N: если $a, b \in N$, то aR^Nb тогда и только тогда, когда в M справедливо aRb. Иначе, учитывая, что $N \times N$ является подмножеством множества $M \times M$,

$$R^N = R \cap (N \times N).$$

Легко проверить следующие равенства:

$$(R^N)^{-1} = (R^{-1})^N; \quad \left(\bigcap_{i \in I} R_i\right)^N = \bigcap_{i \in I} R_i^N; \quad \left(\bigcup_{i \in I} R_i\right)^N = \bigcup_{i \in I} R_i^N.$$

Понятие бинарного отношения допускает различные обобщения. Так, рассмотрим n-ю степень M^n множества M, т. е. множество всех упорядоченных систем $(a_1, a_2, \ldots, a_n)$ из n элементов множества M. Тогда любое подмножество R множества M^n определяет в множестве M n-е отношение (при $n = 3$ — тернарное отношение). Множества, в которых задано некоторое число таких отношений, называются моделями и являются предметом изучения самостоятельной теории.

2.5. ОТНОШЕНИЕ ЭКВИВАЛЕНТНОСТИ

Отношение эквивалентности является такого типа бинарным отношением, которое обладает свойствами рефлексивности, транзитивности и симметричности (рис. 2.3). Из многочисленных конкретных примеров таких отношений напомним два: равенство дробей и сравнимость целых чисел по некоторому модулю. Для обозначения отношения эквивалентности используются обозначения $\sim$ и $\equiv$ [45].

Отношение эквивалентности, определенное на множестве M, тесно связано с разбиением множества M на непересекающиеся классы. Под разбиением понимается такой выбор в множестве M системы подмножества (классов этого разбиения), что всякий элемент из M принадлежит к одному и только одному из этих подмножеств.

Всякое разбиение π множества M определяет в M отношение эквивалентности. Действительно, если $a, b \in M$ и если положим $a \sim b$ в том и только в том случае, когда a и b принадлежат к одному классу разбиения π, то получим в M бинарное

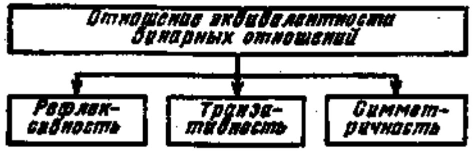


Рис. 2.3. Пояснение к отношению эквивалентности бинарных отношений в множествах

отношение, удовлетворяющее, очевидно, всем требованиям отношения эквивалентности, и, наоборот всякое отношение эквивалентности R, заданное в множестве M, определяет разбиение этого множества.

Действительно, назовем классом элемента a и обозначим через K_a множество всех тех элементов x из M, для которых aRx. Из рефлексивности отношения R вытекает включение $a \in K_a$, т. е. система классов K_a, а $a \in M$ покрывает все множество M. Далее, симметричность отношения R обеспечивает, что из $b \in K_a$ следует $a \in K_b$, транзитивность же отношения R приводит к тому, что если $b \in K_a$, то из $c \in K_b$ следует $c \in K_a$, т. е. $K_b \subseteq K_a$. Эти последние замечания приводят к тому, что если $b \in K_a$, то $K_b = K_a$, т. е. класс определяется любым своим элементом.

Если K_a и K_b — два произвольных класса с непустым пересечением, содержащим, например, элемент c, то $K_a = K_c$ и $K_b = K_c$, т. е. классы K_a и K_b совпадают. Тем самым мы показали, что система всех классов вида K_a является разбиением множества M. Очевидно, что переход от разбиения π множества M к определяемому им отношению эквивалентности, а затем к определяемому последним разбиению множества M приводит снова к разбиению π.

Таким образом, между отношениями эквивалентности в множестве M и разбиениями множества M на непересекающиеся классы установлено взаимно однозначное соответствие.

Если в множестве M заданы отношения эквивалентности R_i, $i \in I$, то их пересечение также будет отношением эквивалентности. В самом деле, из aR_ia для всех $i \in I$ следует $a\left(\bigcap_{i \in I} R_i\right)a$. Далее, если $a\left(\bigcap_{i \in I} R_i\right)b$ и $b\left(\bigcap_{i \in I} R_i\right)c$, то aR_ib и bR_ic для всех $i \in I$, откуда aR_ic для всех $i \in I$, а поэтому $a\left(\bigcap_{i \in I} R_i\right)c$. Наконец, если $a\left(\bigcap_{i \in I} R_i\right)b$, то aR_ib для всех $i \in I$, т. е. bR_ia для всех $i \in I$, откуда $b\left(\bigcap_{i \in I} R_i\right)a$.

Можно убедиться, что если отношениям эквивалентности R_i, $i \in I$ соответствуют разбиения π_i множества M, то отношению эквивалентности $\bigcap_{i \in I} R_i$ соответствует разбиение, классами которого служат все непустые пересечения классов, взятых по одному в каждом из разбиений π_i, $i \in I$. Это разбиение назовем пересечением разбиений π_0, $i \in I$. Если в множестве M заданы отношения эквивалентности R_i, $i \in I$, то их объединение, понимаемое в смысле объединения бинарных отношений, уже не будет, вообще говоря, отношением эквивалентности.

В множестве M существует, однако, такое отношение эквивалентности, которое включает в себя все отношения R_i, $i \in I$ (в смысле включения бинарных отношений), но само отношение включается в любое другое отношение эквивалентности, включающее в себя все R_i, $i \in I$. Это отношение эквивалентности может быть названо объединением отношений эквивалентности R_i, $i \in I$.

Для доказательства определим в множестве M бинарное отношение S следующим образом: aSb тогда и только тогда, когда в M можно хотя бы одним способом выбрать такую конечную систему элементов

$$a = c_0,\ c_1,\ c_2,\ .\ .\ .\ .\ ,\ c_{n-1},\ c_n = b, \tag{2.11}$$

что для $k=1,\ 2\ldots$ существует хотя бы один такой индекс $i_k \in I$, для которого $C_{k-1}R_{i_k}C_k$. Отношение S является, очевидно, рефлексивным, транзитивным и симметричным. Если T — любое отношение эквивалентности, включающее в себя все R_i, $i \in I$, и если aSb, причем этим элементам соответствует система элементов (2.11), то из $C_{k-1}R_{i_k}C_k$ следует $C_{k-1}TC_k$ ($k=1,\ 2,\ \ldots$ $\ldots, n$), а поэтому ввиду транзитивности отношения T имеет место aTb.

Произведение RS двух отношений эквивалентности R и S тогда и только тогда является отношением эквивалентности, когда отношения R и S перестановочны, т. е. $RS = SR$. Если это имеет место, то объединение отношений эквивалентности R и S совпадает с их произведением как бинарных отношений. Множество классов разбиения, соответствующего данному отношению эквивалентности R в множестве M, будем обозначать через M/R и называть фактор-множеством множества M по отношению к эквивалентности R. Отображение множества M на фактор-множество M/R, сопоставляющее всякому элементу $a \in M$ тот класс разбиения, соответствующего R, в котором лежит элемент a, называется естественным отображением M на M/R. Между отношениями эквивалентности, имеющимися в множестве M, и отображениями этого множества на некоторые другие множества существует тесная связь, являющаяся прототипом так называемых «теорем о гомоморфизмах». Если дано отображение φ множества M на некоторое множество N, то ему соответствует вполне определенное отношение эквивалентности R в множестве M (т. е. разбиение этого множества): для элементов a, $b \in M$ полагаем aRb в том и только в том случае, если $a\varphi = b\varphi$. Сопоставляя каждому элементу x из N класс тех элементов из M, которые имеют x своим образом при отображении φ, получаем взаимно однозначное отображение ζ множества N на множество M/R, причем произведение $\varphi\zeta$ совпадает с естественным отображением M на M/R.

2.6. ЧАСТИЧНО УПОРЯДОЧЕННЫЕ МНОЖЕСТВА (ЧУМ)

Частичная упорядоченность является другим очень важным типом бинарных отношений, которые обладают свойствами рефлексивности, транзитивности и антисимметричности (рис. 2.4). Множество M с заданной в нем частичной упорядоченностью называется *частично упорядоченным*. Для записи частичной упорядоченности будет использоваться символ $\leqslant$; если $a, b \in M$ и $a \leqslant b$, то в зависимости от обстоятельств будем говорить, что a меньше или равно b, a содержится в b, a предшествует b. Если $a \leqslant b$ и $a \neq b$, то будем писать $a < b$ и говорить, что a меньше b, a строго содержится в b и т. д. Бинарное отношение $a < b$ уже не будет рефлексивным.

Через знаки $\geqslant$ и $>$ будут обозначаться отношения, обратные отношениям $\leqslant$ и $<$, т. е., например, $a \geqslant b$ (a больше или равно b, a содержит b, a следует за b) тогда и только тогда, когда $b \leqslant a$.

Пусть в множестве M задана частичная упорядоченность. Элементы a и b этого множества называются сравнимыми, если $a \leqslant b$ или $b \leqslant a$. Необязательно всякие два элемента из M должны быть сравнимыми, поэтому и используется термин «частичная» упорядоченность. Так можно получить тривиальную частичную упорядоченность множества M, если положить, что $a \leqslant b$ лишь при $a = b$; различные элементы из M будут в этом случае несравнимыми. Частично упорядоченное множество, в котором любые два элемента сравнимы, называется упорядоченным множеством или линейно упорядоченным множеством или цепью (рис. 2.5).

В качестве первых примеров упорядоченных множеств можно указать множество натуральных чисел и множество точек на прямой линии (множество всех действительных чисел), оба в их естественной упорядоченности. Следующими примерами ЧУМ являются:

Рис. 2.4. Пояснение к частично упорядоченным бинарным отношениям в множествах

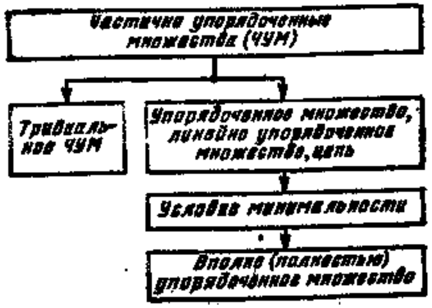


Рис. 2.5. Разновидности упорядоченных множеств

1) множество $\tilde{N}$ всех подмножеств некоторого данного множества N с отношением теоретико - множественного включения $\subseteq$ в качестве отношения частичной упорядоченности;

2) множество всех непрерывных действительных функций, определенных на отрезке [0,1], если $f \leqslant g$ означает, что для всех x этого отрезка $f(x) \leqslant g(x)$;

3) множество всех натуральных чисел, если $a \leqslant b$ понимать в том смысле, что b делится нацело на a.

Пусть между ЧУМ M и ЧУМ M' установлено взаимно однозначное соответствие φ:

$$a\varphi = a'; \quad a \in M; \cdot a' \in M'.$$

Если из $a \leqslant b$, где $a, b \in M$, всегда следует $a\varphi \leqslant b\varphi$ и наоборот, то φ называется *изоморфизмом* между M и M', а сами множества M и M' — изоморфными ЧУМ. Очевидно, что в тех случаях, когда частичная упорядоченность является самостоятельным объектом изучения, а природа элементов, из которых составлены множества, не играет роли, изоморфные множества можно считать тождественными. Из предыдущего известно, что частичная упорядоченность, заданная на множестве M, индуцирует во всяком подмножестве этого множества некоторое бинарное отношение; нетрудно убедиться, что оно будет обладать свойством частичной упорядоченности. Будем считать, что ЧУМ M изоморфно вкладывается в ЧУМ N, если существует изоморфное отображение множества M на некоторое подмножество N' множества N, причем N' рассматривается с частичной упорядоченностью, индуцированной в нем частичной упорядоченностью множества N.

Всякое ЧУМ M изоморфно вкладывается в множество $\tilde{N}$ всех подмножеств некоторого множества N, частично упорядоченного по включению. В качестве множества N можно взять, например, само M.

Действительно, поставим в соответствие каждому элементу a из M подмножество A, составленное из всех таких элементов $x \in M$, что $x \leqslant a$. Пусть $a, b \in M$, а A, B — соответствующие им подмножества. Если $A = B$, то $b \leqslant a$, $a \leqslant b$, откуда $a = b$. Этим доказано, что соответствие $a \to A$ является взаимно однозначным отображением множества M в множество $\tilde{M}$ всех его подмножеств. Если далее $a \leqslant b$, то из $x \leqslant a$ будет следовать $x \leqslant b$, т. е. $A \subseteq B$. Наоборот, если $A \subseteq B$, то $a \subseteq b$, т. е. $a \leqslant b$. Таким образом, соответствие $a \to A$ является изоморфным вложением M в $\widetilde{M}$.

Отношение, обратное к отношению частичной упорядоченности, само будет частично упорядоченным. Частично упорядоченные множества M и M' называются *инверсно изоморфными*, если одно из них изоморфно другому, взятому с обратной частичной упорядоченностью, т. е. если между ними существует такое взаимно однозначное соответствие φ

$$a\varphi = a',\ a \in M,\ a' \in M',$$

что $a \leqslant b$, где a, $b \in M$ тогда и только тогда, когда $a\varphi \geqslant b\varphi$.

Условие минимальности вводится следующим образом. Элемент a ЧУМ M называется минимальным элементом этого множества, если в M нет ни одного элемента x, удовлетворяющего условию $x < a$. Ясно, что M может содержать много различных минимальных элементов, но может также не иметь ни одного такого элемента. К примеру, множество $\widetilde{N}$ всех подмножеств некоторого множества N обладает единственным минимальным элементом, которым является пустое подмножество. В множестве всех непустых подмножеств множества N минимальными элементами являются все подмножества, состоящие из одного элемента. Наконец, если множество N бесконечное, то множество всех его бесконечных подмножеств вообще не имеет минимальных элементов. С помощью понятия минимального элемента можно определить один специальный класс частично упорядоченных множеств, более широкий, чем класс конечных ЧУМ. Этот класс ЧУМ удовлетворяет следующим условиям.

Условие минимальности. Всякое непустое подмножество N ЧУМ M обладает хотя бы одним минимальным (bN) элементом.

Условие обрыва убывающих цепей. Всякая строго убывающая цепь элементов ЧУМ множества M

$$a_1 > a_2 > \ldots > a_n > \ldots$$

обрывается на конечном месте. Иными словами, для всякой убывающей цепи элементов

$$a_1 \geqslant a_2 \geqslant \ldots \geqslant a_n \geqslant \ldots$$

существует такой индекс n, на котором эта цепь стабилизируется, т. е.

$$a_n = a_{n+1} = \ldots$$

Условие индуктивности. Все элементы ЧУМ M обладают некоторым свойством ε, если этим свойством обладают

Рис. 2.6. Пояснение к свойствам частично упорядоченных множеств (ЧУМ)

все минимальные элементы этого множества (в случае, когда они существуют) и если из справедливости свойства ε для всех элементов M, строго предшествующих некоторому элементу a, может быть выведена справедливость этого свойства для самого элемента a.

Докажем эквивалентность указанных трех условий (рис. 2.6). Из условия минимальности вытекает условие индуктивности. Действительно, пусть ЧУМ M удовлетворяет условию минимальности и пусть в нем для некоторого свойства выполняются посылки условия индуктивности. Если тем не менее в M существуют элементы, которые не обладают свойством ε, то пусть a будет одним из минимальных среди этих элементов — существование элемента a вытекает из условия минимальности. Элемент a не может быть минимальным во всем M, что следует из первой посылки условия индуктивности, а так как все элементы, строго предшествующие a, свойством ε уже обладают, то по второй посылке условия индуктивности и сам элемент a должен обладать свойством ε, т. е. приходим к противоречию.

Из условия индуктивности вытекает условие обрыва убывающих цепей.

Действительно, пусть ЧУМ M удовлетворяет условию индуктивности. Применим это условие к следующему свойству: элемент a обладает свойством ε, если всякая строго убывающая цепь элементов, начинающаяся от элемента a, обрывается на конечном месте. Этим свойством обладают, очевидно, все минимальные элементы множества M, если они существуют. С другой стороны, пусть все элементы, строго предшествующие элементу a, обладают свойством ε. В этом случае второй член любой строго убывающей цепи, которая начинается от элемента a, будет обладать свойством ε, а поэтому рассматриваемая цепь должна обрываться, т. е. элемент a тоже обладает свойством ε. Из условия индуктивности теперь следует, что указанным свойством обладают вообще все элементы множества M, т. е. в M обрывается любая строго убывающая цепь — она начинается с некоторого элемента.

Из условия обрыва убывающих цепей вытекает условие минимальности. Для доказательства этого положения предположим, что ЧУМ M условию минимальности не удовлетворяет, а именно: пусть его непустое множество N не имеет минимальных элементов. Пользуясь аксиомой выбора, отметим по одному элементу в каждом непустом подмножестве из N, а затем следующим образом построим последовательность элементов:

$$a_n, \quad n=1,\ 2\ \ldots \tag{2.12}$$

В качестве a_1 возьмем элемент, отмеченный в самом подмножестве N. Если элемент a_n уже построен и $a_n\in N$, то в качестве a_{n+1} берем элемент, отличный в непустом (так как N не имеет минимальных элементов) множестве элементов из N, строго предшествующих a_n. Последовательность (2.12) является, очевидно, бесконечной строго убывающей цепью, т. е. множество M не может удовлетворять условию обрыва убывающих цепей. Условие индуктивности позволяет проводить не только доказательство по индукции, но и построение по индукции. Именно, пусть M—ЧУМ с условием минимальности, и нужно определить на этом множестве функцию $\varphi(x)$, которая ставит в соответствие каждому элементу x из M некоторый элемент вспомогательного множества S'. Будем считать, что $\varphi(x)$ должна удовлетворять некоторым рекуррентным соотношениям, т. е. соотношениям, однозначно определяющим для всякого $a\in M$ значение $\varphi(a)$ по значениям $\varphi(b)$ для всех b, строго меньших a. Докажем, что существует, притом единственная, функция $\varphi(x)$, определенная на всем множестве M, которая удовлетворяет указанным рекуррентным соотношениям и принимает произвольные заданные значения на всех минимальных элементах множества M.

Доказательство начнем с единственности. Пусть на M существуют две различные функции $\varphi(x)$ и $\psi(x)$, удовлетворяющие нашим условиям. В непустом множестве тех элементов x, для которых $\varphi(x)\neq\psi(x)$, существует в соответствии с условием минимальности хотя бы один минимальный элемент a. Этот элемент не может быть минимальным во всем M, так как на минимальных элементах множества M функции $\varphi(x)$ и $\psi(x)$ по условию совпадают. Существуют, следовательно, такие элементы b, что $b<a$, причем для всех этих элементов $\varphi(b)=\psi(b)$. Рекуррентные отношения однозначно определяют значения рассматриваемых функций для $x=a$ по их значениям для всех $b<a$, а поэтому $\varphi(a)=\psi(a)$, т. е. мы пришли к противоречию.

Для доказательства существования искомой функции $\varphi(x)$ предполагаем, что на минимальных элементах ее значения уже заданы. Будем говорить, что элемент $a\in M$ обладает свойством ε, если на множестве A всех таких x, что $x\leqslant a$, может быть определена функция $\varphi_a(x)$, удовлетворяющая заданным рекуррентным соотношениям и принимающая заданные значения на минимальных элементах из M, содержащих a.

Все минимальные элементы из M обладают свойством ε. С другой стороны, если a и b обладают свойством ε и $b<a$, то, применяя доказанное ранее положение о единственности искомых функций вместо M к множеству B тех x, для которых $x\leqslant b$, получаем, что для всех этих x

$$\varphi_b(x)=\varphi_a(x).$$

Отсюда следует, что если все элементы b, строго предшествовавшие данному элементу a, обладают свойством ε, то этим свойством обладает и сам элемент a: получим функцию $\varphi_a(x)$, удовлетворяющую всем требованиям, если для всякого b ($b>a$) положим

$$\varphi_a(b)=\varphi_b(b),$$

а в качестве $\varphi_a(a)$ возьмем то значение, которое однозначно определяется рекуррентными соотношениями.

На основании условия индуктивности, которое выполняется в множестве M, можно теперь утверждать, что все элементы этого множества обладают свойством ε. Если предположим, наконец, для всех $a\in M$, что

$$\varphi(a)=\varphi_a(a),$$

то определим функцию $\varphi(x)$, обладающую всеми нужными свойствами, и этим закончим доказательство теоремы.

Линейно упорядоченное множество, удовлетворяющее условию минимальности, а поэтому и двум другим условиям, с ним эквивалентным, называется *вполне упорядоченным*. Примером вполне упорядоченного множества служит множество натуральных чисел в его естественной упорядоченности. Всякое подмножество вполне упорядоченного множества само вполне упорядочено. Из вполне упорядоченного множества следует, что оно обладает единственным минимальным элементом. Во вполне упорядоченном множестве для всякого элемента a существует элемент, непосредственно следующий за a (за единственным возможным исключением, если a является максимальным элементом). Элемент a может и не иметь непосредственно предшествующего элемента, в этом случае он называется *предельным элементом*.

ЧУМ тогда и только тогда удовлетворяет условию минимальности, если все его цепи (т. е. линейно упорядоченные подмножества) вполне упорядочены.

Действительно, если ЧУМ M удовлетворяет условию минимальности, то это же верно для всех его подмножеств, в частности, для всех цепей. Обратное утверждение вытекает из того, что в формулировке условия обрыва убывающих цепей, эквивалентного условию минимальности, используются лишь цепи множества M. В ЧУМ M можно перейти к обратной частичной упорядоченности. Минимальные элементы этой обратной упорядоченности называются *максимальными элементами множества* M в его исходной упорядоченности, а убывающие цепи в обратной упорядоченности — *возрастающими цепями множества* M. Вообще для всякого понятия (или утверждения), связанного с частичной упорядоченностью, можно получить двойственное понятие (утверждение).

Пусть ЧУМ M удовлетворяет условию минимальности. Беря в M обратную частичную упорядоченность, получаем ЧУМ, удовлетворяющее условию минимальности. Для множеств с условием максимальности после замены отношения $\leqslant$ на отношение $\geqslant$ и обратно остается справедливым все сказанное выше о множествах с условием минимальности. Помимо вполне частично упорядоченных множеств, которые обладают одним максимальным или минимальным элементом, нас будет интересовать полностью частично упорядоченные множества (ПЧУМ) [46, 82].

Определение 2.6.1. Пусть $D=(D, \leqslant)$ — частично упорядоченное множество (ЧУМ) с рефлексивным отношением $\leqslant$.

1. Подмножество $X \subseteq D$ называется *направленным*, если $X \neq \varnothing$ и

$$(\forall x, y \in X)(\exists z \in X)[x \leqslant z \text{ и } y \leqslant z].$$

2. Множество D называется *полным*, если:

1) имеется наименьший элемент $\bot \in D$, т. е. имеет место $\forall x \in D(\bot \leqslant x)$, такой элемент $\bot$ называется *дном;*

2) для любого направленного подмножества $X \subseteq D$ существует точная верхняя грань (супремум) $\sqcup X \in D$.

Полная решетка — это по определению такое ЧУМ D, в котором *любое* подмножество $X \subseteq D$ имеет супремум. Каждая полная решетка есть ПЧУМ, так как $\bot = \sqcup \varnothing$.

Пример 2.7. 1. Пусть X — множество, а $P(X)$ — его множество-степень, т. е. множество всех подмножеств множества X. Тогда $(P(X), \subseteq)$ — ПЧУМ (и даже полная решетка).

2. Пусть $\Pi = (\{\bot, t, f\}, <)$, причем $\bot < t$, $\bot < f$ (и нет никаких других соотношений для $<$). Тогда Π — ПЧУМ.

Определение 2.6.2. Пусть $(D, \sqsubseteq)$ — ПЧУМ.

Топология Скотта на D определяется следующим образом.

Множество $O \subseteq D$ считается открытым, если:

1) $x \in O \wedge x \leqslant y \Rightarrow y \in O$;

2) $\sqcup X \in O$, где $X \subseteq D$ и X направлено $\Rightarrow X \cap O = \varnothing$.

Очевидно, что $\varnothing$ и D открыты и что открытость сохраняется при произвольных объединениях. Если множества O_1 и O_2 открыты, то $O_1 \cap O_2$ является открытым из-за того, что X в посылке условия 2 направленно. Поэтому определение 2.6.2 действительно вводит топологию. Обозначим через CPO (complete partial order) категории, объектами которой являются ПЧУМ, а морфизмами — непрерывные отображения.

Теорема 2.1. CPO — декартово-замкнутая категория [46].

Обычно всегда считается, что если специально не оговорено, то ПЧУМ всегда будет наделено топологией Скотта. В дальнейшем, как правило, символы $D=(D, \sqsubseteq)$, $D'=(D', \leqslant')\dots$ будут обозначать ПЧУМ.

Прямое произведение двух ПЧУМ можно снова считать ПЧУМ.

Предложение. Для данных D, D' определим частичное упорядочение их декартова произведения $D \times D'$ соотношением $\langle x, x' \rangle \leqslant \langle y, y' \rangle$, когда $x \leqslant y$ и $x' \leqslant y'$. Тогда $D \times D'$ — ПЧУМ и для направленного подмножества $X \subseteq D \times D'$ имеет место $\sqcup X = \langle \sqcup X_0, \sqcup X_1 \rangle$, где $X_0 = \{x \in D / \exists x' \in D'(\langle x, x' \rangle \in X)\}$ и $X_1 = \{x' \in D' / \exists x \in D \langle x, x' \rangle \in X))\}$.

Доказательство. Действительно, $\langle \bot, \bot' \rangle$ — для произведения $D \times D'$. Далее замечаем, что если множество $X \subseteq D \times D'$ направленно, то направленны также X_0, X_1.

Топология на $D_0 \times D'$ не является в общем случае произведением топологий, имевшихся на D и D'.

2.7. МОЩНОСТИ, ПОРЯДКОВЫЕ ТИПЫ И ЧИСЛА В ТЕОРИИ МНОЖЕСТВ

Каждому множеству A поставим в соответствие объект $|A|$ так, что $|A| = |B|$ тогда и только тогда, когда существует взаимно однозначно отображение множества A на множество B [44].

Мощности множеств называются также кардинальными числами, или кардиналами. Кардинал множества A обозначается Card A. Так, кардиналом пустого множества $\varnothing$ является число 0, множества $\{a_1, \dots, a_n\}$ из n элементов ($n=1, 2\dots$) — число n. Кардинал множеств всех натуральных чисел обозначается N_0, а множества всех действительных чисел — N. Множества мощностью N_0 называются счетными, а мощность N называется мощностью континиума.

Положим Card $A \leqslant$ Card B, если существует взаимно однозначное отображение множества A в множество B. Легко проверить, что отношение $\leqslant$ является линейным порядком.

Взаимно однозначное отображение φ линейно упорядоченного множества A на линейно упорядоченное множество B, сохраняющее линейный порядок, т. е. для всех a, $b \in A$ $a \leqslant b$ влечет $\varphi(a) \leqslant \varphi(b)$ и наоборот, называется изоморфизмом A на B.

Если для множеств A и B существует изоморфизм, то множества называются изоморфными.

Каждому линейно упорядоченному множеству A поставим в соответствие объект $O(A)$, называемый порядковым типом, таким образом, чтобы $O(A)=O(B)$ тогда и только тогда, когда линейно упорядоченные множества A и B изоморфны.

Поскольку равенство порядковых типов $O(A)=O(B)$ влечет равенство мощностей $\mathrm{Card}\,A=\mathrm{Card}\,B$, то каждому порядковому типу соответствует некоторая мощность, которая называется мощностью этого типа.

Каждое множество $\{a_1, \ldots, a_4\}$, состоящее из n элементов $(n=1, 2\ldots)$, допускает $n!$ перестановок и поэтому может быть линейно упорядочено $n!$ способами. Однако все получающиеся при этом линейно упорядоченные множества имеют один и тот же порядковый тип, который может быть обозначен через n.

Пустому множеству приписывается порядковый тип 0. Множеству всех натуральных чисел, линейно упорядоченному по возрастанию, приписывается порядковый тип ω, а по убыванию — ω^*.

Линейно упорядоченное множество A, для которого всякое непустое подмножество B — элемент, называется вполне упорядоченным.

Порядковый тип вполне упорядоченного множества называется порядковым (трансфинитным) числом или ординалом.

Нетрудно видеть, что порядковые типы n, ω $(n=0, 1, 2, \ldots)$ являются порядковыми числами, а порядковый тип ω^* порядковым числом не является.

Нетрудно показать, что всякое множество ординалов равно, как и всякое множество кардиналов является вполне упорядоченным.

Помимо порядкового типа существуют еще, по крайней мере, два типа в теории множеств: тип реляционной системы (TR) и системный тип (который рассмотрен в § 4.10). Для введения понятия типа рассмотрим реляционные системы TR.

Пусть A — множество, R_0, R_1, …, R_{k-1} — отношения соответственно от P_0, P_1, …, P_{k-1} переменных из A (иначе $R_j \subset \subset A^{P_j}$ для $j<k$). Система $\langle A, R_0, R_1, \ldots, R_{k-1}\rangle$ называется реляционной системой характеристики $(P_0, P_1, \ldots, P_{k-1})$, множество A — полем этой системы.

Так как функция — это отношение специального вида, то можно считать группу реляционной системой характеристики множества 1 и $X \subset 1$, где

$$f_X(x)=\begin{cases} 1, & \text{если } x \in X; \\ 0, & \text{если } x \subset 1-X, \end{cases}$$

а кольцо (см. гл. 3) является реляционной системой соответствующей характеристики [44].

Две реляционные системы $\langle A, R\rangle$ и $\langle B, S\rangle$ называются изоморфными, если существует взаимно однозначная функция f, отображающая A на B так, что для всех $x, y \in A$

$$xRy \equiv f(x)\,Sf(y).$$

В этом случае используются обозначения $\langle A, R\rangle \approx \langle B, S\rangle$ или $R \approx S$, если ясно, какие множества A и B имеются в виду.

Можно показать [44], что каждое свойство системы $\langle A, R\rangle$, которое можно выразить с помощью символов исчисления высказываний и кванторов, ограниченных полем системы, присуще также каждой системе, которая изоморфна $\langle A, R\rangle$, т. е. является инвариантом относительно изоморфизма. О двух изоморфных системах говорят, что они имеют один и тот же тип. Более строго тип вводится следующим образом. Выражение $\alpha TR\langle A, R\rangle$ читается α, является типом реляционной системы $\langle A, R\rangle$, причем считается, что в каждой реляционной системе $\langle A, R\rangle$, где $R \subset A^2$, существует точно один такой объект α, что $\alpha TR\langle A, R\rangle$. Для произвольных $\langle A, R\rangle$ и $\langle B, S\rangle$ справедливо

$$(\alpha TR\langle A, R\rangle) \wedge (\beta TR\langle B, S\rangle) \rightarrow [(\alpha=\beta) \equiv \langle A, R\rangle \approx \langle B, S\rangle].$$

Этот единственный объект α, для которого $\alpha TR\langle A, R\rangle$, обозначается $\overline{\langle A, R\rangle}$, или просто R. Объект α называется реляционным типом, если существует система $\langle A, R\rangle$, для которой $\alpha TR\langle A, R\rangle$.

Глава 3

ОБЩАЯ АЛГЕБРА

3.1. ГРУППОИДЫ, ПОЛУГРУППЫ, ГРУППЫ

В настоящее время под алгеброй понимается теоретико-множественная, аксиоматическая наука, основным объектом изучения которой являются алгебраические операции, производимые над объектами произвольной природы [45]. Наиболее развитыми и традиционными разделами общей алгебры являются теория полей и колец, теория групп. Но для инженера-математика очень важны новые области общей алгебры, условно изображенные на рис. 3.1. Эта схема в значительной степени является связующей между гл. 3 и 4, которая посвящена теории категорий, представляющей дальнейшее развитие структурных методов в алгебре. Большое направление гомологической алгебры обеспечивает единый подход к геометрическим разделам данной книги (см. гл. 13, 14).

Топологическая алгебра и алгебраическая геометрия, проективная и дифференциальная алгебра послужили основой для

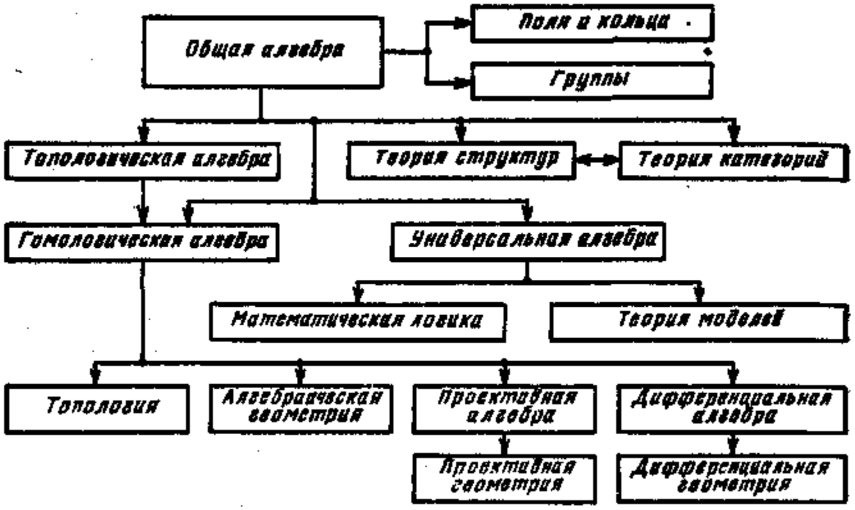


Рис. 3.1. Структура общей алгебры

написания раздела по алгебраическим, аффинной, эвклидовой, проективной и дифференциальной геометриям (см. гл. 14). И наконец, универсальная алгебра дает (устанавливает связь) «выход» на математическую логику и теорию моделей.

Среди общего набора алгебраических операций остановимся пока только на бинарных операциях. Под этими операциями будет пониматься правило, по которому некоторым упорядоченным парам элементов данного множества M ставятся в соответствие элементы множества M (один или много). Назовем эту операцию умножением и будем употреблять для нее обычную мультипликативную запись, тогда равенство

$$ab = c \tag{3.1}$$

будет иметь тот смысл, что для пары элементов a, b из M произведение определено и что одним из значений этого произведения служит элемент c [45].

Понятие бинарной алгебраической операции, рассматриваемое в этом широком смысле, равносильно понятию заданного в множестве M тернарного отношения (см. § 2.4). Действительно, если в M задана бинарная операция, то введем в M тернарное отношение R, полагая, что $R(a, b, c)$ имеет место тогда и только тогда, когда имеет место равенство (3.1). Наоборот, если в M задано тернарное отношение R, то будем считать, что равенство (3.1) имеет место тогда и только тогда, когда справедливо $R(a, b, c)$. В дальнейшем бинарная алгебраическая операция будет пониматься в более узком смысле — как умножение, которое должно быть определено для любой упорядоченной пары элементов и должно быть однозначным. Всякое непустое множество, в котором задана алгебраическая операция этого типа, называется *группоидом*. Более узким понятием будет понятие *полугруппы*, т. е. группоида, в котором выполняется закон ассоциативности для любых элементов a, b и c:

$$(ab)\,c = a\,(bc). \tag{3.2}$$

Это равенство придает однозначный смысл произведению abc любых трех элементов полугруппы. Отсюда следует, что при всех n произведение $a_1a_2\ldots a_n$ любых n элементов, взятых в указанном порядке, также будет однозначно определенным элементом полугруппы. Еще более узким является понятие *группы*. Группой называется полугруппа, в которой выполнимы обратные операции, т. е. для любых элементов a и b каждое из уравнений

$$ax = b; \quad ya = b \tag{3.3}$$

обладает решением, притом единственным (рис. 3.2, a). Единственность решений каждого из уравнений (3.3) позволяет производить в группе левосторонние и правосторонние сокращения: если $ab_1 = ab_2$ или $b_1a = b_2a$, то $b_1 = b_2$. Решения x и y уравнений (3.3) в случае произвольной группы не обязаны совпадать, так как алгебраическая операция не предполагается нами коммутативной, т. е. произведение может зависеть от порядка сомножителей.

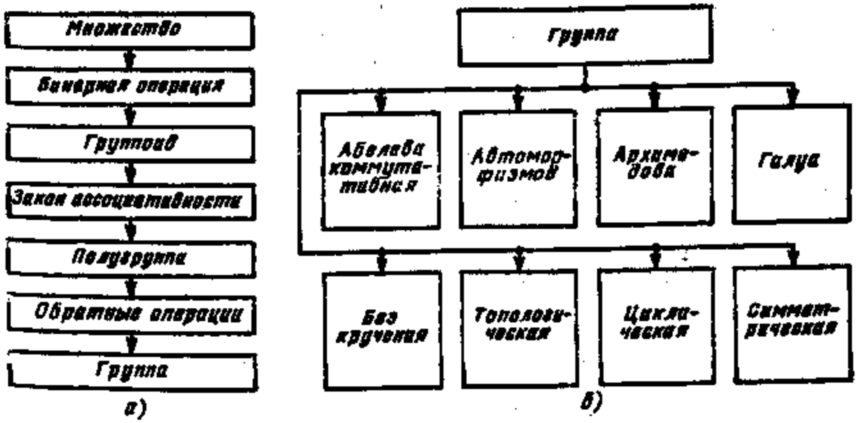


Рис. 3.2. Пояснение к определению группы (*а*) и классификация группы (*б*)

Группа (или полугруппа, или группоид), для любых двух элементов a, b которой выполняется закон коммутативности

$$ab = ba,$$

называется *коммутативной* или *абелевой*.

Теорема 3.1. Всякая группа G обладает однозначно определенным элементом e, удовлетворяющим условию

$$ae = ea = a$$

для всех элементов $a \in G$. Действительно, из определения группы следует существование в G для любого элемента a такого элемента e'_a, что $ae'_a = a$, причем этот элемент e'_a однозначно определен. Если b — любой другой элемент группы G, а y — элемент группы, удовлетворяющей равенству $ya = b$, то в соответствии с законом ассоциативности

$$be'_e = (ya)\, e'_a = y\,(ae'_a) = ya = b,$$

откуда $e'_b = e'_a$. Элемент e'_a не зависит, следовательно, от элемента a. Обозначим его через e'. Таким образом,

$$a'_e = a \quad \text{для всех } a \in G. \tag{3.4}$$

Аналогично доказывается существование и единственность такого элемента e', что

$$e''a = a \quad \text{для всех } a \in G. \tag{3.5}$$

Применив, однако, равенства (3.4) и (3.5) к произведению $e''e'$, получим $e''e' = e''$ и $e''e' = e'$, откуда $e'' = e'$. Тем самым теорема 3.1 доказана. Элемент e, существование и единственность которого утверждаются в этой теореме, называется *единицей группы* G и обычно обозначается символом 1.

Для всякого элемента a группы G существует такой однозначно определенный элемент a^{-1}, что

$$aa^{-1} = a^{-1}a = 1. \tag{3.5'}$$

Действительно, из определения группы вытекает существование таких однозначно определенных элементов a и a', что

$$aa' = 1; \quad a'a = 1.$$

Однако применяя закон ассоциативности, получаем

$$a''aa' = a''(aa') = a'' \cdot 1 = a'';$$
$$a''aa' = (a''a)\, a' = 1a' = a',$$

откуда $a'' = a'$. Вводим обозначение, полагая $a'' = a' = a^{-1}$. Элемент a^{-1} называется ***обратным элементом*** для a. Очевидно, что обратным элементом для a^{-1} служит сам элемент a и что $1^{-1} = 1$. Легко видеть, что для любых элементов $a_1, a_2, \ldots, a_n$

$$(a_1a_2 \ldots a_{n-1}a_n)^{-1} = a_n^{-1}a_{n-1}^{-1} \ldots a_2^{-1}a_1^{-1}.$$

Докажем теорему, которая облегчает проверку того, что данная полугруппа является группой. Полугруппа G тогда и только тогда будет группой, когда в G существует по меньшей мере одна правая единица, обладающая свойством

$$ae = a \quad \text{для всех } a \in G, \tag{3.5''}$$

причем это e можно выбрать так, что для всякого $a \in G$ существует по меньшей мере один правый обратный элемент a^{-1}, удовлетворяющий условию

$$aa^{-1} = e. \tag{3.5'''}$$

В одну сторону эта теорема доказана выше. Пусть теперь дана полугруппа G, удовлетворяющая условиям теоремы. Покажем, что элемент e будет и левой единицей для G. Если $a \in G$ и a^{-1} — один из его правых обратных элементов, то

$$eaa^{-1} = ee = e = aa^{-1}.$$

Умножая обе части этого равенства справа на один из элементов, правых обратных для a^{-1}, и используя однозначность произведения в полугруппе, получаем

$$eae = ae,$$

откуда $ea = a$, что и требовалось доказать. Если теперь e' — любая правая единица для G, e'' — любая левая единица, то, как и в доказательстве соотношений (3.4) и (3.5), получим, что $e'' = e'$, т. е. докажем существование и единственность в G единицы e.

Пусть $a \in G$ и a^{-1} — один из правых обратных элементов для a. Умножив равенство $aa^{-1} = e$ слева на a^{-1}, получим

$$a^{-1}aa^{-1} = a^{-1}.$$

Умножив это последнее равенство справа на один из правых обратных элементов для a^{-1}, придем к равенству

$$a^{-1}ae = e,$$

откуда $a^{-1}a = e$. Элемент a^{-1} оказался и левым обратным для a.

Теперь, так же как при доказательстве (3.5′), легко проверяется, что любой левый обратный элемент для a равен любому правому обратному элементу. Отсюда следует существование в G для всякого элемента a однозначно определенного обратного элемента a^{-1}. Для завершения доказательства этой теоремы укажем, что решениями уравнений (3.3) служат соответственно элементы

$$x = a^{-1}b \text{ и } y = ba^{-1}.$$

Единственность этих решений следует из того, что если, например,

$$ax_1 = ax_2,$$

то, умножая это равенство слева на a^{-1}, получаем $x_1 = x_2$. Иногда, в частности при изучении абелевых групп, используется не мультипликативная, а аддитивная запись: групповая операция называется сложением, сумма записывается через $a+b$, единица группы называется нулем, а вместо обратного элемента говорят о противоположном элементе и обозначают его через $-a$. При аддитивной записи для абелевых групп обратная операция (будучи единственной) называется вычитанием. Решение уравнения

$$a + x = b$$

называется разностью и записывается в виде $b-a$. Ясно, что

$$b - a = b + (-a),$$

и поэтому

$$b - (a_1 + a_2) = b - a_1 - a_2.$$

Многочисленные важные примеры абелевых групп дают нам обычные операции над числами. Так, беря все целые числа (положительные, нуль и отрицательные) и рассматривая в этом множестве операцию сложения, получаем абелеву группу, называемую аддитивной группой целых чисел.

Абелеву группу по сложению составляют также все рациональные числа — это аддитивная группа рациональных чисел.

Можно говорить об аддитивных группах всех действительных и всех комплексных чисел.

Существует много разновидностей групп, основные из которых представлены на рис. 3.2, *б*.

Если иметь в виду лишь натуральные числа, то по сложению они составляют полугруппу, называемую аддитивной полугруппой натуральных чисел, но не группу, так как вычитание здесь не всегда выполнено. При составлении из чисел групп по умножению следует иметь в виду, что ни в одну из них не должен входить нуль, так как деление на нуль невозможно. Мультипликативную группу составляют все отличные от нуля рациональные числа, так же как строго положительные рациональные числа. Полугруппами (но не группами) по умножению будут системы всех целых чисел, всех целых неотрицательных чисел и всех натуральных чисел.

К примеру, конечной абелевой группой является мультипликативная группа корней n-й степени из единицы. Порядком конечной группы называется число ее элементов. Порядок указанной группы равен n.

Рассмотрим кратко некоммутативные группы и полугруппы. Преобразованием множества M называется любое отображение множества в себя (т. е. в некоторое его подмножество). Частным случаем преобразования является подстановка, т. е. взаимно однозначное отображение множества M на себя. В теории множества (см. гл. 2) вводится понятие умножения отображений, означающее их последовательное выполнение, и доказывается ассоциативность этой операции. Относительно операции последовательного выполнения можно сделать заключение, что все преобразования данного множества M составляют полугруппу; она называется симметрической полугруппой на множестве M.

Последовательное выполнение двух подстановок множества M снова будет подстановкой. Поэтому можно говорить о полугруппе подстановок на M. Тождественная подстановка, оставляющая на месте каждый элемент из M, является единицей в этой полугруппе. Если x — произвольная подстановка, переводящая всякий элемент a из M в элемент ax, то обратное преобразование ax в a для всех $a \in M$ также будет подстановкой: она служит для x обратной подстановкой. Таким образом, по операции последовательного выполнения все подстановки данного множества M составляют группу; она называется симметрической группой на множестве M.

Если множество M конечно и состоит из n элементов, то симметрическая группа на M, называемая симметрической группой n-й степени и обозначаемая S_n, будет конечной и имеет порядок $n!$ Конечной будет и симметрическая полугруппа на конечном множестве M.

Другим примером некоммутативной группы является совокупность всех невырожденных квадратных матриц порядка $n (n \geqslant 2)$ с действительными элементами, рассматриваемая относительно операции умножения матриц. Целесообразно применить теорию алгебраических систем, использующую понятия носителя и сигнатуры.

Сигнатура — это термин, принятый в теории алгебраических моделей, которые рассматриваются как пара: *носитель* модели (объекты, элементы множества) и *сигнатура* (операции) $<M, G>$.

Сигнатура любой алгебры должна быть полной, независимой и непротиворечивой. Сигнатура является полной, если любая другая формула может быть представлена в виде пропозициональной формы с помощью ее элементов.

Сигнатура называется независимой, если не найдется в ней элемента, выводимого с помощью правил вывода из других элементов сигнатуры.

Сигнатура непротиворечива, если не найдется формула F, которая имеет место одновременно с формулой $\sim F$.

При данном подходе *алгеброй Буля* называется алгебра вида

$$A_Б = \langle M, \ \cup, \ \cap, \ - \rangle,$$

элементы носителя которой принимают два значения: 0 или 1, а сигнатура $\cup$, $\cap$ обладает свойствами операций дистрибутивных решеток с дополнениями.

Пример 3.1. Можно определить и другие двузначные алгебры:

Вебба (Никода) [30]: $A = \langle M, \circ \rangle$;

Шеффера: $A = \langle M, 1 \rangle$;

импликативную: $A = \langle M, \to, - \rangle$;

компликативную: $A = \langle M, \rightleftharpoons, 1 \rangle$;

Жегалкина: $A = \langle M, \wedge, +, 1 \rangle$.

Действия этих операций сведены в табл. 3.1.

Константа 1 необходима для образования полного набора операции (базиса) в соответствии с теоремой Поста — Яблонского, которая определяет, что, для того чтобы система операций была полной, она должна содержать хотя бы одну функцию, не сохраняющую константу 1, константу 0, несамодвойственную, нелинейную и немонотонную.

Поэтому набор операций в алгебре Жегалкина кроме конъюнкции и сложения по mod 2 должен иметь константу 1.

Если переменные принимают значения не только 0 и 1, но и другие, например 2, ..., k—1, то получаем k-значную логику, в которой определены функции, отображающие множество 0, 1, ..., k—1 в себя, заданные таблицами истинности и идентифицируемые определенными связками.

Наиболее изученными являются следующие алгебры, задающие k-значную логику [30]:

Поста: $A = \langle M, >, \sim \rangle$;

Россера — Тьюкетта: $A = \langle M, \wedge, \vee, j_i, i \rangle$, $0 < i < k-1$;

Вебба: $A = \langle M, \circ \rangle$,

где действия операций этих алгебр и их названия представляют собой следующее:

Таблица 3.1

x_1	x_2	$x_1 \circ x_2$	$x_1 x_2$	$x_1 \to x_2$	$x_1 \rightleftharpoons x_2$	$x_1 + x_2$
0	0	1	1	1	0	0
0	1	0	1	0	0	1
1	0	0	1	1	1	1
1	1	0	0	1	0	0

дизъюнкцию: $x_a \vee x_b = \max(x_a, x_b)$;

цикл: $x = x + 1 (\mathrm{mod}_k)$;

конъюнкцию: $x_a \wedge x_b = \min(x_a, x_b)$;

характеристические функции:

$$j_i(k) = \begin{cases} k-1, & x = 1; \\ 0, & x \neq i, \end{cases}$$

функцию Вебба:

$$x_a \circ x_b = \max(x_a, x_b) + 1 (\mathrm{mod}_k).$$

Пример 3.2. Проиллюстрируем понятие группы на примере группы подстановок, содержащей шесть элементов.

Подстановкой n-й степени называется взаимно однозначное отображение множества из n элементов на себя.

Рассмотрим три элемента x_1, x_2, x_3. Существуют шесть перестановок из трех элементов:

$$x_1x_2x_3; \ x_2x_3x_1; \ x_1x_3x_2; \ x_3x_1x_2; \ x_2x_1x_3; \ x_3x_2x_1.$$

Запишем для перестановки из трех элементов друг под другом:

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ x_2 & x_1 & x_3 \end{pmatrix}.$$

Эта запись означает, что x_1 переходит в x_2, x_2 — в x_1, x_3 — в x_3 посредством этой подстановки.

Ясно, что число подстановок равно числу перестановок. Введем следующие обозначения для шести возможных подстановок:

$$a = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_1 & x_2 & x_3 \end{pmatrix}; \qquad b = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_1 & x_3 & x_2 \end{pmatrix};$$

$$c = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_3 & x_1 & x_2 \end{pmatrix}; \qquad d = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_2 & x_3 & x_1 \end{pmatrix};$$

$$e = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_2 & x_1 & x_3 \end{pmatrix}; \qquad f = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_3 & x_2 & x_1 \end{pmatrix}.$$

Введем операцию умножения «·» над подстановками. Произведением подстановок называется подстановка, получаемая в результате последовательного выполнения сначала первой (написанной слева), затем второй из перемножаемых подстановок. Например,

$$cb = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_3 & x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 \\ x_1 & x_3 & x_2 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_2 & x_1 & x_3 \end{pmatrix} = e.$$

В рассматриваемой алгебре $\langle M, \cdot \rangle$ выполняется ассоциативность, но отсутствует коммутативность.

3.2. КОЛЬЦА, ТЕЛА, ПОЛЯ

Вторым важнейшим алгебраическим понятием наряду с группой является понятие кольца.

Кольцом называется множество R, в котором заданы (рис. 3.3) две бинарные алгебраические операции — сложение и умножение, причем по сложению это должна быть абелева группа — аддитивная группа кольца R, а умножение должно быть связано со сложением законами дистрибутивности:

$$a(b+c)=ab+bc; \quad (b+c)a=ba+ca. \tag{3.6}$$

Заметим, что операция умножения в общем случае не накладывает никаких ограничений, т. е. кольцо R по умножению является лишь группоидом — это мультипликативный группоид кольца R. Если умножение в кольце ассоциативно, то называем кольцо ассоциативным и говорим о его мультипликативной полугруппе; если же умножение в кольце и ассоциативно, и коммутативно, то кольцо называется ассоциативно-коммутативным.

Во всяком кольце законы дистрибутивности выполняются и для разности, т. е.

$$a(b-c)=ab-ac; \quad (b-c)a=ba-ca. \tag{3.7}$$

Действительно,

$$c+(b-c)=b.$$

Умножая обе части этого равенства слева на a, а затем применяя в левой части равенства первый из законов дистрибутивности (3.6), получаем

$$ac+a(b-c)=ab,$$

откуда вытекает первое равенство (3.7). Можно утверждать, что всякая абелева группа G служит аддитивной группой некоторого кольца. Для доказательства достаточно предположить, что групповая операция группы G записывается аддитивно, а затем ввести в G нулевое умножение, т. е. положить

$$ab=0$$

для любых a и b из G. Очевидно, что законы дистрибутивности (3.6) выполняются. Это нулевое кольцо с аддитивной группой G будет ассоциативно-коммутативным.

В теории колец нулевые кольца играют такую же роль, какую в теории групп играют абелевы группы.

В качестве первого примера ненулевого ассоциативно-коммутационного кольца служит кольцо целых чисел. В качестве примера ассоциативного, но не коммутативного кольца приведем кольцо квадратных матриц порядка n (где $n\geqslant 2$) с действительными элементами; операциями в этом кольце служат сложение и умножение матриц.

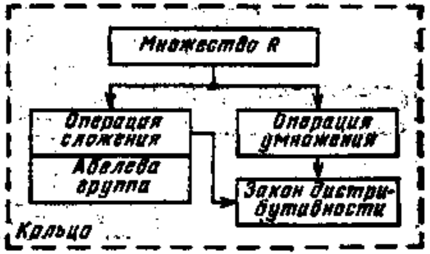


Рис. 3.3. Пояснение к определению кольца

Следующим примером является неассоциативное кольцо векторов трехмерного евклидова пространства, причем операциями служат стандартные для аналитической геометрии сложение и векторное умножение векторов. Это умножение не будет ни ассоциативным, ни коммутативным, но со сложением связано законами дистрибутивности. Можно также убедиться, что в рассмотренном нами кольце векторов для любых векторов a, b, c выполняется равенство:

$$a^2=0 \tag{3.8}$$

и тождество Якоби

$$(ab)c+(bc)a+(ca)b=0. \tag{3.9}$$

Заметим, что из (3.8) вытекает закон антикоммутативности

$$ba=-ab.$$

Действительно, так как

$$a^2=b^2=(a+b)^2=0,$$

то

$$0=(a+b)^2=a^2+ab+ba+b^2=ab+ba.$$

Всякое кольцо, удовлетворяющее условиям (3.8) и (3.9), называется лиевым кольцом. Лиевы кольца составляют важный класс в теории неассоциативных колец, которому принадлежат все нулевые кольца (рис. 3.4).

Связь ассоциативных и лиевых колец устанавливает следующее положение:

если R — произвольное ассоциативное кольцо, то, сохраняя аддитивную группу этого кольца, а операцию умножения ab заменяя операцией коммутирования

$$a\circ b=ab-ba,$$

получаем лиево кольцо $R^{(-)}$.

Действительно, проверим справедливость законов дистрибутивности. Проверяя первый закон (3.7), получаем

$$a\circ(b+c)=a(b+c)-(b+c)a=ab+ac-ba-ca=$$
$$=(ab-ba)+(ac-ca)=a\circ b+a\circ c.$$

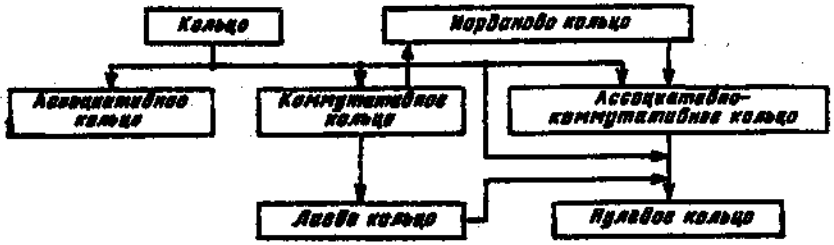


Рис. 3.4. Классификация колец

Поэтому множество R, рассматриваемое с операциями сложения и коммутирования, оказывается кольцом. Осталось проверить (3.8) и (3.9). Действительно,

$$a \circ a = aa - aa = 0;$$

$$(a \circ b) \circ c + (b \circ c) \circ a + (c \circ a) \circ b + = (ab - ba)c - c(ab - ba) +$$
$$+ (bc - cb)a - a(bc - cb) + (ca - ac)b - b(ca - ac) = 0.$$

Кольцо $R^{(-)}$ является лиевым. Если в ассоциативном кольце R сохранить его аддитивную группу, а операцию умножения ab заменить операцией симметрирования

$$a \cdot b = ab + ba,$$

то будет получено кольцо $R^{(+)}$, в котором для любых элементов a, b выполняются равенства

$$a \cdot b = b \cdot a; \tag{3.10}$$

$$[(a \cdot a) \cdot b] \cdot a = (a \cdot a) \cdot (b \cdot a). \tag{3.11}$$

Действительно, проверим первый закон дистрибутивности (3.6):

$$a \cdot (b + c) = a(b + c) + (b + c)a = ab + ac + ba + ca = (ab + ba) +$$
$$+ (ac + ca) = a \cdot b + a \cdot c.$$

Проверим справедливость равенств (3.10) и (3.11):

$$a \cdot b = ab + ba = ba + ab = b \cdot a;$$

$$[(a \cdot a) \cdot b] - a = [(aa + aa)b + b(aa + aa)]a + a[(aa + aa)b +$$
$$+ b(aa + aa)] = aaba + aaba + baaa + baaa + aaab + aaab +$$
$$+ abaa + abaa = (aa + aa)(ba + ab) + (ba + ab)(aa + aa) = (a \cdot a) \cdot (b \cdot a).$$

Всякое кольцо, удовлетворяющее условиям (3.10) и (3.11), называется йордановым.

В общем случае оно неассоциативное; к йордановым кольцам принадлежат все ассоциативно-коммутативные кольца.

В кольце R любое произведение, в котором хотя бы один из сомножителей равен нулю, само равно нулю. Иначе

$$a \cdot 0 = 0 \cdot a = 0 \tag{3.12}$$

для всех элементов кольца R.

В самом деле, если x — произвольный элемент кольца R, то с учетом (3.7)

$$a \cdot 0 = a(x - x) = ax - ax = 0.$$

Если a и b — любые элементы произвольного кольца R, то

$$(-a)b = a(-b) = -ab; \tag{3.13}$$

$$(-a)(-b) = ab. \tag{3.14}$$

Действительно,

$$ab + (-a)b = [a + (-a)]b = 0 \cdot b = 0.$$

Аналогичным образом проверяется вторая половина равенства (3.13). С другой стороны, используя (3.13), получаем

$$(-a)(-b) = -[a(-b)] = -(-ab) = ab,$$

что и доказывает (3.14).

Обратное (3.12) утверждение, справедливое в случае числовых колец, в общем случае не имеет места; существуют кольца, обладающие делителями нуля, т. е. такими отличными от нуля элементами a и b, произведение которых равно нулю:

$$ab = 0.$$

Примерами таких колец помимо нулевого кольца являются кольца матриц, в частности кольца действительных квадратных матриц. Если R — произвольное кольцо, то можно рассмотреть всевозможные квадратные матрицы порядка n с элементами из R. Определим для них обычным способом сложение и умножение; нетрудно проверить, что кольца будут ассоциативные, если ассоциативно исходное кольцо R. Нулем этого кольца служит нулевая матрица, составленная из нулей. Построенное кольцо называется полным кольцом матриц порядка n над кольцом R и обозначается через R_n.

Если $n \geqslant 2$, а кольцо R состоит не только из одного нуля, то полное кольцо матриц R_n обладает делителями нуля. Действительно, если a — отличный от нуля элемент из R, то матрицы

$$\begin{pmatrix} a & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 0 \\ 0 & \ldots & 0 & a \end{pmatrix}$$

отличны от нулевой матрицы и их произведение равно нулю.

Примерами колец с делителями нуля являются также полные кольца функций. Пусть даны произвольное множество M и произвольное кольцо R. Рассмотрим множество всевозможных функций на M со значениями в R, т. е. всевозможных отображений f множества M в кольцо R; образ элемента $x \in M$ при отображении f будет обозначаться через $f(x)$. Это множество функций превращается в кольцо, если сумма и произведение функций будут определены, как обычно, равенствами

$$(f+g)(x)=f(x)+g(x);$$

$$(fg)(x)=f(x)g(x),$$

т. е. при помощи сложения и умножения значений заданных функций для всех x из M. Нетрудно проверить, что все требования, входящие в определение кольца, выполняются и что полученное кольцо будет ассоциативным и коммутативным, если только асоциативно и соответственно коммутативно исходное кольцо R.

Построенное кольцо называется *полным кольцом функций на множестве* M со значениями в кольце R. Если M — множество точек числовой прямой, а R — совокупность действительных чисел, то наше кольцо будет обычным кольцом всех действительных функций действительного переменного.

Всякое полное кольцо функций на множестве M, содержащем не менее двух элементов, со значениями в кольце R, которое состоит не из одного нуля, обладает делителями нуля.

Действительно, роль нуля в этом кольце играет нулевая функция, которая тождественно (т. е. для всех x из M) равна нулю. Если разобьем множество M на два непустых непересекающихся подмножества A и B, то существуют, очевидно, ненулевые функции f и g, такие, что f принимает нулевые значения на A, а g — на B. Очевидно, что произведение fg будет нулевой функцией.

Ассоциативно-коммутационное кольцо без делителей нуля называется *областью целостности*; таковы, в частности, всевозможные числовые кольца. Интересные примеры областей целостности можно получить, используя следующую конструкцию.

Если R — произвольное ассоциативно-коммутационное кольцо, то можно рассмотреть всевозможные многочлены

$$a_0+a_1x+a_2x^2+\ . \ . \ . \ +a_nx^n,\ n \geqslant 0,$$

относительно неизвестного x с коэффициентами a_0, a_1, ..., a_n из R; если $a_n \neq 0$, то n будет степенью этого многочлена. Определяя сложение и умножение так же, как это делается в курсе высшей алгебры, получаем кольцо, называемое *кольцом многочленов* $R[x]$ от неизвестного x над кольцом R; это кольцо также будет ассоциативно-коммутативным. Нулем кольца многочленов служит многочлен, все коэффициенты которого равны 0.

Аналогично определяется кольцо многочленов $R[x_1, x_2, \ldots, x_n]$ от любого конечного числа неизвестных; оно будет при этом просто кольцом многочленов от одного неизвестного x_n над кольцом $R[x_1, \ldots, x_{n-1}]$. Можно говорить о кольце многочленов над R от любого бесконечного множества неизвестных, считая при этом, что отдельный многочлен зависит лишь от некоторого конечного числа неизвестных.

Если R — область целостности, то любое кольцо многочленов над R также будет областью целостности.

Справедливость этого утверждения для кольца многочленов $R[x]$ от одного неизвестного вытекает из того, что если многочлены f и g отличны от нуля, а в R нет делителей нуля, то степень произведения fg равна сумме степеней сомножителей, т. е. это произведение также отлично от нуля.

Переход к случаю любого конечного числа неизвестных достигается простой индукцией, а в случае колец многочленов от бесконечного множества неизвестных достаточно учесть, что всякий многочлен записывается через конечное число неизвестных.

Над произвольным ассоциативно-коммутативным кольцом R можно рассматривать не только многочлены, но и степенные ряды

$$a_0+a_1x+a_2x^2+\ . \ . \ . \ +a_kx^k+\ . \ . \ . \ =\sum_{k=0}^{\infty} a_kx^k \qquad (3.15)$$

от переменной x.

Определение операций переносится с многочленов на степенные ряды без всяких затруднений:

$$\sum_{k=0}^{\infty} a_kx^k+\sum_{k=0}^{\infty} b_kx^k=\sum_{k=0}^{\infty} (a_k+b_k)\,x^k;$$

$$\sum_{k=0}^{\infty} a_kx^k \sum_{l=0}^{\infty} b_lx^l=\sum_{m=0}^{\infty} c_mx^m,$$

где

$$c_m=\sum_{k+l=m} a_kb_l.$$

Как и в случае многочленов, проверяется получение ассоциативно-коммутационного кольца; оно называется *кольцом степенных рядов от неизвестного* x *над кольцом* R и обозначается через $R\{x\}$.

Переход к кольцам степенных рядов над R от любого конечного числа, а затем и бесконечного множества неизвестных осуществляется так же, как и для колец многочленов, причем в случае бесконечного множества переменных целесообразно наложить требование зависимости каждого степенного ряда лишь от конечного числа неизвестных.

Если R — область целостности, то любое кольцо степенных рядов над R также будет областью целостности.

Достаточно рассмотреть случай кольца $R\{x\}$ от одного неизвестного. Если назовем нижней степенью степенного ряда (3.15) число n в том случае, когда

$$a_0 = a_1 = \ldots = a_{n-1} = 0, \quad a_n \neq 0,$$

то всякий ненулевой степенной ряд будет обладать нижней степенью. Ввиду отсутствия в R делителей нуля нижняя степень произведения двух степенных рядов равна сумме нижних степеней сомножителей. Понятие единицы, введенное в § 3.2 для группы, переносится на любой группоид или кольцо R следующим образом: это будет элемент 1, такой, что для всех элементов a из R

$$a \cdot 1 = 1 \cdot a = a.$$

Если единица в кольце R существует, то только одна. Ее может и не быть, как это имеет место для случая кольца четных чисел.

Всякое лиево кольцо, состоящее не только из одного нуля, является кольцом без единицы.

Действительно, пусть лиево кольцо L обладает единицей 1. Тогда по (3.3), полагая $a=1$, получаем $1^2=0$. Таким образом, $1=0$, откуда следует, что кольцо L состоит лишь из нуля.

Если кольцо R обладает единицей, то единицами обладают также всевозможные кольца матриц над R, кольца функций со значениями в R, а в ассоциативно-коммутативном случае — и кольца многочленов, и кольца степенных рядов над R от любого числа неизвестных.

Действительно, единицей кольца матриц служит единичная матрица, по главной диагонали которой стоит единица, а все элементы вне диагонали равны нулю.

Единицей кольца является функция, тождественно равная единице. Наконец, единицей кольца многочленов (или кольца степенных рядов) служит многочлен (или степенной ряд), все коэффициенты которого равны нулю, кроме a_0, равного единице.

Всякое кольцо является по умножению группоидом, а ассоциативное кольцо — полугруппой. Никакое кольцо, состоящее не только из нуля, не может быть по умножению группой, как вытекает из мультипликативного свойства нуля (3.12). Однако может оказаться, что все отличные от нуля элементы кольца составляют группу по умножению. Такое кольцо, которое обязательно должно быть ассоциативным, называется *телом*, а группа по умножению его отличных от нуля элементов — мультипликативной группой этого тела (рис. 3.5). Тело с коммутативным умножением называется *полем*. Основными примерами полей являются поле рациональных чисел, поле действительных чисел и поле комплексных чисел.

Из определения тела следует, что тело не содержит делителей нуля. Кроме того, всякое тело обладает единицей. Действительно, единица мультипликативной группы этого тела служит в соответствии с (3.12) единицей всего тела.

Наконец, во всяком теле каждое из уравнений

$$ax = b; \quad ya = b, \quad \text{где } a \neq 0, \tag{3.16}$$

обладает решением, притом единственным.

Действительно, если $b \neq 0$, то оба уравнения (3.16) обладают однозначно определенными решениями в мультипликативной группе тела, а нуль не может удовлетворять ни одному из этих уравнений. Если же $b=0$, то нуль является решением для каждого из этих уравнений и никаких других решений нет ввиду отсутствия делителей нуля.

Обратное утверждение — ассоциативное кольцо R будет телом, если в нем каждое из уравнений вида (3.16) при любом $a \neq 0$ и произвольном b обладает хотя бы одним решением. Покажем прежде всего, что в R нет делителей нуля. Если

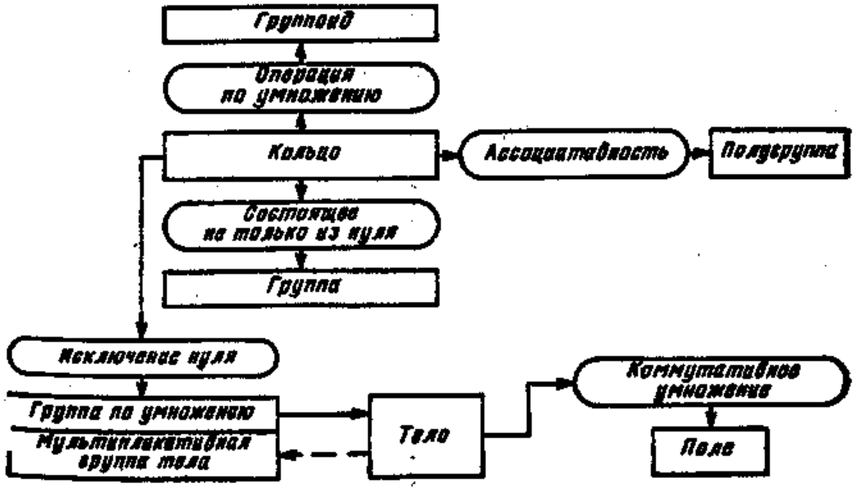


Рис. 3.5. Пояснение к понятию тела и поля

$a \neq 0$ и $b \neq 0$, но $ab = 0$, то обозначим через e одно из решений уравнения $ax = a$, а через c — одно из решений уравнения $bx = e$. Тогда

$$0 = 0 \cdot c = abc = ae = a,$$

что противоречит исходному условию, что множество отличных от нуля элементов кольца R будет мультипликативной полугруппой. Оно будет также группой, так как уравнения (3.16) обладают при $a \neq 0$ и $b \neq 0$ решениями в самом этом множестве, а единственность этих решений следует из отсутствия в R делителей нуля: если, например, $ax_1 = ax_2$ и $a \neq 0$, то $a(x_1 - x_2) = 0$, и поэтому $x_1 = x_2$. Тем самым обратное утверждение доказано.

3.3. ПОДГРУППЫ, ПОДКОЛЬЦА

Непустое подмножество A группоида G называется подгруппоидом в G, если произведение любых двух элементов из A само принадлежит A. Если G — полугруппа, то подгруппоид A естественно называть подполугруппой. Если же G — группа, то подгруппой этой группы назовем такую подполугруппу A, которая сама является группой относительно операции, определенной в группе G; для этого достаточно, чтобы подполугруппа A вместе со всяким своим элементом a содержала и его обратный элемент a^{-1} (рис. 3.6).

Не всякая подполугруппа некоторой группы будет подгруппой этой группы, к примеру аддитивная полугруппа натуральных чисел является подполугруппой, но не подгруппой аддитивной группы всех целых чисел.

Имеет смысл говорить о подгруппах любой полугруппы. К примеру, симметрическая на произвольном множестве M подгруппа является подгруппой симметрической полугруппы на этом же множестве. Подкольцом кольца R называется всякое подмножество A этого кольца, которое само является кольцом относительно операций, определенных в R. Иначе, A должно быть подгруппой аддитивной группы кольца R и подгруппоидом мультипликативного группоида этого кольца; законы дистрибутивности, будучи справедливы в R, выполняются и в A. Можно говорить о подкольце тела и поля.

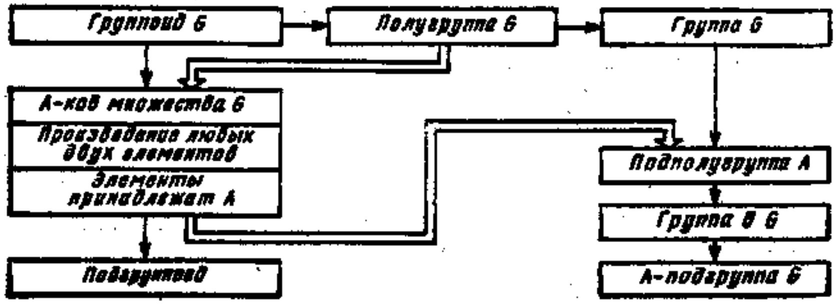


Рис. 3.6. Пояснение к определению подгруппы

Аналогично определяются подтела и подполя, при этом можно говорить о подтеле (подполе) не только тела, но и любого кольца. Так, в кольце многочленов $P[x]$ над произвольным полем многочлена P многочлены нулевой степени, к которым добавлен нуль, составляют подполе.

Подгруппой всякой группы G являются сама эта группа и единичная подгруппа E, составленная из одной единицы. Аналогично подкольцами всякого кольца R служат само это кольцо и нуль-подкольцо, составленное из одного нуля.

Для демонстрации примеров подгрупп введем понятие о степени элементов. Пусть G — произвольная полугруппа, а a — ее элемент. Ассоциативность умножения позволяет обычным путем определить положительные степени a^n элемента a, $n = 1, 2, \ldots$, при этом, как обычно,

$$a^k \cdot a^l = a^{k+l}; \qquad (3.17)$$

$$(a^k)^l = a^{kl}. \qquad (3.18)$$

Из (3.17) следует, что все положительные степени элемента a составляют в полугруппе G абелеву подполугруппу, которая называется циклической подполугруппой элемента a.

Если G — группа, то положим $a^0 = 1$, а затем введем отрицательные степени элемента a. Если n — любое натуральное число, то легко проверить справедливость равенства

$$a^n (a^{-1})^n = 1,$$

из которого следует, что

$$(a^{-1})^n = (a^n)^{-1}.$$

Элемент, равный обеим частям этого равенства, обозначим через a^{-n}. Равенства (3.17) и (3.18) остаются справедливыми для любых степеней элемента a группы G, а потому все степени элемента a, включая нулевую и отрицательные, составляют в группе G абелеву подгруппу, называемую циклической подгруппой элемента a и обозначаемую через $\{a\}$. При аддитивной записи групповой операции вместо степени элемента следует говорить о кратных этого элемента.

Заметим, что приводимые рассуждения являются как бы предпосылками для введения операции экспоненцирования и понятия экспоненциала в теории категорий (см. гл. 4). Степени элемента a группы G, имеющие разные показатели, не обязаны быть различными элементами этой группы, что вытекает хотя бы из существования конечных групп. Если все степени элемента a действительно различны, то a называется элементом бесконечного порядка, в противном случае — элементом конечного порядка. Во втором случае существуют такие целые числа k и l, что $k>l$, но

$$a^k = a^l.$$

Отсюда $a^{k-l}=1$, причем $k-l>0$, т. е. существуют равные единице степени элемента a с натуральными показателями. Наименьший среди таких показателей называется порядком элемента a.

Если a — элемент конечного порядка k, то степени

$$1=a^0,\ a,\ a^2,\ .\ .\ .\ .\ ,\ a^{n-1} \qquad (3.19)$$

будут различными элементами группы. Всякая другая степень a^k элемента a, положительная или отрицательная, равна одному из элементов (3.19). Действительно, если

$$k=nq+r,\quad 0\leqslant r\leqslant n,$$

то в соответствии с (3.16) и (3.17) и равенством $a^n=1$

$$a^k=(a^n)^q a^r=a^r.$$

Отсюда следует, что порядок n элемента конечного порядка a совпадает с порядком его циклической подгруппы $\{a\}$. Группа, все элементы которой имеют конечные порядки, необязательно ограниченные в совокупности, называется *периодической*. Группа называется *группой без кручения*, если все ее элементы, кроме 1, бесконечного порядка. В случае колец представляет интерес вопрос об аналоге циклической подгруппы, т. е. о минимальном подкольце, содержащем данный элемент. Пусть R — произвольное (необязательно ассоциативное) кольцо и a — его элемент. Всякое подкольцо кольца R, содержащее a, содержит также все возможные произведения по n множителей, равных $a(n=1,\ 2,\ 3\ldots)$; эти произведения играют роль положительных степеней элемента a. Ввиду возможной неассоциативности умножения во всяком таком произведении должны быть некоторым образом распределены скобки, причем так, чтобы каждый раз перемножались лишь два элемента кольца. Для $n=3$ таких произведений будет два — $(aa)a$ и $a(aa)$, для $n=4$ — пять и т. д.

Всякое подкольцо, содержащее a, содержит и всевозможные суммы любого конечного числа указанных произведений, взятых с любыми целыми коэффициентами. Элементы кольца R, записываемые в виде таких сумм (быть может, и неоднозначно), сами составляют, однако, подкольцо в R. Это и будет искомое подкольцо, порожденное элементом a.

Рис. 3.7. Пояснение к определению подкольца кольца R

Из сказанного следует, что если кольцо R ассоциативно, то подкольцо, порожденное элементом a, состоит из всех тех элементов кольца R, которые хотя бы одним способом записываются в виде суммы (с целыми коэффициентами) положительных степеней элемента a. Это подкольцо будет, таким образом, коммутативным.

Всякая подгруппа группы G содержит единицу этой группы, всякое подкольцо кольца R — нуль этого кольца (рис. 3.7). Пересечение любой системы подгрупп группы G (любой системы подколец кольца R) будет, следовательно, непустым.

Пересечение любой системы подгрупп группы G является подгруппой этой группы. Аналогично пересечение любой системы подколец кольца R будет подкольцом этого кольца, пересечение любой системы подгрупп некоторой полугруппы, если оно не пусто, будет подполугруппой и т. д.

Докажем первое утверждение. Пусть в группе G взята произвольная система подгрупп $A\alpha$, где α пробегает некоторое множество индексов, и пусть D — пересечение этих подгрупп. Если b и c — любые элементы из D, то их произведение bc содержится в каждой из подгрупп $A\alpha$, т. е. содержится в D. С другой стороны, для любого $b\in D$ элемент b^{-1} принадлежит каждой из подгрупп $A\alpha$, и поэтому $b^{-1}\in D$. Множество D будет, следовательно, подгруппой группы G.

Пусть G будет или группой, или кольцом, или телом, или полугруппой, или алгебраической системой какого-либо другого типа из встречавшихся нам типов. Если M — любое подмножество из G, то существует минимальная подгруппа (соответственно подкольцо, подтело и т. д.), содержащая подмножество M, — подгруппа (подкольцо, подтело), порожденная множеством M, обозначается она через $\{M\}$. Именно это будет пересечением всех подгрупп (подколец, подтел) группы (кольца, тела) G, целиком содержащих M, причем по меньшей мере одна такая подгруппа (подкольцо, подтело) существует, а именно сама G. Легко проверить, обобщая данное ранее определение циклических подгрупп, что подгруппа $\{M\}$ состоит из тех и только тех элементов группы G, которые хотя бы одним способом могут быть записаны в виде некоторого произведения степеней конечного числа элементов из M.

Аналогично, обобщая ранее полученные результаты, можно утверждать, что подкольцо, порожденное в кольце R множеством M, состоит из всех тех элементов кольца R, которые хотя бы одним способом могут быть записаны в виде суммы (с целыми коэффициентами) произведений конечного числа элементов из M; в неассоциативном случае эти произведения рассматриваются, конечно, с некоторым распределением скобок.

Если применить эти результаты, когда в группе G задана некоторая система подгрупп A_i, $i \in I$, а множество M является объединением всех A_i, придем к понятию *подгруппы, порожденной заданной системой подгрупп*; обозначается она через $\{A_i, i \in I\}$, а в частном случае — через $\{A, B\}$, если речь идет об объединении двух подгрупп A и B, и т. д. Аналогичный смысл имеет понятие *подкольца, порожденного в кольце заданной системой подколец*.

Если G — группа или кольцо и т. д., то в G существуют такие множества M (само G, например), что

$$\{M\} = G.$$

Всякое такое подмножество называется *системой образующих* для G. Если G обладает хотя бы одной конечной *системой образующих*, то говорят, что G — *группа* (или полугруппа, или кольцо) *с конечным числом образующих*. Система образующих может состоять, в частности, из одного элемента. Группы, обладающие одним образующим элементом, т. е. совпадающим с одной из своих циклических подгрупп, называются *циклическими группами*.

3.4. ИЗОМОРФИЗМЫ

Понятие изоморфизма может быть введено для каждого из рассматриваемых в этой главе типов алгебраических систем, причем оно аналогично понятию изоморфизма для частично упорядоченных множеств. В частности, изоморфные группы могут рассматриваться как тождественные, как два экземпляра одной и той же группы всякий раз, когда изучается сама групповая операция, а природа элементов, из которых группы составлены, не имеет значения.

Два группоида G и G' называются *изоморфными*, если существует такое взаимно однозначное отображение φ группоида G на группоид G', что для любых элементов a, $b \in G$

$$(ab)\varphi = a\varphi \cdot b\varphi.$$

Само отображение φ с этими свойствами называется *изоморфным отображением*. Это свойство группоида симметрично, т. е. отображение, обратное изоморфному, само будет изоморфным и *транзитивным*. Кроме того, оно является также *рефлексивным*. Чтобы убедиться в этом, достаточно рассмотреть отображение группоида на себя. Изоморфизм группоидов G и G' записывается обычно с помощью символа $\cong$:

$$G \cong G'.$$

Этот же символ будет использоваться для обозначения изоморфизма в случае других алгебраических образований. Очевидно, что при изоморфном отображении имеют место свойства, сформулированные на языке операции, которая задана в этом группоиде, такие, как ассоциативность, коммутативность, существование единицы и обратных элементов (рис. 3.8). На примере покажем, как доказываются эти положения. Пусть G — коммутативный группоид, а φ — его изоморфное отображение на группоид G'. Если a' и b' — любые элементы из G', a и b — такие элементы из G, что

$$a\varphi = a'; \quad b\varphi = b',$$

то

$$(ab)\varphi = a'b'; \quad (ba)\varphi = b'a'$$

и из равенства $ab = ba$ и однозначности отображения вытекает равенство $a'b' = b'a'$. Отсюда следует, что изоморфные образы *полугрупп, групп и абелевых групп будут соответственно полугруппами, группами и абелевыми группами.*

Два кольца называются *изоморфными*, если между ними можно установить взаимно однозначное соответствие, которое

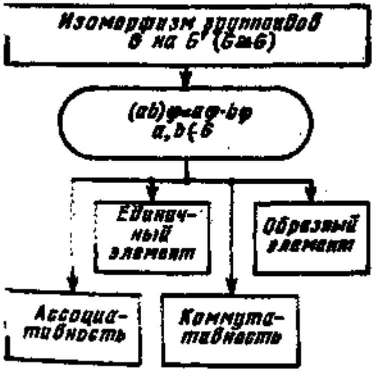


Рис. 3.8. Пояснение к понятию изоморфизма

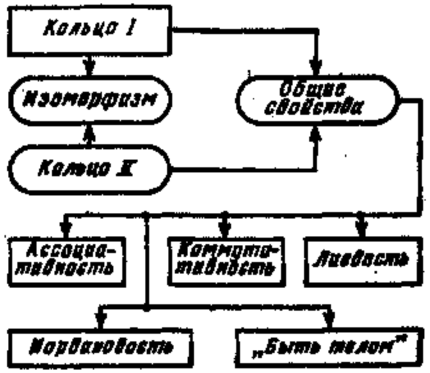


Рис. 3.9. Пояснение к сохранности свойств кольца при изоморфизме

является изоморфизмом как для аддитивных групп, так и для мультипликативных группоидов этих колец. Очевидно, что при изоморфизме сохраняются свойства кольца быть ассоциативным, коммутативным, лиевым или йордановым, а также свойство быть телом (рис. 3.9).

Приведем еще один пример изоморфизма группы: *мультипликативная группа положительных действительных чисел изоморфна аддитивной группе всех действительных чисел.*

Действительно, если поставить в соответствие каждому положительному числу его логарифм по некоторому фиксированному основанию, то получим взаимно однозначное отображение первой из указанных групп на вторую. Изоморфность этого отображения вытекает из того факта, то логарифм произведения равен сумме логарифмов сомножителей.

Рассмотрим циклические группы (рис. 3.10). Аддитивная группа целых чисел служит примером бесконечной циклической группы, так как всякое целое число кратно числу 1. Мультипликативная группа корней n-й степени из единицы служит примером конечной циклической группы порядка n, что следует из существования первообразного корня n-й степени из единицы.

Все бесконечные циклические группы изоморфны аддитивной группе целых чисел и поэтому изоморфны между собой. Все конечные циклические группы порядка n изоморфны мультипликативной группе корней n-й степени из единицы и поэтому изоморфны между собой.

Докажем справедливость первого утверждения. Если G — бесконечная циклическая группа с образующим элементом a, то соответствие $a^k \to k$ будет взаимно однозначным отображением группы G на всю аддитивную группу целых чисел. Изоморфность этого отображения вытекает из формулы (3.3).

Для доказательства второго утверждения достаточно установить соответствие между степенями (с одинаковыми показателями) образующего элемента заданной циклической группы и соответствующего первообразного корня из единицы.

Согласно этой теореме для обозрения всех подгрупп бесконечной циклической группы достаточно рассмотреть аддитивную группу целых чисел.

Все ненулевые подгруппы аддитивной группы целых чисел исчерпываются совокупностями чисел, кратных некоторому натуральному числу n.

Действительно, все целые числа, кратные натуральному числу n, составляют подгруппу, а именно циклическую подгруппу, порожденную числом n, причем при различных n эти подгруппы различны. С другой стороны, если A — любая ненулевая подгруппа аддитивной группы целых чисел, то она не

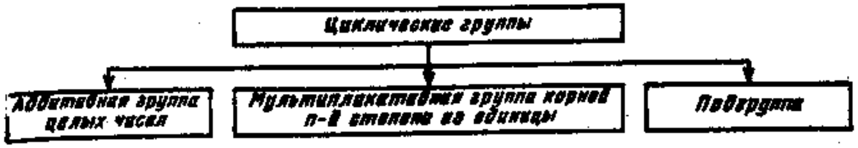


Рис. 3.10. Циклические группы

может состоять лишь из отрицательных чисел, так как со всяким числом должна содержать и число, ему противоположное. Пусть n — наименьшее натуральное число, содержащееся в подгруппе A. Если a — любое число из A, то пусть

$$a = qn + r, \quad 0 \leqslant r < n,$$

тогда

$$r = a - qn \in A,$$

а поэтому ввиду выбора числа n $r=0$, где $a=qn$, что и требовалось доказать.

Таким образом, *всякая бесконечная подгруппа циклической группы* сама является циклической. Это же справедливо для конечных циклических групп.

Группоид G изоморфно вкладывается в группоид G', если существует изоморфное отображение группоида G на некоторый подгруппоид группоида G'. Это понятие переносится, конечно, на случай колец.

Всякое кольцо R изоморфно вкладывается в полное кольцо матриц R_n. Действительно матрицы, имеющие на главной диагонали один и тот же элемент a, а вне этой диагонали нули, составляют в R_n подкольцо, изоморфное кольцу R.

Всякое кольцо R изоморфно вкладывается в полное кольцо функций на данном множестве M со значениями в R.

Действительно, функции, принимающие для всех x из M одно и то же значение $a \in R$, составляют в кольце подкольцо, изоморфное кольцу R.

Для произвольного ассоциативно-коммутативного кольца R кольцо многочленов $R[x]$ изоморфно вкладывается в кольцо степенных рядов $R\{x\}$.

Действительно, ряды, имеющие не более конечного числа отличных от нуля коэффициентов, составляют в $R\{x\}$ подкольцо, изоморфное кольцу $R[x]$.

Всякое кольцо R изоморфно вкладывается в кольцо с единицей. Если при этом кольцо R ассоциативно или коммутативно, то его можно вложить соответственно в ассоциативное и коммутативное кольцо с единицей.

Рассмотрим множество всевозможных упорядоченных пар (a, k), где $a \in R$, k — целое число. Определим сумму и произведение таких пар равенствами

$$(a, k)+(b, l)=(a+b, k+l); \tag{3.20}$$

$$(a, k)(b, l)=(ab+la+kb, kl). \tag{3.21}$$

По сложению получается абелева группа, так как

$$[(a, b)+(b, l)](c, m)=(a+b, k+l)(c, m)=((a+b)c+$$
$$+m(a+b)+(k+l)c, (k+l)m)=(ac+ma+kc+bc+mb+lc,$$
$$km+lm)=(al+ma+kl, km)+(bc+mb+lc, lm)=(a, k)(c, m)+$$
$$+(b, l)(c, m).$$

Аналогичным образом проверяется второй закон дистрибутивности. Тем самым построено кольцо. Далее в соответствии с (3.21) имеем

$$(a, k)(0, 1)=(0, 1)(a, k)=(a, k),$$

т. е. пара (0, 1) служит единицей этого кольца. Наконец, в соответствии с (3.20) и (3.21)

$$(a, 0)+(b, 0)=(a+b, 0);$$
$$(a, 0)(b, 0)=(ab, 0),$$

т. е. пары вида $(a, 0)$ составляют в нашем кольце пар подкольцо, изоморфное кольцу R. Тем самым доказана первая часть теоремы. Вторая часть непосредственно следует из (3.21).

Заметим, что построенное кольцо не является единственным (или минимальным) кольцом с единицей, содержащим в смысле изоморфного вложения заданное кольцо R, так как уже само кольцо R могло обладать единицей.

Всякий группоид изоморфно вкладывается в мультипликативный группоид некоторого кольца, причем ассоциативный или коммутативный группоид может быть вложен в кольцо, обладающее таким же свойством.

Для доказательства рассмотрим всевозможные суммы вида

$$\sum_{a \in G} k_a a, \tag{3.22}$$

где a пробегает все элементы данного группоида G, а коэффициенты k_a являются целыми числами, причем *не более конечного числа коэффициентов, отличного от нуля.* Определим

следующим образом сложение и умножение сумм вида (3.22):

$$\sum_{a \in G} k_a a+\sum_{a \in G} l_a a=\sum_{a \in G}(k_a+l_a)a; \tag{3.23}$$

$$\sum_{a \in G} k_a a \sum_{a \in G} l_a a=\sum_{c \in G} m_c c, \tag{3.24}$$

где m_c является суммой всех отличных от нуля произведений $k_a l_b$ для таких a и b, что $ab=c$. Очевидно, что правые части равенств (3.23) и (3.24) являются суммами вида (3.22). К примеру, равенство (3.24) имеет тот смысл, что конечные суммы, входящие в качестве множителей в левую часть этого равенства, должны перемножаться почленно, далее

$$k_a a \cdot l_b b=(k_a l_a)(ab),$$

причем произведение следует понимать в смысле операции, заданной в группоиде G, а затем выполняется произведение подобных членов.

По сложению получаем абелеву группу. Проверка законов дистрибутивности, а также доказательство того, что из ассоциативности или коммутативности умножения в группоиде G следует это же для умножения сумм вида (3.22), делаются способом, аналогичным вышеиспользованному.

Таким образом, всевозможные суммы вида (3.22) составляют кольцо относительно операций, определенных равенствами (3.23) и (3.24). Суммы вида (3.22), у которых $k_a=1$ для какого-нибудь одного элемента из G, а все остальные коэффициенты равны нулю, составляют, как это следует из (3.24), подгруппоид мультипликативного группоида этого кольца, изоморфный заданному группоиду G.

Построенное кольцо называется *целочисленным группоидным кольцом* группоида G. Если группоид является полугруппой или группой, то говорят соответственно о *целочисленном групповом кольце. Всякая группа G изоморфно вкладывается в симметрическую группу на некотором множестве M. В качестве M можно взять при этом множество элементов самой группы G.*

Действительно, поставим в соответствие каждому элементу a группы G преобразование этой группы, которое переводит любой элемент x из G в элемент xa. Так как из $x \neq y$ следует $xa \neq ya$ и, кроме того, для любого x из G имеет место равенство

$$(xa^{-1})a=x,$$

то преобразование $x \to xa$ отображает G на себя взаимно однозначно, т. е. является подстановкой. Кроме того, если $a \neq b$, то и подстановки, соответствующие этим элементам, будут различными, так как, к примеру, $1 \cdot a \neq 1 \cdot b$. Наконец, равенство

$$(xa)\, b = x\,(ab)$$

показывает, что подстановка, соответствующая произведению ab, совпадает с результатом последовательного выполнения подстановок, соответствующих элементам a и b.

Всякая полугруппа G изоморфно вкладывается в симметрическую полугруппу на некотором множестве M.

Для доказательства изоморфно вложим подгруппу G в мультипликативную полугруппу некоторого ассоциативного кольца R, изоморфно вложим затем это кольцо в ассоциативное кольцо $\overline{R}$ с единицей и обозначим через $\overline{G}$ мультипликативную полугруппу этого последнего кольца. Так как в $\overline{G}$ существует такой элемент 1, что из $a \neq b$ следует $1 \cdot a \neq 1 \cdot b$, то, повторяя, что полугруппа $\overline{G}$ изоморфно вкладывается в симметричную полугруппу на самом множестве $\overline{G}$, достигаем искомого вложения для заданной полугруппы G.

Заметим, что возможность вложения любого группоида (и, в частности, любой группы) в группоид (полугруппу) с единицей можно было бы доказать без перехода к кольцам.

3.5. ВЛОЖЕНИЕ ПОЛУГРУПП В ГРУППЫ И КОЛЕЦ В ТЕЛА

Не всякая полугруппа G может быть изоморфно вложена в какую-либо группу. Необходимым условием для этого является выполнимость в G *закона сокращения* (рис. 3.11).

Из $ac = bc$, а также из $ca = cb$ следует $a = b$. Легко построить пример полугруппы, не удовлетворяющей этому условию.

Аналогично не всякое ассоциативное кольцо R может быть изоморфно вложено в какое-либо тело, для этого необходимо отсутствие в R делителей нуля (рис. 3.12).

Данные необходимые условия не обязательно являются достаточными в общем случае.

Необходимые и достаточные условия для вложения полугруппы в группу выражаются в виде бесконечного множества требований вида «из данной системы равенств следует такое равенство» и не могут быть записаны в виде конечного числа таких требований.

Существуют ассоциативные кольца без делителей нуля, которые не могут быть вложены в тело.

Докажем, что *приведенные выше необходимые условия для вложимости полугруппы в группу и ассоциативного кольца в тело будут в коммутативном случае достаточными.* Прежде всего, докажем следующее утверждение:

Рис. 3.11. Пояснение к изоморфной вложимости подгруппы в группу

Рис. 3.12. Вложение ассоциативного кольца в тело

пусть в абелевой подгруппе G выделена подполугруппа S, причем в G можно выполнять сокращение на элементы из S, т. е. из $ax = bx$, где *$x \in S$, a, $b \in G$, всегда следует $a = b$. Тогда полугруппу G можно изоморфно вложить в такую абелеву полугруппу $\overline{G}$ с единицей, что всякий элемент из S обладает в $\overline{G}$ обратным элементом.*

Рассмотрим все множество дробей вида a/x, где $a \in G$, $x \in S$, понимая под этим просто упорядоченную пару элементов a, x. Дроби a/x и b/y будем считать равными

$$\frac{a}{x} = \frac{b}{y}$$

в том и только в том случае, если $ay = bx$. Это отношение равенства будет, очевидно, и рефлексивным, и симметричным. Оно транзитивно, так как если также

$$\frac{b}{y} = \frac{c}{z},$$

т. е. $bz = cy$, то

$$ayz = bxz = cyx,$$

откуда после сокращения на $y \in S$ получаем $az = cx$ или

$$\frac{a}{x} = \frac{c}{z}.$$

Таким образом, все множество дробей распадается на пересекающиеся классы равных дробей. Множество этих классов обозначим через $\overline{G}$.

Определим умножение дробей равенством

$$\frac{a}{x}\,\frac{b}{y} = \frac{ab}{xy}. \tag{3.25}$$

Это определение имеет смысл, так как $xy \in S$. Если

$$\frac{a}{x} = \frac{a_1}{x_1}; \quad \frac{b}{y} = \frac{b_1}{y_1}, \tag{3.26}$$

т. е.

$$ax_1 = a_1x; \quad by_1 = b_1y, \tag{3.27}$$

то

$$abx_1y_1 = a_1b_1xy,$$

откуда

$$\frac{ab}{xy} = \frac{a_1b_1}{x_1y_1}.$$

Заменив в левой части равенства (3.25) множители равными им дробями, получим произведение, равное правой части равенства (3.25). Это позволяет рассматривать равенство (3.25) как определение операции умножения классов равных дробей. Ассоциативность и коммутативность этого умножения очевидны, и поэтому $\overline{G}$ превращено нами в абелеву полугруппу.

Все дроби вида z/z, $z \in S$ равны между собой. С другой стороны, если

$$\frac{a}{x} = \frac{z}{z},$$

то $az = zx$, т. е. после сокращения на z $a = x$. Дроби вида z/z составляют, следовательно, отдельный класс. Этот класс играет роль единицы полугруппы $\overline{G}$. Действительно, так как дроби можно сокращать на общий множитель, принадлежащий S,

$$\frac{az}{xy} = \frac{a}{x}, \quad z \in S,$$

ввиду $(az)x = a(xz)$, то

$$\frac{a}{x}\,\frac{z}{z} = \frac{az}{xz} = \frac{a}{x}.$$

Если a — фиксированный элемент из G, то все дроби вида ax/x равны между собой: из $(ax)y = (ay)x$ следует $\frac{ax}{x} = \frac{ay}{y}$.

С другой стороны, если

$$\frac{ax}{x} = \frac{b}{y},$$

то $axy = bx$, т. е. $b = ay$. Наконец, если

$$\frac{ax}{x} = \frac{by}{y},$$

то $axy = byx$, откуда $a = b$. Таким образом, ставя в соответствие каждому элементу a из G класс равных между собой дробей вида ax/x, получаем взаимно однозначное отображение полугруппы G в полугруппу $\overline{G}$. Изоморфность этого отображения вытекает из равенства

$$\frac{ax}{x}\,\frac{by}{y} = \frac{(ab)(xy)}{xy}.$$

Для завершения доказательства заметим, что элементу $z \in S$ соответствует класс дробей вида zy/y. Существование в полугруппе $\overline{G}$ обратного элемента для этого класса вытекает из следующего замечания.

Класс дробей, равных дроби x/y, где $x, y \in S$, обладает в полугруппе $\overline{G}$ обратным элементом. Им служит класс дробей, равный y/x. Действительно,

$$\frac{x}{y}\,\frac{y}{x} = \frac{xy}{yx} = \frac{t}{t}, \qquad t \in S.$$

Из доказанного утверждения и последнего замечания вытекает следующая теорема (случай $S = G$): *всякая абелева полугруппа с законом сокращения может быть вложена в абелеву группу* (рис. 3.13).

К примеру, применяя эту конструкцию к аддитивной полугруппе натуральных чисел, которая является абелевой полугруппой с законом сокращения, получаем известное вложение этой полугруппы в аддитивную группу целых чисел.

Пусть в ассоциативно-коммутативном кольце R выделено множество N элементов, отличных от нуля и не являющихся делителями нуля. Тогда кольцо R можно изоморфно вложить в такое ассоциативно-коммутационное кольцо $\overline{R}$ с единицей, что всякий элемент из R обладает в $\overline{R}$ обратным элементом.

Прежде всего, заметим, что подполугруппа S мультипликативной полугруппы кольца R, порожденная множеством N, также не содержит ни нуля, ни делителей нуля. Действительно, S состоит из всевозможных произведений элементов из N. Однако если $a \neq 0$, $b \neq 0$ и оба эти элемента не являются делителями нуля, то $ab \neq 0$ и из $(ab)c = 0$ следует $bc = 0$, и поэтому $c = 0$.

Таким образом, в мультипликативной полугруппе кольца R можно выполнить сокращение на элементы из S,

Рис. 3.13. Вложение абелевой подгруппы в абелеву группу

а поэтому эта полугруппа вкладывается в такую коммутативную полугруппу $\overline{R}$ с единицей, в которой всякий элемент из S и, в частности, из N обладает обратным элементом (рис. 3.14).

Для превращения $\overline{R}$ в кольцо определим сложение дробей вида $\frac{a}{x}$, $a\in R$, $x\in S$ следующим образом:

$$\frac{a}{x}+\frac{b}{y}=\frac{ay+bx}{xy}. \qquad (3.28)$$

Это определение имеет смысл, так как $xy\in S$. Если справедливы равенства (3.26) и поэтому (3.27), то

$$(ay+bx)\,x_1y_1=ayx_1y_1+bxx_1y_1=a_1xyy_1+b_1yxx_1=(a_1y_1+b_1x_1)\,xy,$$

т. е.

$$\frac{ay+bx}{xy}=\frac{a_1y_1+b_1x_1}{x_1y_1}.$$

Таким образом, равенство (3.28) можно рассматривать как определение операции сложения *классов* равных дробей, т. е. сложение в $\overline{R}$. Коммутативность этого сложения очевидна, а ассоциативность легко проверяется на основании (3.28).

Все дроби $0/z$ равны между собой и составляют полный класс равных дробей. Этот класс играет в $\overline{R}$ роль нуля, так как

$$\frac{a}{x}+\frac{0}{z}=\frac{az}{xz}=\frac{a}{x}.$$

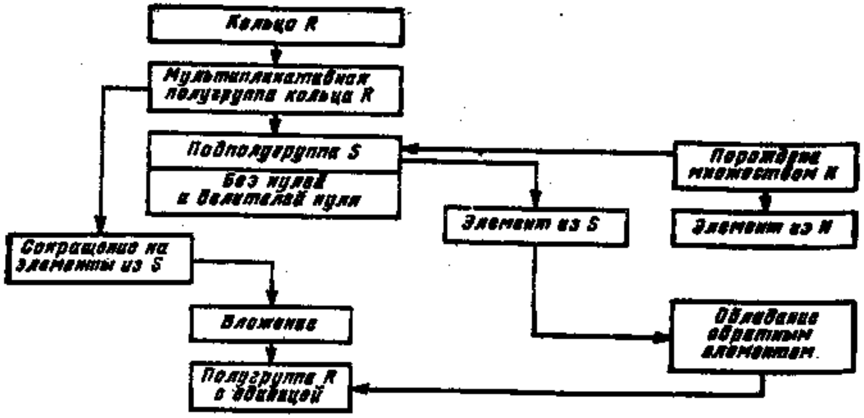


Рис. 3.14. Построение кольца дробей $\overline{R}$ по множеству неделителей нуля N (или мультипликативной полугруппе неделителей нуля S)

Далее, всякий элемент из $\overline{R}$ обладает противоположным элементом, так как

$$\frac{a}{x}+\frac{(-a)}{x}=\frac{ax+(-a)\,x}{x^2}=\frac{0}{x^2}.$$

Отсюда получаем, что $\overline{R}$ является по сложению абелевой группой.

Сложение и умножение связаны в $\overline{R}$ законом дистрибутивности:

$$\frac{a}{x}\,\frac{c}{z}+\frac{b}{y}\,\frac{c}{z}=\frac{ac}{xz}+\frac{bc}{yz}=\frac{acyz+bcxz}{xyz^2}=\frac{acy+bcx}{xyz}=$$
$$=\frac{ay+bx}{xy}\,\frac{c}{z}=\left(\frac{a}{x}+\frac{b}{y}\right)\frac{c}{z}.$$

Таким образом, $\overline{R}$ оказывается ассоциативно-коммутативным кольцом. Известно, что отображение, которое переводит каждый элемент $a\in R$ в класс равных между собой дробей вида ax/x, является изоморфизмом относительно умножения. Оно также изоморфизм относительно сложения, как следует из равенства

$$\frac{ax}{x}+\frac{by}{y}=\frac{axy+byx}{xy}=\frac{(a+b)\,xy}{xy}.$$

Тем самым доказательство теоремы закончено.

Построенное кольцо $\overline{R}$ называется *кольцом дробей* кольца R по множеству неделителей нуля N (или по мультипликативной полугруппе неделителей нуля S). Известно, что в *кольце дробей обратными элементами обладают не только все элементы из S, но и вообще все элементы, представимые дробями вида x/y, где x, $y\in S$.*

Применим полученные результаты к случаю, когда R является областью целостности и можно положить $N=R\backslash 0$, причем $S=N$. В результате можно сформулировать теорему: *всякая область целостности R изоморфно вкладывается в поле.*

Действительно, в этом случае в виде x/y записывается всякий элемент из $\overline{R}$, где x, $y\in S$ отличны от нуля, т. е. всякий такой элемент обладает в $\overline{R}$ обратным элементом.

Кольцо $\overline{R}$ будет, следовательно, полем, это *поле дробей* области целостности R.

Если ассоциативное кольцо R без делителей нуля удовлетворяет условию, выполняющемуся в коммутативном случае всегда и состоящему в том, что для любых элементов a и b из R, отличных от нуля, в R можно найти такие отличные от нуля элементы x и y, что $ax=by$, тогда R изоморфно вкладывается в тело.

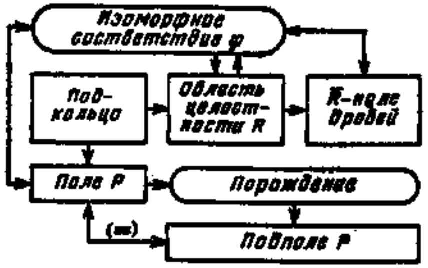


Рис. 3.15. Пояснение к теореме единственности (тождественное отображение кольца на себя)

Рассмотренная выше теорема существования поля дробей дополняется следующей теоремой единственности: *Пусть R — область целостности, $\overline{R}$ — ее поле дробей; пусть, с другой стороны, R является подкольцом некоторого поля P, причем подполе поля P, порожденное множеством R, совпадает с самим P. Тогда между $\overline{R}$ и P существует изоморфное соответствие φ, тождественно отображающее кольцо R на себя, причем это соответствие φ однозначно определено* (рис. 3.15).

Элемент a кольца R, рассматриваемый как элемент поля P, обозначим через a'. Таким образом, соответствие $a \to a'$ является тождественным отображением кольца R на себя.

Покажем, что если изоморфное отображение φ поля $\overline{R}$ на поле P, тождественное на R, существует, то оно однозначно определено. Действительно, если a/b — произвольная дробь, то из $b\frac{a}{b}=a$ следует

$$b\varphi\left(\frac{a}{b}\right)\varphi=a\varphi,$$

т. е., ввиду того что $a\varphi=a'$, $b\varphi=b'$,

$$b'\left(\frac{a}{b}\right)\varphi=a'.$$

. Таким образом, изоморфизм φ переводит элемент a/b поля $\overline{R}$ в частное элементов a' и b' поля P и поэтому определен однозначно.

Переходим к доказательству основного утверждения теоремы. Поставим в соответствие элементу a/b поля $\overline{R}$ частное a'/b' элементов a' и b' поля P, причем положим $\frac{a'}{b'}=\left(\frac{a}{b}\right)\varphi$. Если $a/b=c/d$, т. е. $ad=bc$, то в P справедливо равенство $a'd'=b'c'$, а поэтому ввиду равенства $b'\frac{a'}{b'}=a'$ будет

$$b'd'\frac{a'}{b'}=a'd'=b'c',$$

откуда

$$d'\frac{a'}{b'}=c'.$$

Отображение φ не зависит, следовательно, от выбора записи элемента поля $\overline{R}$ в виде дроби, т. е. является однозначным отображением поля $\overline{R}$ в поле P (рис. 3.15). Оно будет также взаимно однозначным отображением, так как если в поле P имеет место равенство $a'/b'=c'/d'$, то $a'd'=b'c'$, а поэтому в кольце R будет $ad=bc$, откуда для поля $\overline{R}$ справедливо равенство $a/b=c/d$.

Изоморфность отображения φ следует из того, что равенства (3.25) и (3.28), которые определяют умножение и сложение дробей, заведомо справедливы для частных в произвольном P.

Мы получили изоморфное отображение поля дробей $\overline{R}$ в поле P. Оно будет тождественным на кольце R, так как для любого $a \in R$

$$\left(\frac{ab}{b}\right)\varphi=\frac{a'b'}{b'}=a' \in P.$$

Наконец, мы знаем, что изоморфным *образом* поля всегда является поле. Поэтому те элементы поля P, которые записываются в виде частных элементов из R, составляют в поле P подполе, содержащее все кольца R и, следовательно, по условию теоремы совпадающее с P. Нами построено, следовательно, изоморфное отображение φ поля $\overline{R}$ на все поле P, тождественное на R.

Полем дробей для кольца целых чисел служит поле рациональных чисел. Если P — произвольное поле, то кольцо многочленов $P[x]$ является областью целостности и поэтому вкладывается в поле дробей. Это поле обозначается через $P(x)$ и называется *полем рациональных дробей* от неизвестного x над полем P. Его элементы имеют вид дробей $f(x)/g(x)$, где $f(x)$ и $g(x)$ — многочлены из $P[x]$, причем $g(x)\neq 0$, а равенство этих дробей в операции над ними определяется в соответствии с ранее изложенным.

Аналогичный смысл имеет поле рациональных дробей $p(x_1, x_2, \ldots, x_n)$ от нескольких переменных.

Поле дробей для кольца степенных рядов $P\{x\}$ над полем P изоморфно *полю «лорановых» степенных рядов над P*, т. е. рядов вида

$$a_n x^n + a_{n+1} x^{n+1} + \ . \ . \ . \ , \qquad (3.29)$$

где n может быть больше, равно или меньше нуля, а все коэффициенты принадлежат P. Ряд (3.29) в общем случае, следовательно, содержит конечное число членов с отрицательными степенями неизвестного x и бесконечно много членов с его положительными степенями. Операции над этими рядами выполняются по правилам, естественно обобщающим правила операций в кольце $P\{x\}$.

Глава 4

ТЕОРИЯ КАТЕГОРИЙ

4.1. ВВЕДЕНИЕ В ТЕОРИЮ КАТЕГОРИЙ

Прежде всего заметим, что существуют философское и математическое понятия категорий. Появление современной теории категорий можно рассматривать как еще один шаг в наступлении, которое ведет точная наука математика на «качественную» неточную науку философию. Однако можно сказать, что математическая теория категорий недостаточно отображает (моделирует) философскую теорию категорий, многие положения которой необходимы инженеру-кибернетику. Поэтому, по-видимому, целесообразно говорить еще об одной теории категорий — системотехнической, системологической [6, 16], которая широко используется инженером, максимально приближена к философии, но не может рассматриваться как строгая математическая теория.

В современной математической теории категорий имеются, по крайней мере, два подхода. В одном, условно называемом алгебраическим, остаются в основном на алгебраических позициях [46]. Во втором существенно влияние геометрического топологического подхода [46].

Инженеру и инженеру-математику, специалисту по кибернетике математическая теория категорий необходима, по крайней мере, по трем соображениям. Во-первых, она представляет собой более мощное средство для описания семантики кибернетических систем, в том числе лингвистических. Во-вторых, она позволяет описывать и анализировать внутренние противоречивые системы типа систем искусственного интеллекта. Математическую теорию категорий можно рассматривать как некоторый аппарат для описания диалектических, противоречивых, динамических философских логик (типа логик Гегеля, Канта). Классические логики, например такая, как исчисление высказываний и предикатов, недостаточны для описания поведения современных систем.

Наконец, в-третьих, важная особенность заключается в предоставлении новых (по сравнению, в частности, с λ-исчислением), более совершенных, чем функция и аппликация, средств для описания последовательности получения процедурности той или иной зависимости. Поэтому нет ничего удивительного в том, что аппарат теории категорий применяется для описания процедурности и непроцедурности кибернетических (в частности, программных) систем [49, 50]. Эти обстоятельства стали причиной построения теоретико-категориального обоснования математики в противовес теоретико-множественному. Кроме того, рассмотренные особенности позволили более эффективно по сравнению с теоретико-множественным подходом анализировать логические системы вывода.

При теоретико-множественном подходе используется концепция отношения принадлежности данного элемента (к множеству) и свойства совокупности, причем множества определяются через его элементы с помощью его внутренней структуры. При теоретико-категориальном подходе свойства совокупности определяются через внешние связи с другими совокупностями. Это очень похоже на моделирование, на анализ семантики в теории искусственного интеллекта с помощью семантических сетей [30]. Поэтому теория категорий содержит дополнительные возможности по математическому описанию семантики кибернетических систем. Эти два подхода позволяют построить разными путями основания математики с формированием различных математических конструкций и выражением свойств математических объектов. Связь между совокупностью выражается с помощью стрелок (или преобразований, или морфизмов), которые представляют некоторую абстрацию от понятий функции или отображения. Категория может рассматриваться как универсум (универсальное понятие) для определенного рода математических рассуждений. Этот универсум состоит из объектов и стрелок. Так, категория для топологических исследований состоит из топологических пространств (объектов) и непрерывных функций (стрелок). Для линейной алгебры категория состоит из объектов — векторных пространств и стрелок — линейных преобразований этих пространств. В теории групп стрелками категорий служат гомоморфизмы групп. Очень важно отметить, что истоки теории категорий лежат в алгебраической топологии (близкой к комбинаторной топологии [39, 46]), в которой рассматриваются построения, конструкции, средства, связывающие топологию с алгеброй, а именно с теорией групп. Теоретико-стрелочные термины и мышление, предлагаемые в теории категорий, оказались очень полезными в прикладных областях, в инженерных приложениях. Появился целый набор категорий, основные из которых представлены в табл. 4.1.

Пример 4.1. В категории множеств Set объектами являются всевозможные непустые множества, а стрелками (морфизмами) — всевозможные отображения множеств друг в друга. Основные теоретико-множественные операции и понятия (пустое множество, пересечение, произведение и т. д.) описываются с помощью стрелок в Set. Но этот аппарат слабее, неполно описывает исходную ситуацию в теории множеств. Интерпретации теоретико-множественных понятий могут себя вести отлично от соответствующих понятий в категории Set. Поэтому была введена категория топосов, интерпретации которой в основной части совпадают с интерпретациями категории Set.

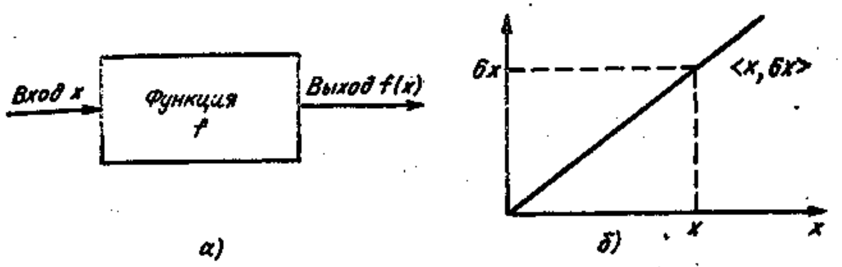
Таблица 4.1

Категория	Объекты	Стрелки (морфизмы)
Set	Все множества	Все функции между множествами
Finset	Все конечные множества	Все функции между конечными множествами
Nonset	Все непустые множества	Все функции между непустыми множествами
Top	Все топологические пространства	Все непрерывные функции между топологическими пространствами
Vect	Векторные пространства	Линейные преобразования
Grp	Группы	Гомоморфизмы групп
Mon	Моноиды	Гомоморфизмы моноидов
Met	Метрические пространства	Сжатия
Man	Многообразия	Гладкие отображения
Top Grp	Топологические группы	Непрерывные гомоморфизмы
Pos	Частично упорядоченные множества	Монотонные функции

Термин «топос» означает по-гречески место, участок и был введен в алгебраической геометрии для определения понятия категории пучков (совокупность пучков над топологическим пространством). Они обозначались особой стрелкой, названной классификатором подобъектов, благодаря которой все топосы стали иметь классификаторы подобъектов. Это позволило создать теорию категорий полностью на категориальном языке и установить эквивалентность элементарных топосов.

Категории Set. Топосы имеют общую связь с математической логикой. Классическая логика представлена в Set операциями на некотором множестве двухэлементной булевой алгеброй. Каждый топос имеет аналог этой алгебры. Поэтому каждый топос определяет свое собственное логическое исчисление, и можно говорить о категориальном анализе логики с помощью топосов.

Интересно отметить, что это исчисление может отличаться от классической логики, и можно утверждать, что логические принципы в тополе являются инструментом интуиционистской логики [46]. Теперь множественная интуиционистская логика Крипке может быть использована для отыскания топосов, в которых логические построения оказываются переформулировкой моделей Крипке (в виде категории пучков). Для первичного знакомства с понятием категории нам потребуется несколько необычное определение функции. Вместо традиционного «функцией называется множество упорядоченных пар, таких, что...» рассмотрим следующее определение функции. Введем в рассмотрение для множества A и B произведение или декартово произведение этих множеств как множество всех упорядоченных пар, первый элемент которых принадлежит A, второй — B. Обозначим его через $A\times B$, т. е.



а) б)

Рис. 4.1. Пояснение к понятию функции в теории категорий

$$A\times B=\{\langle x,\ y\rangle : x\in A \quad \text{и} \quad y\in B\}.$$

Определим функцию как тройку:

$$f=\langle A,\ B,\ R\rangle,$$

где $R\subseteq A\times B$ — отношение между A и B (в частности, график функции f), такое, что для каждого $x\in A$ существует ровно один $y\in B$ с $<x,\ y>\in R$.

С помощью этого определения с самого начала в него вводятся области определений и значений A и B. Это определение все же остается теоретико-множественным и представляет функцию как неизменное множество пар (рис. 4.1), т. е. как фиксированный, статический объект, и не отражает динамику, операционный, процедурный характер данного понятия.

Введем понятие *образа функции* f. Область определения функции f (множество входов), которая задается как множество упорядоченных пар, восстанавливается как

$$\mathrm{dom}\, f=\{x:\langle x,\ y\rangle\in f \ \text{для некоторого}\ y\}.$$

Областью значений заданной функции может быть любое множество значений, включающее все выходы функции $f(x)$. Все выходы образуют так называемую область изменений или образ функции f:

$$\mathrm{Im}\, f=\{y:\langle x,\ y\rangle\in f \quad \text{для} \quad \text{некоторого}\ \vec{x}\in A\}.$$

В общем случае f можно называть функцией из множества A в множество B, где $\mathrm{A}=\mathrm{dom}\, f$, а $\mathrm{Im}\, f\subseteq B$. Следовательно, для функции, заданной как множество пар, область значений не определяется однозначно. Поэтому приходится в этом случае вво-

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "g\circ f"'] & B \arrow[d, "g"] \\ & C \end{tikzcd}$$

Рис. 4.2. Пояснение к операции композиции двух функций

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "g\circ f" near start] \arrow[d, bend right=60, "h\circ(g\circ f)"'] \arrow[d, bend right=30, "(h\circ g)\circ f"'] & B \arrow[d, "g"] \arrow[dl, "h\circ g" near end] \\ D & C \arrow[l, "h"] \end{tikzcd}$$

Рис. 4.3. Пояснение к операции композиции трех функций

дить тождественную функцию. Тождественной функцией множества A называется функция id_A, областью значений которой является данное множество A. Образ id_A совпадает с A, с теоретико-множественной точки зрения $\mathrm{id}_A=\{<x,\ x>: x\subseteq A\}$.

Рассмотрим случай, когда A — подмножество множества B. Тогда правило $f(x)=x$ задает функцию из A в B. Эта функция называется включением подмножества A в множество B. Термин предполагает, что функция включает элементы множества A в совокупность элементов множества B (подчеркивается история, процедурность). Поэтому рассмотрим понятие композиции. Пусть $f: A\to B$ и $g: B\to C$ — две функции, такие, что цель первой из них совпадает с источником другой. Переход от x к $g(f(x))$ определяет функцию с областью определения A и областью значения C, которая называется композицией f и g и символически обозначается $g\circ f: g\circ f(x)=g(f(x))$. Графически композиция представлена на рис. 4.2. Допустим, имеются три функции $f: A\to B$, $g: B\to C$ и $L: C\to D$, области определения и значений которых связаны таким образом, что можно применить последовательно все три функции и получить в результате функцию из A в D. Это можно сделать двумя способами, образовав на первом этапе композиции $g\circ f: A\to C$ или $h\circ g: B\to D$. Далее, по правилу «применить f, а затем $h\circ g$» получаем функцию $(h\circ g)\circ f$, а по правилу «применить $g\circ f$, а затем h» — композицию $h\circ(g\circ f)$. С помощью диаграммы все эти процессы могут быть представлены в виде, изображенном на рис. 4.3. Эти две функции, полученные различными способами, совпадают, т. е.

$$[h\circ(g\circ f)](x)=h(g\circ f(x))=h(g(f(x)));$$
$$[(h\circ g)\circ f](x)=h\circ g(f(x))=h(g(f(x))).$$

Нами установлен закон ассоциативности для композиции функций:

$$h\circ(g\circ f)=(h\circ g)\circ f.$$

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[dr, "h"'] & B \arrow[d, "g"] \\ & C \end{tikzcd}$$

Рис. 4.4. Пояснение к коммутативной диаграмме

Этот закон справедлив не для любых трех функций, а только для таких, которые образуют путь (т. е. источник одной функции является стоком другой). Диаграмма рис. 4.3, называемая коммутативной диаграммой, широко используется в теории категорий и является как бы категориальным представлением композиции трех функций.

Не всякая диаграмма коммутативна. Так, для случая, представленного на рис. 4.4, она коммутативна только при $h=g\circ f$, т. е. если путь по дуге h совпадает с путем по дугам g и f.

В заключение введения в теорию категорий полезно напомнить о том, как «совершенствовалось» понятие функции с помощью более понятной аппликации, продукции (в том числе подстановки, операции переписывания). Категория является еще одним способом расширения функционального подхода. Изложение элементов теории категорий и топосов в дальнейшем следует схеме [46].

4.2. ОСНОВНЫЕ ПОНЯТИЯ. ОПРЕДЕЛЕНИЕ КАТЕГОРИИ

Категория K есть совокупность класса Ob_K, элементы которого называются объектами K, и класса Mor_K, элементы которого называются морфизмами категории K. Сразу заметим, что на неформальном уровне морфизмы — это преобразования или (в более узком смысле правил применительно к лингвистике) правила вывода. Объекты и морфизмы должны быть связаны между собой следующими условиями.

1. Каждой упорядоченной паре объектов A, $B\in Ob_K$ сопоставлено некоторое множество $\mathrm{Mor}_K(A, B)$ морфизмов категории K.

2. Каждый морфизм категории K принадлежит одному и только одному из множеств Mor_K.

3. В классе морфизмов Mor_K введена частичная бинарная операция композиции: композиция $\alpha\beta$ морфизмов $\alpha\in\mathrm{Mor}_K(A, B)$ и $\beta\in\mathrm{Mor}_K(C, D)$ определена тогда и только тогда, когда объект B совпадает с объектом C, и в этом случае морфизм $\alpha\beta\in\mathrm{Mor}_K(A, D)$ — ассоциативная частичная композиция, т. е. $(\alpha\beta)\gamma=\alpha(\beta\gamma)$ для любых трех морфизмов $\alpha\in\mathrm{Mor}_K(A, B)$, $\beta\in\mathrm{Mor}_K(B, C)$ и $\gamma\in\mathrm{Mor}_K(C, D)$.

4. В каждом множестве $\mathrm{Mor}_K(A, A)$, $A\in Ob_K$ содержится морфизм I_A, называемый тождественным или единичным морфизмом объекта A, такой, что $\alpha I_A=\alpha$ и $I_A\beta=\beta$ для любых морфизмов $\alpha\in\mathrm{Mor}_K(X, A)$ и $\beta\in\mathrm{Mor}_K(A, Y)$.

Для $\alpha \in \mathrm{Mor}_K(A, B)$ введем обозначение в виде стрелки $\alpha : A \to B$ и будем использовать терминологию «морфизм (преобразование) α с началом A и концом B», или «из A в B».

Пример 4.2. Каждая грамматика $G=(N, \Sigma, P, S)$ представляет собой некоторую категорию KG. Объектами этой категории являются цепочки $(N \cup \Sigma)$, состоящие из терминальных (N) и нетерминальных символов. Неформально морфизмы — это правила вывода (P), но требуется уточнить понятие вывода; в частности, в выводе должны учитываться последовательности применения продукций и контекстов.

Морфизмы могут обладать свойством изоморфизма, которое заключается в следующем. Морфизм $\eta : A \to B$ категории K называется изоморфизмом, если существует такой морфизм $\delta : B \to A$, что $\eta\delta = 1_A$ и $\delta\eta = 1_B$.

В разных руководствах по теории категорий применяются разные терминологии и точки зрения (аспекты). Вместо термина «морфизм» используется понятие «стрелка», и то и другое подразумевает преобразование или функцию:

МОРФИЗМ = СТРЕЛКА = ПРЕОБРАЗОВАНИЕ.

Поэтому дадим еще одно аксиоматическое (пять аксиом) определение категории.

Категория K включает в себя:

1) совокупность предметов, называемых K-объектами;

2) совокупность предметов, называемых K-стрелками;

3) операции, ставящие в соответствие K-стрелке f K-объект $\mathrm{dom}\, f$ (начало стрелки f) и K-объект $\mathrm{cod}\, f$ (конец стрелки f). То, что $a = \mathrm{dom}\, f$ и $b = \mathrm{cod}\, f$, изображается следующим образом: $f : a \to b$ или $a \xrightarrow{f} b$;

4) операцию, ставящую в соответствие каждой паре $<g, f>$ K-стрелок с $\mathrm{dom}\, g = \mathrm{cod}\, f$ K-стрелку $g \circ f$, называемую композицией f и g, с $\mathrm{dom}\,(g \circ f) = \mathrm{dom}\, f$ и $\mathrm{cod}\,(g \circ f) = \mathrm{cod}\, g$, т. е. $g \circ f : \mathrm{dom}\, f \to \mathrm{cod}\, g$, причем выполняется следующий закон ассоциативности.

Пусть $a \xrightarrow{f} b \xrightarrow{g} c \xrightarrow{h} d$ — конфигурация K-объектов и K-стрелок, тогда $h \circ (g \circ f) = (h \circ g) \circ f$. Закон ассоциативности утверждает, что диаграмма на рис. 4.5 всегда коммутативна;

5) сопоставление каждому K-объекту b K-стрелки $1b : b \to b$, называемой единичной или тождественной стрелкой, так что выполняется закон тождества: для любых K-стрелок $f : a \to b$ и $g : b \to c$

$$1_b \circ f = f \quad \text{и} \quad g \circ 1_b = g$$

и диаграмма, представленная на рис. 4.6, всегда коммутативна.

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[d, shift right=2, "h \circ (g \circ f)"'] \arrow[d, "(h \circ g) \circ f"] \arrow[dr, "g \circ f"] & b \arrow[d, "g"] \arrow[dl, "h \circ g"'] \\
d & c \arrow[l, "h"']
\end{tikzcd}$$

Рис. 4.5. Коммутативная диаграмма категории

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[dr, "f"'] & b \arrow[d, "1_b"] \arrow[dr, "g"] & \\
 & b \arrow[r, "g"'] & c
\end{tikzcd}$$

Рис. 4.6. Пояснение к закону тождества

Пусть K — произвольная категория. Множество D объектов и морфизмов категории K называется диаграммой, если с каждым морфизмом $\alpha : A \to B$, принадлежащим D, в D содержатся объекты A и B и морфизмы 1_A и 1_B. Пусть D — некоторая диаграмма в категории K и $A, B \in Ob_K$. Будем говорить, что в диаграмме D существует путь из объекта A в объект B, если диаграмма D содержит некоторую последовательность морфизмов

$$A \xrightarrow{\alpha_1} A_1, \qquad A_1 \xrightarrow{\alpha_2} A_2, \ldots, A_{n-1} \xrightarrow{\alpha_n} B.$$

Этот путь из объекта A в объект B обозначается $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Диаграмма D считается коммутативной, если для двух объектов A и B, принадлежащих D, и любых двух путей $(\alpha_1, \ldots,$

Рис. 4.7. Пояснение к примеру 4.3 →

$$\begin{tikzcd}
X_1 \arrow[r, "f_1"] \arrow[d, "\alpha_1"'] & Y_1 \arrow[d, "\alpha_2"] \\
X_2 \arrow[r, "f_2"] & Y_2
\end{tikzcd}$$

Рис. 4.8. Композиция двух морфизмов

$$\begin{tikzcd}
X_1 \arrow[r, "f_1"] \arrow[d, "\alpha_1"'] & Y_1 \arrow[d, "\alpha_2"] & X_2 \arrow[r, "f_3"] \arrow[d, "\beta_1"'] & Y_2 \arrow[d, "\beta_2"] & X_1 \arrow[r, "f_1"] \arrow[d, "\alpha_1\beta_1"'] & Y_1 \arrow[d, "\alpha_2\beta_2"] \\
X_2 \arrow[r, "f_2"] & Y_2 & X_3 \arrow[r, "f_4"] & Y_3 & X_3 \arrow[r, "f_2"] & Y_3
\end{tikzcd}$$

а) б) в)

..., α_n) и ($\beta_1, \ldots, \beta_m$) из A и B в категории K имеет место равенство

$$\alpha_1\alpha_2 \ . \ \ . \ \ . \alpha_n = \beta_1\beta_2 \ . \ \ . \ \ . \ \beta_m.$$

Пример 4.3. Пусть объектами категории являются тройки (X, f, Y). Морфизмы категориальной пары (α_1, α_2) такие, что диаграмма, представленная на рис. 4.7, коммутативна, т. е. $f_1\alpha_2 = \alpha_1 f_2$. В этом примере первый и третий компоненты — структурированные множества, а второй компонент — функция. Компоненты морфизма — тоже функции. Композиция морфизмов (α_1, α_2) и (β_1, β_2), которым соответствуют диаграммы на рис. 4.8, *а*, *б*, является морфизмом ($\alpha_1\beta_1$, $\alpha_2\beta_2$), которому соответствует диаграмма на рис. 4.8, *в*. Нетрудно убедиться, что объекты и морфизмы диаграммы на рис. 4.8, *в* образуют категорию.

Можно составить тройки из морфизмов α и β некоторой категории T, они тоже составят категорию.

4.3. ХАРАКТЕРНЫЕ ПРЕДСТАВИТЕЛИ КАТЕГОРИЙ

Рассмотрим примеры основных категорий.

1. Категория 1. Она имеет единственный объект a и стрелку f. Этим категория определяется полностью. Для этой категории всегда справедливо $\operatorname{dom} f = \operatorname{cod} f = a$. Так как f — единственная стрелка, то в качестве единичной стрелки можно взять только ее, поэтому полагаем $1_a = f$. Единственной парой, для которой надо определить композицию, является пара $<f, f>$, для которой $f \circ f = f$. Отсюда получаем закон тождества $1_a \circ f = f \circ 1_a = f \circ f = f$. Диаграмма для этой категории представлена на рис. 4.9. В качестве представителя этой категории можно рассмотреть категорию, единственным объектом которой является число 0, а единственной стрелкой — упорядоченная пара $<0, 0>$.

2. Категория 2. Эта категория имеет два объекта и три стрелки (рис. 4.10). Если в качестве двух объектов взять 0 и 1, а в качестве стрелок — пары $<0,0>$, $<0,1>$ и $<1,1>$, то

$$\langle 0,\ 0\rangle : 0 \to 0;$$
$$\langle 0,\ 1\rangle : 0 \to 1;$$
$$\langle 1,\ 1\rangle : 1 \to 1.$$

Рис. 4.9. Диаграмма категории 1

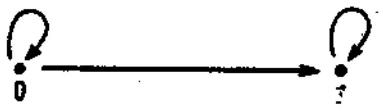


Рис. 4.10. Диаграмма категории 2

Тогда

$$\langle 0,\ 0\rangle = 1_0 \text{ (единичная стрелка на 0)};$$
$$\langle 1,\ 1\rangle = 1_1.$$

Композиция на этом множестве вводится только единственным способом:

$$1 \circ 1_0 = 1_0; \qquad \langle 0,\ 1\rangle \circ 1_0 = \langle 0,\ 1\rangle; \qquad 1_1 \circ \langle 0,\ 1\rangle = \langle 0,\ 1\rangle$$
$$\text{и} \quad \langle 1_1 \circ 1_0\rangle = 1_1.$$

3. Категория 3. Эта категория состоит из трех объектов и шести стрелок. Диаграмма категории приведена на рис. 4.11. Для этой категории композиция определяется однозначным способом.

4. Категория предпорядка. В рассмотренных трех категориях композиция определялась однозначно благодаря тому, что любые два объекта связаны не более чем одной стрелкой. В общем случае категория, в которой любые два объекта p и q связаны не более чем одной стрелкой $p \to q$, называется категорией предпорядка. Пусть p — совокупность объектов категории предпорядка, тогда на ней можно определить бинарное отношение, или, иначе, множество, $R \subseteq P \times P$; $<p, q> \in R$ тогда и только тогда, когда в данной категории существует стрелка $p \to q$.

Если ввести обозначение pRq вместо $<p, q> \in R$, то отношение P будет иметь следующие свойства:

а) рефлексивность, т. е. для каждого p выполняется pRp;

б) транзитивность, т. е. если pRq и qRs, то pRs.

Транзитивное и рефлективное бинарное отношение называют отношением предпорядка.

Предыдущие рассуждения убеждают нас в том, что категория предпорядка определяет единственное отношение предпорядка на своих объектах, откуда и следует ее название.

Если проводить обратные рассуждения, начиная с множества P предупорядоченного отношения R (что означает, что $R \subseteq P \times P$ — рефлексивное и транзитивное отношение), то можно построить категорию предпорядка следующим образом. Ее объектами являются элементы p из множества P, а стрелками — пары $<p, q>$, для которых pRq. Пара $<p, q>$ является стрелкой из p в q. Для компонующихся пар справедливо

$$p \xrightarrow{\langle p,\ q\rangle} q \xrightarrow{\langle q,\ s\rangle} s.$$

Положим

$$\langle q,\ s\rangle \circ \langle p,\ q\rangle = \langle p,\ s\rangle.$$

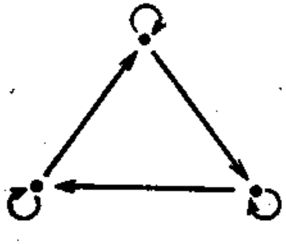


Рис. 4.11. Диаграмма категории 3

Так как $<p, q>$ и $<q, s>$ — стрелки, то справедливо, что pRq, qRs и pRs (транзитивность), откуда следует, что $<p, s>$ — тоже стрелка. Заметим, что имеется не более одной стрелки из p в q, причем ее наличие зависит от условия выполнения pRq; по условию транзитивности существует единственная композиция стрелок. По условию рефлективности $<p, p>$ — тоже всегда стрелка для любого элемента p. Кроме того, $<p, p> = 1_p$.

Ранее были определены особые категории предпорядка, для которых отношение предпорядка удовлетворяет еще одному условию — антисимметричности;

в) антисимметричность, т. е. если справедливо, что pRq и qRq, то $p=q$.

Антисимметричное отношение предпорядка называется отношением частичного порядка. Этот тип отношения обычно обозначается как $\leqslant$, т. е. пишется $p\leqslant q$ вместо pRq. Поэтому можно определить ЧУМ или ЧУМ как пару $P=<p, \leqslant>$, состоящую из множества P и отношения частичного порядка $\leqslant$ на P. Эта структура имеет большое значение в теории топосов. Множество $\{0\}$ превращается в ЧУМ, если положить $0\leqslant 0$. Соответствующей категорией предпорядка является категория 4. Категория предпорядка 2 соответствует отношению частичного порядка на множестве $\{0,1\}$ с $0\leqslant 1$, а также с $0\leqslant 0$ и $1\leqslant 1$.

В данном случае это включает в себя обычный числовой порядок $\leqslant$ для чисел 0 и 1, который означает «меньше или равно». Но следует помнить, что отношение предпорядка справедливо и для логических, семантических отношений.

Категория 3 соответствует обычному упорядочению множества $\{0, 1, 2\}$ и т. д. В общем случае можно при данном n построить категорию предпорядка n, использовав обычный порядок на $\{0, 1, 2, \ldots, -1\}$. Продолжив это конструирование, рассмотрим бесконечную совокупность $\omega=\{0, 1, 2, 3, \ldots, n-1\}$ всех натуральных чисел с обычным отношением порядка. Ей соответствует категория предпорядка с диаграммой

$$0\to 1\to 2\to 3\to \,.\;.\;.\;,$$

где композиция и единичные стрелки опущены. В заключение приведем пример категории предпорядка, которая не удовлетворяет условию частичного порядка (рис. 4.12). Для этой категории справедливо pRq и qRp, но $p\neq q$ (никогда).

5. Дискретная категория. Пусть b — объект некоторой категории K. Тогда K-стрелка 1_b определяется однозначно

Рис. 4.12. Диаграмма категории предпорядка, для которой не удовлетворяется условие частичной упорядоченности

в соответствии со свойством закона тождества. Действительно, если стрелка $1':b\to b$ обладает тем свойством, что диаграмма на рис. 4.13, *а* коммутативна для любых K-стрелок f и q указанного выше вида, то в частном случае, когда $f=1'$ и $g=1_b$, будет также коммутативна диаграмма на рис. 4.13, *б*. Отсюда $1_b=1_b\circ 1'$ (правый треугольник рис. 4.13, *б*). Но по закону тождества (для $f=1'$) $l_b\circ 1'=1'$. Значит, $1_b=1'$.

Так как единичная стрелка 1_b определяется однозначно по объекту b, то очень часто отождествляют объект b со стрелкой 1_b и пишут $b:b\to b$, $b\circ f$ и т. д. Поэтому будем считать, что в соответствии с аксиомами теории категорий совокупность K-стрелок включает, по крайней мере, единичную стрелку для каждого K-объекта, причем различные объекты должны иметь разные стрелки. Категория K называется дискретной, когда она имеет только единичные стрелки (каждая стрелка является единичной для некоторого объекта). Дискретная категория является категорией предпорядка, так как в ней каждый объект имеет лишь одну стрелку. Отождествляя объекты с соответствующими единичными стрелками, убеждаемся, что дискретная категория является совокупностью объектов. Действительно, любое множество X можно превратить в дискретную категорию, если добавить единичные стрелки $x\to x$ для каждого $x\in X$. Тем самым X превращается в категорию предпорядка, соответствующую отношению $R\subseteq X\times X$, такому, что соотношение xRy справедливо тогда и только тогда, когда $x=y$.

6. Категория N. Категория имеет более одной стрелки между объектами. Пусть категория имеет один объект N и бесконечную совокупность стрелок из N в N. По определению этими стрелками будут натуральные числа 0, 1, 2, 3... Каждая

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[dr, "f"'] & b \arrow[d, "f'"] \arrow[dr, "g"] & \\
 & b \arrow[r, "g"'] & c
\end{tikzcd}
\qquad
\begin{tikzcd}
b \arrow[r, "1'"] \arrow[dr, "1'"'] & b \arrow[d, "1'"] \arrow[dr, "1_b"] & \\
 & b \arrow[r, "1_f"'] & b
\end{tikzcd}$$

а) б)

Рис. 4.13. Пояснение к понятию дискретной категории

$$\begin{tikzcd} N \arrow[r, "n"] \arrow[dr, "m+n"'] & N \arrow[d, "m"] \\ & N \end{tikzcd}$$

а)

$$\begin{tikzcd} N \arrow[r, "m"] \arrow[dr, "m"'] & N \arrow[d, "0"] \arrow[dr, "n"] & \\ & N \arrow[r, "n"'] & N \end{tikzcd}$$

б)

Рис. 4.14. Пояснение к понятию категории N

из этих стрелок имеет одни и те же начало и конец — объект N; композиция двух стрелок (чисел) m и n есть снова число. Положим

$$m \circ n = m + n.$$

Тогда соответствующая диаграмма на рис. 4.14, *б* коммутативна по определению. Закон ассоциативности следует из свойства ассоциативности для определения сложения, т. е. из того, что $m+(n+k)=(m+n)+k$ для любых m, n и k. Единичная стрелка объекта N задается числом 0. Диаграмма на рис. 4.14, *б* коммутативна, так как

$$0+m=m \quad \text{и} \quad n+0=n.$$

7. **Моноид.** Категория N является категорией потому, что структура $(N, +, 0)$ является примером моноида, известного из абстрактной алгебры.

Моноидом называется тройка

$$M=(m, \ * , \ e),$$

где M — некоторое множество; $*$ — бинарная операция на M, т. е. функция, ставящая в соответствие каждой паре $<x, y> \in M \times M$ элемент из M, причем $*$ ассоциативна, т. е. $x*(y*z)=(x*y)*z$ для любых $x, y, z \in M$; e — элемент множества M, называемый единицей моноида, т. е. $e*x=x*e=x$ при всех $x \in M$.

По любому моноиду M строится категория с одним объектом, как в случае 6. В качестве объекта берется множество M, а в качестве стрелок $M \to M$ — элементы из M, кроме того, предполагается, что $e=1_M$. Композиция стрелок $x, y \in M$ задается правилом $x \circ y = x * y$.

Справедливо и обратное утверждение, что если K — категория с единственным объектом a и M — совокупность ее стрелок, то тройка $(M, \ \circ, 1_a)$ является моноидом. Все стрелки имеют одни и те же начало и конец, поэтому для них всегда определена композиция, которая представляет собой функцию из $M \times M$ в M, т. е. бинарную операцию на M, ассоциативную по закону ассоциативности для категорий.

По закону тождества 1_a является единицей данного моноида.

8. **Категория** Matr (K). Пусть K — коммутативное кольцо. Тогда множество всех матриц над K можно объединить в категорию Matr (K). Ее объектами являются целые положительные числа 1, 2, 3, ..., а стрелками $m \to n$ — матрицы типа $n \times m$ с элементами из K. Для пары компонуемых стрелок

$$m \xrightarrow{B} n \xrightarrow{A} p,$$

т. е. для матрицы **A** типа $p \times n$ и матрицы **B** типа $n \times m$ композиция $A \circ B$ определяется как произведение $\mathbf{A} \times \mathbf{B}$ матриц **A** и **B**, которое является $(p \times n)$ матрицей и задает стрелку вида $(m \to p)$. Категориальная ассоциативность получается из ассоциативности операции матричного умножения. Единичной стрелкой является 1_m — единичная матрица порядка m.

Эта категория нам понадобится для установления связи с тензорной алгеброй и исчислением.

9. **Подкатегория.** Пусть K — категория, a, b — ее объекты. Обозначим через $K(a, b)$ совокупность K-стрелок с началом a и концом b, т. е.

$K(a, b) = \{f : b$ является K-стрелкой вида $a \xrightarrow{f} b\}$.

Категория K называется подкатегорией (в обозначениях это $K \in D$), если:

а) каждый K-объект является D-объектом;

б) для любых двух K-объектов a и b справедливо включение

$$K(a, \ b) \subseteq D(a, \ b),$$

т. е. все K-стрелки $a \to b$ представлены в D.

Так, Finset$\subseteq$Set и Nonset$\subseteq$Set, хотя ни Finset, ни Nonset не являются подкатегориями одна другой;

в) если для любых K-объектов a и b выполняется равенство

$$K(a, \ b) = D(a, \ b),$$

то K называется полной подкатегорией D.

Пусть D — некоторая категория, а C — произвольная совокупность D-объектов. Тогда имеется полная подкатегория K в D, K-стрелками которой являются все D-стрелки между элементами из C. Нетрудно убедиться, что Finset и Nonset — полные подкатегории Set. Одной из важных полных подкатегорий в Finset (а значит, и в Set) является категория Finord всех ординалов. Под конечными ординалами понимаются множества, которые обычно используются в основаниях теории множеств как представители натуральных чисел. Натуральные числа берутся в качестве имен этих множеств, а именно:

$$0 \text{ для } \varnothing \text{ (пустое множество);}$$
$$1 \text{ для } \{0\} \ (=\{\varnothing\});$$
$$2 \text{ для } \{0,\ 1\} \ (=\{\varnothing,\ \{\varnothing\}\});$$
$$3 \text{ для } \{0,\ 1,\ 2\} \ (=\{\varnothing,\ \{\varnothing,\ \{\varnothing\{\varnothing\}\}\}\})$$

и т. д.

Продолжая этот процесс индуктивно, положим, что натуральное число n есть имя множества

$$\{0,\ 1,\ 2,\ .\ .\ .\ ,\ n-1\}.$$

Образованная таким образом последовательность конечных множеств состоит из всех конечных ординалов, которые и составляют объекты категории Finord, а стрелками являются все функции между конечными ординалами.

10. Произведение категорий. Объектами категории Set^2 пар множеств служат все пары $<A, B>$ множеств. Стрелками в Set^2 из $<A, B>$ в $<C, D>$ являются пары $<f, g>$ функций вида $f: A\to C$ и $g: B\to D$. Композиция определяется правилом $<f,\ g>\circ<f',\ g'>=<f\circ f',\ g\circ g'>$, где $f\circ f'$ и $g\circ g'$ — обычные композиции функций. Единичной стрелкой объекта $<A, B>$ будет пара $<\mathrm{id}_A,\ \mathrm{id}_B>$.

11. Категория стрелок. Объектами этой категории функций, обозначаемой $\mathrm{Set}^{\to}$, являются все функции $f: A\to B$. Стрелка из $\mathrm{Set}^{\to}$-объекта $f: A\to B$ в $\mathrm{Set}^{\to}$-объект $g: C\to D$ представляет собой пару функций $<h, k>$, такую, что диаграмма на рис. 4.15, *а* коммутативна, т. е. $g\circ h=k\circ f$. Композицию задаем правилом $<j,\ l>\circ<h,\ k>=<j\circ h,\ l\circ k>$. Корректность ее поясняется диаграммой на рис. 4.15, *б*. Единичной стрелкой $\mathrm{Set}^{\to}$-объекта $f: A\to B$ является пара тождественных функций $<\mathrm{id}_A,\ \mathrm{id}_B>$.

$$\begin{tikzcd}
A \arrow[r, "h"] \arrow[d, "f"'] & C \arrow[d, "q"] \\
B \arrow[r, "k"] & D
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "h"] \arrow[d, "f"'] & C \arrow[r, "j"] \arrow[d] & E \arrow[d, "i"] \\
B \arrow[r, "k"] & D \arrow[r, "e"] & F
\end{tikzcd}$$

а) *б)* *в)*

Рис. 4.15. Пояснение к понятию категории стрелок

$$\begin{tikzcd}
A \arrow[rr, "k"] \arrow[dr, "f"'] & & B \arrow[dl, "g"] \\
& R &
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[rr, "k"] \arrow[dr, "f"'] & & B \arrow[dl, "g"] \\
& R &
\end{tikzcd}$$

а) *в)*

$$\begin{tikzcd}
A \arrow[rr, "l_{ok}"] \arrow[r, "k"] \arrow[dr, "f"'] & {} \arrow[r, "R"] \arrow[d, "g"] & {} \arrow[dl, "h"] \\
& R &
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "k"] \arrow[d, "f"'] & B \arrow[d, "g"] \\
R \arrow[r, "id"] & R
\end{tikzcd}$$

б) *г)*

Рис. 4.16. Пояснение к относительной категории

12. Относительная категория. Эту категорию можно рассматривать как специальный случай категории стрелок, когда накладывается ограничение, что начало и конец стрелки фиксированы. Если взять множество вещественных чисел R, то получим категорию вещественнозначных функций $\mathrm{Set}\downarrow R$. Ее объектами являются все функции из $f: A\to R$ с областью значений R. В качестве стрелок из $f: A\to R$ в $g: B\to R$ берутся функции $k: A\to B$, для которых коммутативен треугольник на рис. 4.16, *а*. Объекты категории $\mathrm{Set}\downarrow R$ иногда удобно представить в виде пар (A, f) для $f: A\to R$. Тогда композиция стрелок

$$(A,\ f)\xrightarrow{k}(B,\ g)\xrightarrow{l}(C,\ h)$$

в $\mathrm{Set}\downarrow R$ определяется как стрелка $l\circ k$: $(A,\ f)\to(C,\ h)$ (рис. 4.16, *б*). Единичной стрелкой объекта $f: A\to R$ будет стрелка $\mathrm{id}_A: (A,\ f)\to(A,\ f)$. Выведенная таким образом категория $\mathrm{Set}\downarrow R$ не является подкатегорией категории $\mathrm{Set}^{\to}$, так как в ней имеются совершенно другие стрелки. Однако $\mathrm{Set}\downarrow R$-стрелку $k: (A, f)\to(B, g)$ можно отождествить с $\mathrm{Set}^{\to}$-стрелкой $<k,\ \mathrm{id}_R>$, так как треугольник на рис. 4.16, *в* коммутативен тогда и только тогда, когда коммутативен квадрат на рис. 4.16, *г*. Это позволяет рассматривать категорию $\mathrm{Set}\downarrow R$ как (неполную) подкатегорию в $\mathrm{Set}^{\to}$.

Аналогично для любого множества X можно определить категорию $\mathrm{Set}\downarrow X$ функций со значением в X. В более общем случае K — некоторая категория и a — ее объект. Тогда объек-

$$\begin{tikzcd}
b \arrow[rr, "k"] \arrow[dr, "f"'] & & c \arrow[dl, "g"] \\
& g &
\end{tikzcd}
\qquad
\begin{tikzcd}
& a \arrow[dl, "f"'] \arrow[dr, "g"] & \\
b \arrow[rr, "k"'] & & c
\end{tikzcd}$$

а) *б)*

Рис. 4.17. Пояснение к общему определению относительных категорий

тами категории $k\downarrow a$ объектов над a являются все K-стрелки с концом в объекте a и стрелками из $f:b\to a$ в $g:c\to a$ являются такие K-стрелки $k:b\to c$, что треугольник на рис. 4.17, *а* коммутативен, т. е. $g\circ k=f$. Категории этого типа играют большую роль в теории топосов.

Если обратиться к началам стрелок, можно определить категорию $k\uparrow a$ объектов под a, объектами которых являются все K-стрелки с началом a со стрелками из $f:a\to b$ в $g:a\to c$ с такими K-стрелами $k:b\to c$, что треугольник на рис. 4.17, *б* коммутативен, т. е. $k\circ f=g$. Категории $k\downarrow a$ и $k\uparrow a$ называются относительными.

4.4. ДВОЙСТВЕННОСТЬ В ТЕОРИИ КАТЕГОРИЙ. ПРОИЗВЕДЕНИЯ

Каждую категорию K можно сопоставить с новой категорией K^0. Объекты и морфизмы категории K, рассматриваемые как объекты и морфизмы категории K^0, будут обозначаться соответственно как A^0, B^0... и α^0, β^0... Для любой пары объектов A^0, $B^0\in Ob\,K^0$ по определению $\mathrm{Mor}_{K^0}(A^0, B^0)=\mathrm{Mor}_k(B, A)$. Для любых морфизмов $\alpha^0:A^0\to B^0$ и $\beta^0:B^0\to C^0$ по определению

$$\alpha^0\beta^0=(\beta\alpha)^0.$$

Такая категория K^0 называется двойственной к категории K. Так, для категории примера 4.2, которая обозначается как KG, двойственной будет категория обращения выводов и разборов.

Рассмотрим подробнее понятие двойственности. Если Σ — предложение категориального языка, то двойственным назовем предложение, которое получается из Σ заменой dom на cod, а cod на dom и $h=g\circ f$ на $h=f\circ g$, т. е. все стрелки и композиции, входящие в Σ, поворачиваются в Σ^{op} в другую сторону. Понятия и конструкции, описываемые в Σ^{op}, называются двойственными к понятию (или конструкции), описываемому Σ.

Так, *эпистрелка* является двойственной к *монострелке*. Двойственным к начальному объекту является конечный объект и т. д. Стрелка $f:a\to b$ называется *мономорфной* или *монострелкой* в K, если для любой пары g, $h:c\rightrightarrows a$ K-стрелок из равенства $f\circ g=f\circ h$ следует $g=h$. Для указания того, что f мономорфна, будет использоваться обозначение $f:a\rightarrowtail b$.

Пример 4.4. Допустим, что функция $f:A\to B$ взаимно однозначна или инъективна, если не существует двух различных входов, дающих один и тот же выход, т. е. если для любых x, $y\in A$ из $f(x)=f(y)$ следует $x=y$.

Допустим, что $f:A\to B$ инъективна, а две параллельные функции g, $h:c\rightrightarrows A$ выбраны так, что диаграмма на рис. 4.18 коммутативна, т. е. $f\circ g=f\circ h$. Поэтому для всякого $x\in c$ имеем $g\circ g(x)=f\circ h(x)$, т. е. $f(g(x))=f(h(x))$. Но так как f инъективна, то $g(x)=h(x)$, т. е. g и h, имеющие один и тот же выход, при каждом входе совпадают. Тем самым показано, что на инъективные функции можно «сокращать слева», т. е. если $g\circ g=f\circ h$, то $g=h$. И наоборот, если f такова, что на f можно сократить слева, то она должна быть инъективна. Нетрудно убедиться, что в категории Set стрелки такие, что на них можно сокращать слева, т. е.

$$f:a\rightarrowtail b.$$

Пример 4.5. В категории N каждая стрелка мономорфна, сократимость слева означает, что из $m+n=m+p$ следует $n=p$, а это является очевидным фактом из теории сложения натуральных чисел.

Стрелка $f:a\to b$ называется *эпиморфной* или *эпистрелкой* (сократимость справа) в категории K, если для произвольной пары K-стрелок g, $h:b\rightrightarrows c$ из равенства $g\circ f=h\circ f$ следует $g=h$, т. е. всякий раз, когда диаграмма на рис. 4.19 коммутативна, справедливо $g=h$. Для эпистрелки используется обозначение $f:a\twoheadrightarrow b$. Функция $f:A\to B$ является *наложением*, или она *сюръективна*, если область ее значений B совпадает с множеством всех ее значений, т. е. для каждого $y\in B$ существует некоторый $x\in A$, такой, что $y=f(x)$, т. е. каждый элемент из

$$\begin{tikzcd}
C \arrow[r, "g"] \arrow[d, "h"'] & A \arrow[d, "f"] \\
A \arrow[r, "f"'] & B
\end{tikzcd}$$

Рис. 4.18. Пояснение к понятию «монострелка»

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[d, "f"'] & b \arrow[d, "g"] \\
b \arrow[r, "h"'] & c
\end{tikzcd}$$

Рис. 4.19. Пояснение к понятию «эпистрелка»

B появится в качестве выхода. Очевидно, что эпистрелка получается обращением моностредки. В категории Set стрелки являются сюръективными отображениями (эпистрелками). В категории N все стрелки эпиморфны, так как равенство $n+m=$ $=p+m$ влечет за собой $n=p$. В категории предкатегорий все стрелки *эпиморфны*. В этом смысле различия между моно- и эпи- в этих категориях отсутствуют, и их введение как бы излишне. Произвольная стрелка $f:a\to b$ называется *изострелкой* или *обратимой* K-стрелкой, если существует K-стрелка $g:b\to a$, такая, что $g\circ f=1_a$ и $f\circ g=1_b$.

Функция, которая является как инъективной, так и сюръективной, называется *биекцией*.

Пример 4.6. Если стрелка $f:p\to q$ из категории предпорядка, соответствующей ЧУМ $P=(P, <)$, имеет обратную стрелку $f^{-1}:g\to p$, то $p<g$ и $g<p$, поэтому в силу антисимметричности $p=g$. Но тогда f должна быть единственной стрелкой 1_p из pbg. Таким образом, в ЧУМ, рассматриваемом как категория (которая будет называться *категорией порядка*), каждая стрелка мономорфна и эпиморфна, но изострелками являются только единичные стрелки.

Объекты a и b называются *изоморфными* в K (символически это обозначается как $a\cong b$), если существует K-стрелка $f:a\to b$, являющаяся изострелкой в K, т. е. $f:a\cong b$.

В категории Set отношение $A\cong B$ имеет место, когда существует биекция между A и B, при наличии которой каждое из этих множеств можно представлять себе как переобозначение другого.

В категории Top изоморфные топологические пространства называются гомеоморфными. Это означает, что между ними существует гомеоморфизм, т. е. непрерывная биекция, для которой обратное отображение также непрерывно.

Теория категорий дает абстрактную формулировку идеи математического изоморфизма и исследует понятия, инвариантные при всех видах изоморфизма. В теории категорий слово «изоморфно» является синонимом слова «есть». Объект O называется начальным в категории K, если для каждого объекта из K существует одна и только одна K-стрелка из O в a. Объект 1 называется *конечным* объектом в категории K, если для каждого K-объекта a существует одна и только одна стрелка из a в 1.

Иначе говоря, в Set конечные объекты — это синглетоны, т. е. одноэлементные множества $\{e\}$.

Для данного множества A правило $f(x)=e$ определяет функцию $f:A\to\{e\}$. Так как e является ее единственным возможным значением, эта функция является единственной такой функцией. Таким образом, Set имеет много конечных объектов и все они изоморфны между собой (заметим, что конечные объекты изоморфны в любой категории). Представителем их является $a=$ $=\{o\}$ (отсюда и обозначение конечного объекта).

Очень часто для обозначения единственной существующей стрелки используется восклицательный знак. Так, единственную стрелку из O в a обозначим через $!: o\to a$. Такая стрелка будет обозначаться через O_a, т. е. $O_a\to a$.

Для данной категории K построим двойственную категорию K^{op} следующим образом. Категории K и K^{op} имеют одни и те же объекты. Для каждой K-стрелки $f:a\to b$ вводится K-стрелка $f^{op}:b\to a$ (своя для каждой f). Эти стрелки исчерпывают все стрелки категории K^{op}. Композиция $f^{op}\circ g^{op}$ определена тогда и только тогда, когда в K определена композиция

$$g\circ f \quad \text{и} \quad f^{op}\circ g^{op}=(g\circ f)^{op}.$$

Очевидно, что $\operatorname{dom} f^{op}=\operatorname{cod} f$ и $\operatorname{cod}^{op}=\operatorname{dom} f$ (рис. 4.20).

Пример 4.7. В дискретной категории K

$$K^{op}=K.$$

Пример 4.8. Если K — категория предпорядка (P, K), где $R\subseteq P\times P$, то K^{op} есть категория предпорядка (P, R^{-1}), в которой $pR^{-1}q$ тогда и только тогда, когда qRp, т. е. R^{-1} — инверсия отношения R.

Пример 4.9. Для произвольной категории K справедливо

$$(K^{op})^{op}=K.$$

Иначе говоря, конструкцию, двойственную к выражаемой произведением Σ, можно интерпретировать как первоначальное построение, примененное к двойственной категории. Если Σ истинно в K, то Σ^{op} истинно в K^{op}. К примеру, начальный объект $\varnothing$ в Set является конечным объектом в Set^{op}. Если Σ — теорема в теории категорий, т. е. предложение выводимо из категорных аксиом, то Σ истинно во всех категориях. Поэтому Σ^{op} будет иметь во всех категориях вид K^{op}. Но нетрудно убедиться, что произвольная категория D имеет такой же вид (если положить $K=D^{op}$), поэтому Σ^{op} имеет место во всех категориях. Таким образом, из произвольного истинного в теории

$$\begin{tikzcd}
a \arrow[r, "f"] & b \arrow[l, bend left, "f^{op}"] \arrow[r, "g"] & c \arrow[l, bend left, "g^{op}"]
\end{tikzcd}$$

Рис. 4.20. Пояснение к понятию «двойственная категория»

$$\begin{tikzcd}
 & C \arrow[dl, "f"'] \arrow[d, dashed, "p"] \arrow[dr, "g"] & \\
A & A\times B \arrow[l, "p_A"'] \arrow[r, "p_B"] & B
\end{tikzcd}$$

Рис. 4.21. Категориальное представление произведений множеств категорий предложения получается другое истинное предложение Σ^{op}. В этом заключается принцип двойственности, сокращающий число доказательств вдвое. К примеру, понятие изоморфизма самодвойственно. Двойственная к обратимой стрелке снова обратимая стрелка: $(f^{op})^{-1}=(f^{-1})^{op}$, если доказано, что «два произвольных начальных K-объекта изоморфны», то сразу можно утверждать, что верно и двойственно положение «два произвольных конечных K-объекта изоморфны».

Произведения в теории категорий. Вначале дадим с помощью стрелок категориальное описание и свойства понятия произведения в теории множеств.

Известно, что произведение двух множеств A и B определяется как

$$A\times B=\{\langle x,\ y\rangle : x\in A \quad \text{и} \quad y\in B\}.$$

Это определение представим с помощью стрелок без процедуры упорядочения пар (что можно сделать с точностью до изоморфизма), что иногда называется конструкцией в теории категорий.

Поставим в соответствие произведению $A\times B$ два специальных отображения (проекции)

$$p_A : A\times B\to A;$$
$$p_B : A\times B\to B,$$

которые задаются равенствами

$$p_A(\langle x,\ y\rangle)=x; \qquad p_B(\langle x,\ y\rangle)=y.$$

Допустим, что задано еще одно множество C с парой отображений $f : C\to A$; $g : C\to B$. Определим отображение $p : C\to A\times B$ правилом $p(x)=<f(x),\ g(x)>$, тогда $p_A(p(x))=f(x)$ и $p_B\times(p(x))=g(x)$ для каждого $x\in C$. Таким образом, $p_A\circ p=f$ и $p_B\circ p=g$, т. е. диаграмма, приведенная на рис. 4.21, коммутативна. Кроме того, p является единственной стрелкой, для которой эта диаграмма коммутативна.

Действительно, если $p(x)=<y,\ z>$, то в силу условия $p_A\circ p=f$ будет $p_A(p(x))=f(x)$, т. е. $y=f(x)$. Аналогично, если $p_B\circ p=g$, то $z=g(x)$. Отображение p, построенное по f и g, обозначается через $<f,\ g>$ и называется произведением отображений f и g. В категории Set оно определяется равенством

$$\langle f,\ g\rangle(x)=\langle f(x),\quad g(x)\rangle.$$

После этих предварительных замечаний дадим определение произведения. *Произведением* в категории K двух объектов a и b называется K-объект, обозначаемый через $a\times b$ вместе с парой $(pr_a : a\times b\to a,\ pr_b : b\to a)$ K-стрелок. Существует одна и

только одна стрелка $<f,\ g> : c\to a\times b$, для которой диаграмма на рис. 4.21 коммутативна, т. е.

$$pr_a\circ\langle f,\ g\rangle=f \quad \text{и} \quad pr_b\circ\langle f,\ g\rangle=g.$$

Стрелка $<f,\ g>$ называется *произведением стрелок* f и g относительно *проекций* pr_a, pr_b.

Покажем, что *произведение определено с точностью до изоморфизма.* Допустим, что K-объект d вместе с парой $(p : d\to a;\ g : d\to b)$ также удовлетворяет приведенному определению произведения двух объектов a и b. Тогда для диаграммы на рис. 4.22 при данных проекциях $pr_a : a\times b\to a$ и $pr_b : a\times b\to b$ стрелка $<p,\ q>$ однозначно определена. Стрелка $<pr_a,\ pr_b>$ также является однозначно определенным произведением стрелок pr_a и pr_b относительно $p : d\to a$ и $d\to b$. Так как d является произведением объектов a и b, то существует только одна стрелка $S : d\to d$, для которой диаграмма на рис. 4.23 коммутативна. Очевидно, что при $S=1a$ эта диаграмма коммутативна. Из коммутативности диаграммы на рис. 4.22 диаграмма на рис. 4.23 также будет коммутативна при $S=<pr_a,\ pr_b>\circ<p,\ q>$. В силу единственности S справедливо, что

$$\langle pr_a,\quad pr_b\rangle\circ\langle p,\ q\rangle=1_d.$$

Меняя ролями d и $a\times b$, приходим к аналогичному равенству

$$\langle p,\ q\rangle\circ\langle pr_a,\quad pr_b\rangle=1a\times b.$$

Таким образом, $<p,\ q> : d\cong a\times b$. Композиция изострелки $<p,\ q>$ с проекциями для $<a\times b>$ дает проекции для d, как его поясняет диаграмма на рис. 4.22. На самом деле $<p,\ q>$ является единственной стрелкой $d\to a\times b$ с этим свойством. Таким образом, произведение объектов a и b определено и однозначно с точностью до изоморфизма.

$$\begin{tikzcd}
 & d \arrow[dl] \arrow[dr] \arrow[d, dashed, "<p,q>"] & \\
a \arrow[dr] & {} \arrow[l, "pr_a"'] \arrow[r, "pr_b"] \arrow[d, dashed, "<pr_a,pr_b>"] & b \arrow[dl] \\
 & d &
\end{tikzcd}$$

Рис. 4.22. Пояснение к изоморфизму произведений

$$\begin{tikzcd}
 & d \arrow[dl, "p"'] \arrow[dr, "q"] \arrow[d, dashed, "s"] & \\
a & d \arrow[l, "p"'] \arrow[r, "q"] & b
\end{tikzcd}$$

Рис. 4.23. Пояснение к понятию произведения стрелок

Пример 4.10. Если A и B — конечные множества с m и n элементами соответственно, то произведение $A\times B$ насчитывает $m\times n$ элементов. Отсюда в категории Finord произведение ординальных чисел m и n существует и совпадает с ординалом $m\times n$.

Произведение отображений (функторное произведение). Для двух функций $f:A\to B$ и $g:C\to D$ определяем функцию $f\times g$ из $A\times C$ в $B\times D$ равенством

$$f\times g(\langle x,\ y\rangle)=\langle f(x),\quad g(y)\rangle.$$

Очевидно, что $f\times g$ является произведением двух композиций:

$$f\circ p_A: A\times C\to A\to B;$$

$$g\circ p_C: A\times C\to C\to D.$$

Поэтому можно сделать следующее определение: если $f:a\to b$ и $g:c\to d$ — две K-стрелки, то через $f\times g: a\times c\to b\times d$ обозначим K-стрелку $<f\circ pr_a,\ g\circ pr_b>$ (рис. 4.24). Очевидно, что стрелка $f\times g$ определена только тогда, когда в K существуют произведения $a\times c$ и $b\times d$.

Конечные произведения. Прежде всего распространим понятие произведения множеств на случай трех сомножителей, определив $A\times B\times C$ как множество упорядоченных троек $<x,\ y,\ z>$, в которых на первом месте стоят элементы из A, на втором — из B, на третьем — из C. Сокращенно это запишется в виде

$$A\times B\times C=\{\langle x,\ y,\ z\rangle : x\in A,\quad y\in B,\quad z\in C\}.$$

Аналогично определяется $A_1\times A_2\times\ldots\times A_m$, $\{\{<x_1,\ \ldots,\ x_m>\}: :x_1\in A_1,\ x_2\in A_2,\ldots x_m\in A_m\}$ для всех m-к, т. е. для всех последовательностей длины m, i-е члены которых принадлежат A_i. Частным случаем этого произведения является m-кратное произведение множества A на себя

$$A^m=\{\langle x_1,\ .\ .\ .\ ,\ x_m\rangle : x_1,\ x_2,\ .\ .\ .\ ,\ x_m\in A\},$$

т. е. множество всех m-к, члены которых принадлежат A. Поставим в соответствие множеству A^m m различных отображений проектирования $pr_1{}^m$, ..., $pr_m{}^m$ из A^m в A, задаваемых правилами

$$pr_1^m(\langle x_1,\quad .\ .\ .\ ,\quad x_m\rangle)=x_1;$$

$$pr_2^m(\langle x_1,\quad .\ .\ .\ ,\quad x_m\rangle)=x_2;$$

$$.\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .\ .$$

$$pr_m^m(\langle x_1,\ .\ .\ .\ .\ ,\quad x_m\rangle)=x_m.$$

$$\begin{tikzcd}
& c \arrow[r, "g"] & d \\
a\times b \arrow[ur, "pr_c"] \arrow[dr, "pr_a"'] \arrow[rr, dashed, "{<f\circ pr_a, g\circ pr_c>}"] & & b\times d \arrow[u, "pr_d"'] \arrow[d, "pr_b"] \\
& a \arrow[r, "f"] & b
\end{tikzcd}$$

Рис. 4.24. Пояснение к понятию произведения отображений (функторного произведения)

$$\begin{tikzcd}
& & c \arrow[dll, "f_1"'] \arrow[dl, "f_2"] \arrow[dd, dashed, "{<f_1,\cdots,f_m>}"] \arrow[dr, "f_m"] & \\
a & a & \cdots & a \\
& & a^m \arrow[ull, "pr_1^m"] \arrow[ul, "pr_2^m"'] \arrow[ur, "pr_m^m"'] &
\end{tikzcd}$$

Рис. 4.25. Пояснение к определению конечного произведения стрелок

Для произвольных множеств C и m отображений $f_1: C\to A$, ..., $f_m: C\to A$ можно определить отображение $<f_1,\ \ldots,\ f_m>$ из C в A^m, полагая

$$\langle f_1,\ .\ .\ .\ .\ ,\ f_m\rangle(c)=\langle f_1(c),\ f_2(c),\ .\ .\ .\ .\ ,\ f_m(c)\rangle$$

для каждого $c\in C$. Рассмотренная конструкция существует в любой категории K, в которой существует произведение любых двух K-объектов. Для данного K-объекта a определим m-кратное произведение объекта на себя равенством

$$a^m=\underbrace{a\times a\times\ .\ .\ .\ \times a.}_{m\ \text{раз}}$$

Применяя определение произведения пары объектов к образованию a^m, можно показать, что возникает m стрелок $pr_1{}^m: a^m\to a$, ..., $pr_m{}^m: a^m\to a$ со свойством универсальности, которое состоит в том, что для произвольных K-стрелок $f_1: c\to a$, ..., $f_m: c\to a$, имеющих общее начало, существует одна и только одна стрелка $<f_1,\ \ldots,\ f_m>: c\to a^m$, для которой диаграмма на рис. 4.25 коммутативна. Для $m=1$ в качестве a^1 возьмем a, а $pr_1{}^1: a\to a$ положим равной 1_a. Конечные произведения играют существенную роль в моделировании семантики и категориальном описании тензоров.

Копроизведения. Понятие копроизведения (или суммы) объектов является двойственным к понятию произведения. Его определение получается из определения произведения, если использовать принцип двойственности.

Копроизведением b категории K двух объектов a и b называется K-объект, обозначаемый через $a+b$, вместе с парой $(i_a: a\to a+b,\ i_b: b\to a+b)$ K-стрелок, такой, что для произволь-

$$\begin{tikzcd}
a \arrow[r, "i_a"] \arrow[dr, "f"'] & a+b \arrow[r, "i_b"] \arrow[d, dashed, "{[f,g]}"] & b \arrow[dl, "g"] \\
& c &
\end{tikzcd}$$

Рис. 4.26. Пояснение к определению копроизведения объектов

$$\begin{tikzcd}
E \arrow[rr, hook, "i"] & & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \\
& c \arrow[ul, dashed, "k"] \arrow[ur, "h"'] & &
\end{tikzcd}$$

а)

$$\begin{tikzcd}
o \arrow[rr, "f"] & & a \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & b \\
& c \arrow[ul, dashed] \arrow[ur] & &
\end{tikzcd}$$

б)

Рис. 4.27. Пояснение к понятию «уравнители»

ной пары ($f:a\to c$, $g:b\to c$) K-стрелок существует одна и только одна стрелка $[f, g]:a+b+c$, для которой диаграмма на рис. 4.26 коммутативна, т. е. $[f, g]\circ i_a=f$ и $[f, g]\circ i_b=g$. Стрелка $[f, g]$ называется копроизведением стрелок f и g относительно инъекций i_a и i_b. В категории Set копроизведением объектов A и B является дизъюнктивное объединение $A+B$ (см. гл. 7 по алгебре и исчислению нечетких множеств), т. е. объединение двух множеств, изоморфных A и B соответственно, но не пересекающихся, т. е. пусть

$$A'=\{\langle a, 0\rangle : a\in A\}=A\times\{0\};$$

$$B'=\{\langle b, 1\rangle : b\in B\}=B\times\{1\}.$$

Положим $A+B=A'\cup B'$. Тогда инъекции $i_A:A\to A+B$ и $i_B:B\to A+B$ определяются правилами

$$i_A(a)=\langle a, 0\rangle; \qquad i_B(b)=\langle b, 1\rangle,$$

соответственно.

Дизъюнктивное объединение двух конечных множеств, содержащих m и n элементов соответственно, содержит $m+n$ элементов. В категории Finord копроизведение объектов m и n совпадает с ординальным числом $m+n$. Для ординалов $1=\{0\}$ и $2=\{0,1\}$ скелетальной категории Finord имеет место равенство $1+1=2$, тогда как в категориях Finset и Set в точном смысле справедливо $1+1\cong 2$ (причем копроизведения определены с точностью до изоморфизма).

Уравнители. Пусть $f, g:A\rightrightarrows B$ — пара «параллельных» функций в категории Set; E — подмножество в A, состоящее из всех элементов, на которых f и g совпадают, т. е.

$$E=\{x : x\in A \quad \text{и} \quad f(x)=g(x)\}.$$

Тогда функция включения E в A, обозначаемая i, называется *уравнителем* функций f и g. Это название определяется тем, что

при композиции этих двух функций с i получается равенство $f\circ i=g\circ i$, т. е. i уравнивает эти функции. Говорят даже, что i является «каноническим» уравнителем для f и g. Это означает, что если $h:c\to A$ является произвольным «уравнителем», т. е. $f\circ h=g\circ h$, то h однозначно «пропускается» через i, являющуюся функцией включения E в A, т. е. существует единственная функция $k:c\to E$, такая, что $i\circ k=h$ или иначе: для любой данной стрелки h существует единственная Set стрелка, подстановка которой в диаграмму на рис. 4.27, *а* вместо штриховой линии делает ее коммутативной. Можно показать, что существует не более одной такой стрелки. Действительно, если $i\circ k$ совпадает с h, то для всякого $c\in C$ имеет место равенство $i(k(c))=h(c)$, т. е. $k(c)=h(c)$ (так как i — включение). Из последнего равенства также следует, как надо определить функцию k. Так как

$$f(h(c))=g(h(c)), \quad \text{то} \quad h(c)\in E.$$

Дадим определение уравнителя в более общем случае.

Стрелка $i:e\to a$ из категории K называется *уравнителем* пары $f, g:a\to b$ K-стрелок, если:

а) $f\circ i=g\circ i$;

б) для любой K-стрелки $h:c\to a$, удовлетворяющей равенству $f\circ h=g\circ h$, существует и притом только одна K-стрелка $K:c\to e$, такая, что $i\circ k=h$ (рис. 4.27). Произвольная стрелка называется *уравнителем* в K, если существует пара K-стрелок, уравнителем которой она является. Можно доказать [46] теорему: «Всякий уравнитель является монострелкой». Обратная теорема несправедлива в общем случае. Например, в категории N все стрелки мономорфны, в частности единица является мономорфной. Однако она не может быть уравнителем никакой пары m, n стрелок. Если бы это было справедливо, то имело бы место $m\circ 1=n\circ 1$, т. е. $m+1=n+1$. Поэтому $m=n$. Но тогда $m+0=n+0$ и k должен был бы пропускаться через 1, т. е. существовало бы k, удовлетворяющее равенству $1+k=0$. Но такого натурального числа K не существует. Заметим, что в категории N каждая стрелка аморфна. Вместе с тем 0 является единственной изострелкой. По этому поводу можно доказать [46]: «В произвольной категории эпиморфный уравнитель является изострелкой». В общем случае монострелки могут не быть уравнителями, но в категории Set (и во всех топосах) они являются уравнителями. Действительно, пусть $f:E\to A$ — инъективная функция. Определим $h:A\to\{0, 1\}$ равенством $h(x)=1$ для всех $x\in A$, а функцию $g:A\to\{0, 1\}$ зададим условиями

$$g(x)=\begin{cases} 1, & \text{если} \quad x\in \operatorname{Im} f; \\ 0, & \text{если} \quad x\notin \operatorname{Im} f. \end{cases}$$

Тогда f будет уравнителем пары g и h.

Пределы и копределы. Прежде всего заметим, что определения произведения двух объектов и уравнителя пары стрелок имеют одну и ту же форму. Как произведение, так и уравнители обладают свойством каноничности в том смысле, что через них «пропускаются» указанным выше способом любые другие объекты с такими же свойствами. В случае уравнителя это свойство состоит в уравнивании двух данных стрелок. В случае произведения объектов a и b это свойство заключается в том, чтобы быть началом пары стрелок, концами которых являются a и b. Рассмотренные положения определяют так называемую универсальную *конструкцию*. Определяемый объект является *универсальным* среди объектов, обладающих некоторым свойством.

Конусом для диаграммы D называется такой K-объект c вместе с K-стрелками $f_i: c \to d_i$ для каждого объекта d_i из D, что диаграмма на рис. 4.28 коммутативна для любой стрелки g из G. Конус для диаграммы D будет обозначаться через $\{f_i: c \to d_i\}$. *Пределом* диаграммы D называется D-конус $\{f_i: c \to d_i\}$, такой, что для любого другого D-конуса $\{f_i': c' \to d_i\}$ существует одна и только одна стрелка $f: c' \to c$, для которой диаграмма на рис. 4.29 коммутативна при каждом объекте d_i из D. Предельный конус обладает *свойством универсальности* относительно всех D-конусов. *Он универсален* среди D-конусов. Любой другой D-конус однозначно пропускается через универсальный так, как показано на рис. 4.29. Предел диаграммы D единствен с точностью до изоморфизма: если $\{f_i: c \to d_i\}$ и $\{f_i': c' \to d_i\}$ являются пределами диаграммы D, то однозначно определенная выше стрелка $f: c' \to c$ является изострелкой. При этом надо иметь в виду, что существование обратной к ней стрелки $c \to c'$ следует из того факта, что $\{f_i': c' \to d_i\}$ является пределом.

Пример 4.11. Для данных объектов a и b пусть D является бесстрелочной диаграммой ab. Тогда D-конусом является объект c вместе с двумя стрелками $f: c \to a$, $g: c \to d$ (рис. 4.30). Предельным D-конусом, через который пропускается любой D-конус, является произведение объектов a и b.

Пример 4.12. Пусть D-диаграмма имеет вид

$$a \overset{f}{\underset{g}{\rightrightarrows}} b,$$

тогда D-конус — это пара $h: c \to a$, $j: c \to b$, для которой диаграммы на рис. 4.31 коммутативны. Но отсюда следует, что $j = f \circ h = g \circ h$. Поэтому можно утверждать, что D-конус в этом случае является стрелкой $h: c \to a$, для которой диаграмма

$$c \xrightarrow{h} a \overset{f}{\underset{g}{\rightrightarrows}} b$$

коммутативна, т. е. $f \circ h = g \circ h$. Таким образом, D-предел — это уравнитель пары стрелок f и g.

$$\begin{tikzcd}
d_i \arrow[rr, "g"] & & d_j \\
& c \arrow[ul, "f_i"] \arrow[ur, "f_j"'] &
\end{tikzcd}$$

Рис. 4.28. К определению конуса диаграммы

$$\begin{tikzcd}
& d_i \arrow[dr, "f_i"] & \\
c' \arrow[ur, "f_i'"] \arrow[rr, dashed, "f"'] & & c
\end{tikzcd}$$

Рис. 4.29. К определению предела диаграммы

$$\begin{tikzcd}
a & & b \\
& c \arrow[ul, "f"] \arrow[ur, "g"'] &
\end{tikzcd}$$

Рис. 4.30. Пояснение к примеру 4.11

Пример 4.13. Пусть D — пустая диаграмма, т. е. диаграмма без объектов и стрелок; D-конусом будет K-объект c, нет никаких f_i, так как D не имеет d_i. Предельный конус — это такой объект c, что для любого другого K-объекта (D-конуса) c' существует единственная стрелка $c' \to c$. Это означает, что пределом для пустой диаграммы является конечный объект. Можно двойственным образом определить *конус* $\{f_i: d_i \to c\}$ для диаграммы D, который состоит из объекта c и стрелок $f_i: a \to c$, по одной для каждого объекта d_i из D, удовлетворяющим соответствующим условиям коммутативности.

Копредел для D — это коконус $\{f_i: d_i \to c\}$ со свойством коуниверсальности, которое заключается в том, что для любого коконуса $\{f_i': d_i \to c'\}$ существует единственная стрелка $f: c \to c'$, такая, для которой для каждого d_i из D диаграмма на рис. 4.32 коммутативна. Копределом для диаграммы на рис. 4.30 является копроизведение объектов a и b, в то время как копредел пустой диаграммы — это *начальный объект*.

Коуравнители. Коуравнителем пары параллельных K-стрелок f, g называется копредел диаграммы

$$a \overset{f}{\underset{g}{\rightrightarrows}} g.$$

Коуравнитель можно рассматривать как такую K-стрелку $g: b \to e$, что:

$$\begin{tikzcd}
a \arrow[rr, "f"] & & b \\
& c \arrow[ul, "h"] \arrow[ur, "j"'] &
\end{tikzcd}$$

а)

$$\begin{tikzcd}
a \arrow[rr, "g"] & & b \\
& c \arrow[ul, "h"] \arrow[ur, "f"'] &
\end{tikzcd}$$

б)

Рис. 4.31. Пояснение к примеру 4.12

$$\begin{tikzcd}
& d_i \arrow[dl, "f_i"'] \arrow[dr, "f_i'"] & \\
c \arrow[rr, dashed, "f"] & & c'
\end{tikzcd}$$

$$\begin{tikzcd}
a \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & b \arrow[r, "g"] \arrow[dr, "h"'] & {} \arrow[d, dashed, "k"] \\
& & c
\end{tikzcd}$$

Рис. 4.32. К определению копредела

Рис. 4.33. К определению коуравнителя

а) $g\circ f=g\circ g$;

б) для любой стрелки $h: b\to c$, удовлетворяющей равенству $h\circ f=h\circ g$, существует единственная стрелка $k: e\to c$, для которой диаграмма на рис. 4.33 коммутативна.

Если применить свойства двойственности к понятиям, связанным с уравнителями, то можно утверждать, что коуравнитель является эпистрелкой, обратное утверждение справедливо в Set и мономорфный коуравнитель является изострелкой.

В категории Set уравнители описываются в терминах отношения с помощью очень важного отношения *эквивалентности*. *Отношение эквивалентности на множестве* A — это, по определению, отношение $R\subseteq A\times A$, обладающее следующими свойствами:

а) *рефлективностью*, т. е. aRa для каждого $a\in A$;

б) *транзитивностью*, т. е. если aRb и bRc, то aRc для любых a, b, c из A.

в) *симметричностью*, т. е. если aRb, то bRa для любых a и b из A.

Отношение, которое возникает между двумя объектами, когда с точки зрения их свойств они неразличимы, будет отношением эквивалентности. Ранее это понятие встречалось при обсуждении изоморфных объектов. Два изоморфных объекта можно в некоторой категории рассматривать как один объект во всем, что касается категориальных свойств, отношение

$$\{\langle a,\ b\rangle : a\cong b \quad \text{в} \quad K\}$$

на K-объектах является рефлексивным, транзитивным и симметричным.

Свойство эквивалентности широко используется в кибернетике, в частности в системах искусственного интеллекта при сопоставлении, в программных системах при оценке качества программы и т. д. Процесс отождествления эквивалентных вещей предметной области воплощается в объединении в одну вещь всех вещей, связанных друг с другом отношением эквивалентности (см. метод резольвенций, унификации в теории доказательства, в теории искусственного интеллекта). Возникающие совокупности рассматриваются при этом как новые единые сущности.

Определим формально класс R-эквивалентности для данного $a\in A$ как множество

$$[a]=\{b: aRb\}$$

всех элементов из A, находящихся в R-отношении к a. Одно и то же множество может быть классом эквивалентности различных элементов.

В общем случае:

1) $[a]=[b]$ тогда и только тогда, когда aRb;

2) если $[a]\neq[b]$, то $[a]\cap[b]=\varnothing$;

3) $a\in[a]$.

Утверждение 1 означает, что два эквивалентных элемента находятся в отношении R с одним и тем же множеством элементов. Положение 2 утверждает, что два различных класса эквивалентности не имеют общих элементов. Из этого положения вместе с утверждением с 3 (которое следует из свойства рефлексивности a) следует, что каждый $a\in A$ является элементом одного и только одного класса R-эквивалентности.

Процесс отождествления состоит в переходе от данного множества к новому, элементами которого являются классы R-эквивалентности, т. е. переходим от множества A к множеству

$$A|R=\{[a]: a\in A\}.$$

Этот переход осуществляется при помощи *естественного отображения* $f_R: A\to A/R$, где $f_R(a)=[a]$ для $a\in A$. Поэтому, если aRb, то $f_R(a)=f_R(b)$, т. е. функция f_R отождествляет R-эквивалентные элементы. Оказывается, что f_R является коуравнителем пары f, $g: R\rightrightarrows A$ функций проектирования из R в A, т. е. функций, задаваемых равенствами

$$f(\langle a,\ b\rangle)=a \quad \text{и} \quad g(\langle a,\ b\rangle)=b.$$

Покажем, что диаграмма на рис. 4.34 может быть пополнена только единственной стрелкой k при заданной стрелке h, удовлетворяющей равенству $h\circ f=h\circ g$. Для этого предположим, что $k\circ f_R=h$. Тогда для любого $[a]\in A|R$ имеем

$$k([a])=k(f_R(a))=k\circ f_R(a)=h(a).$$

Таким образом, не имеется выбора в определении значения функции K в точке $[a]$. Единственная проблема состоит в том, будет ли определение этой функции с помощью равенства $k[a]=h[a]$ *корректным*. Если $[a]=[b]$, то значение функции K в точке $[a]=[b]$ должно быть также равно $h(b)$. Для обеспечения единственности значения функции при данном значении аргумента надо бы знать, что в этом случае $h(a)=h(b)$. Но это так, так как если $[a]=[b]$, то $\langle a, b\rangle\in R$ и равенство $h(a)=h(b)$ следует из условия $h\circ f=h\circ g$, налагаемого на функцию h.

$$\begin{tikzcd}
R \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & A \arrow[r, "f_g"] \arrow[dr, "h"'] & A/R \arrow[d, dashed, "k"] \\
 & & B
\end{tikzcd}$$

Рис. 4.34. Пояснение к соотношению $f_R\circ f=f_R\circ g$

Вопрос корректности определения часто встает при рассмотрении так называемых фактор-множеств A/R. Операции на классах R-эквивалентности, а также свойства этих классов определяются с помощью некоторых элементов этих классов эквивалентности, т. е. через их *представителей*, при этом всегда следует убедиться, что определение не зависит от выбора представителя. Другими словами, корректно определенные понятия — это такие понятия, которые устойчивы или инвариантны относительно R, т. е. не изменяются при замене некоторых элементов на R-эквивалентные.

Построим коуравнители в Set с помощью отношения эквивалентности. Для построения коуравнителя пары функций $f, g : A \rightrightarrows B$ необходимо отождествить $f(x)$ с $g(x)$ для $x\in A$. Поэтому рассмотрим отношение

$$S=\{\langle f(x),\quad g(x)\rangle : x\in A\}\subseteq B\times B.$$

Отношение S не может быть отношением эквивалентности на B. Однако можно расширить S до отношения эквивалентности и причем минимальным образом. Существует отношение эквивалентности R на B, такое что:

1) $S\subseteq R$;

2) если T — любое другое отношение эквивалентности на B, такое, что S содержится в T, то $R\subseteq T$ (т. е. R — наименьшее отношение эквивалентности на B, содержащее S) и естественное отображение $f_R : B\to B/R$ будет коуравнителем f и g.

Обратный образ. Обратным образом пары $a\xrightarrow{f}c\xleftarrow{g}b$ K-стрелок с общим концом называется K-диаграмма на рис. 4.35, *а*. Конус этой диаграммы состоит из трех стрелок f', h, g', для которых коммутативна диаграмма на рис. 4.35, *б*, т. е. $h=g\circ f'=f\circ g'$. Поэтому можно сказать, что конус — это пара $a\xleftarrow{g'}d\xleftarrow{f'}b$ K-стрелок, для которых квадрат (рис. 4.35, *в*) коммутативен, т. е. $f\circ g'=g\circ f'$. Таким образом, по определению универсального конуса обратным образом пары

$$a\xrightarrow{f}c\xleftarrow{g}b$$

будет пара K-стрелок $a\xleftarrow{g'}c\xrightarrow{f'}b$, обладающая следующими свойствами:

а) $f\circ g'=g\circ f'$;

б) для любых $a\xleftarrow{h}e\xrightarrow{j}b$, таких, что $f\circ h=g\circ j$, существует и притом только одна стрелка $k : e\to d$, удовлетворяющая равенствам $h=g'\circ k$ и $j=f'\circ k$ (рис. 4.36).

Иначе: для любых h и j, для которых внешний квадрат (т. е. граница приведенной на рис. 4.36 диаграммы) коммутативен, имеется одно и только одно пополнение этой диаграммы стрелкой K, при котором вся диаграмма становится коммутативной. Внутренний квадрат (b, g, f', g') диаграммы на рис. 4.36 называется *декартовым квадратом*. Говорят, что f' — *обратный образ* f *относительно* g и что f' получается подъемом f вдоль g, а g' — *обратный* образ g относительно f и получается *подъемом* g *вдоль* f. Понятие обратного образа включает в себя очень многие математические конструкции, используемые, в частности, в топосах.

$$\begin{tikzcd}
 & b \arrow[d, "g"] \\
a \arrow[r, "f"] & c
\end{tikzcd}$$

а)

$$\begin{tikzcd}
d \arrow[r, "f'"] \arrow[d, "g'"'] \arrow[dr, "h"] & b \arrow[d, "g"] \\
a \arrow[r, "f"] & c
\end{tikzcd}$$

б)

$$\begin{tikzcd}
d \arrow[r, leftrightarrow, "f'"] \arrow[d, "g'"'] & b \arrow[d, "g"] \\
a \arrow[r, "f"] & c
\end{tikzcd}$$

в)

Рис. 4.35. Пояснение к понятию обратного образа

Пример 4.14. В Set обратный образ (рис. 4.37) двух функций f и g может быть определен с помощью равенств

$$D=\{\langle x,\ y\rangle\}:\ x\in A,\ y\in B \text{ и } f(x)=g(y);$$

$$f'(\langle x,\ y\rangle)=y;\quad g'(\langle x,\ y\rangle)=x.$$

Поэтому D является подмножеством произведения $A\times B$, а f' и g' — функции проектирования. Иногда D обозначается через $A\underset{C}{\times}B$ и называется произведением *A и B над C.*

Иногда обратные образы называются *расслоенными произведениями.*

Пример 4.15. Прообразы. Если $f : A\to B$ — произвольная функция и C — некоторое подмножество в B, то прообразом множества C при отображении A, который обозначается через $f^{-1}(C)$, называется подмножество в A, состоящее из всех трех элементов $x\in A$, для которых $f(x)\in C$, т. е.

$$f^{-1}(C)=\{x:\ x\in A \text{ и } f(x)\in C\}.$$

Диаграмма на рис. 4.38 является *декартовым квадратом* в категории Set. На этой диаграмме стрелки с загнутыми концами обозначают, как обычно, включения, а f^* определяется равенством $f^*(x)=f(x)$ для $x\in f^{-1}$ (C)

$$\begin{tikzcd}
e \arrow[rrd, "j"] \arrow[rd, dashed, "k"] \arrow[rdd] & & \\
& d \arrow[r, "f'"] \arrow[d, "g'"] & b \arrow[d, "g"] \\
& a \arrow[r, "f"] & c
\end{tikzcd}$$

Рис. 4.36. К определению обратной пары *K*-стрелок

[т. е. f^* — ограничение функции f на $f^{-1}(C)$], т. е. прообраз C при отображении f получается подъемом C вдоль f (рис. 4.39). В этом примере наблюдаются преимущества рассмотрения (в теории категорий) функции множества пар как метода, хорошо описывающего динамику.

Пример 4.16. Ядро или ядерное отношение. В соответствие каждой функции $f: A \to B$ ставится отношение эквивалентности на A, которое обозначается через R_f и называется *ядерным отношением* или ядром функции f. Напомним, что понятие ядра лежит в основе первой теоремы об изоморфизме из универсальной алгебры. Положим

$$R_f = \{\langle x,\ y\rangle:\ x \in A \text{ и } y \in A \text{ и } f(x) = f(y)\}$$

или xR_fy, если и только если $f(x) = f(y)$.

В соответствии с примером 4.14 (см. рис. 4.37) убеждаемся, что квадрат на рис. 4.40, где $p_1(<x,\ y>)x$ и $p_2(<x,\ y>) = y$, декартов, т. е. R_f получается подъемом f вдоль самого себя. [Этот пример будет использован в дальнейшем для понимания «эпимоноразложения» стрелок в топосе (см. § 4.6).]

Пример 4.17. Ядра. Пусть $f: M \to N$ — гомоморфизм моноидов и

$$k = \{x:\ f(x) = e\}$$

— ядро этого гомоморфизма. Тогда в категории Mon квадрат на рис. 4.41 декартов, где O — одноэлементный моноид (являющийся одновременно начальным и конечным объектом). Эта характеризация ядер применима также к категориям Grp и Vect.

Пример 4.18. В категории предпорядка $(P, <)$ квадрат на рис. 4.42 является декартовым тогда и только тогда, когда s является копроизведением p и q.

$$\begin{tikzcd}
D \arrow[r, "f'"] \arrow[d, "g'"] & B \arrow[d, "g"] \\
A \arrow[r, "f"] & C
\end{tikzcd}$$

Рис. 4.37. Обратный образ в Set

$$\begin{tikzcd}
f^{-1}(C) \arrow[r, "f^*"] \arrow[d, hook] & C \arrow[d, hook] \\
A \arrow[r, "f"] & B
\end{tikzcd}$$

Рис. 4.38. К определению понятия декартова квадрата в категории Set

Пример 4.19. В произвольной категории с конечным объектом, если квадрат (рис. 4.43) декартов, то $(f,\ g)$ будут произведением a и b.

Пример 4.20. В любой категории, если квадрат (рис. 4.44) декартов, то i — уравнитель для f и g.

Пример 4.21. Лемма о квадратах. Пусть диаграмма вида рис. 4.45, *а* коммутативна. Тогда:

а) если два малых квадрата декартовы, то внешний прямоугольник также декартов (нижние и верхние стороны внешнего прямоугольника являются композициями соответственно нижних и верхних сторон малых квадратов);

б) если внешний прямоугольник и правый квадрат декартовы, то декартов и левый квадрат.

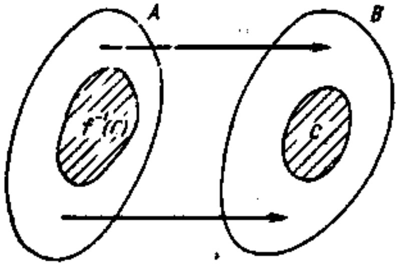


Рис. 4.39. Пояснение к примеру 4.15

$$\begin{tikzcd}
R_f \arrow[r, "p_2"] \arrow[d, "p_1"] & A \arrow[d, "f"] \\
A \arrow[r, "f"] & B
\end{tikzcd}$$

Рис. 4.40. Пояснение к понятию ядерного отношения

$$\begin{tikzcd}
K \arrow[r, hook] \arrow[d, "!"] & M \arrow[d, "f"] \\
O \arrow[r] & N
\end{tikzcd}$$

Рис. 4.41. Пояснение к понятию ядра

$$\begin{tikzcd}
s \arrow[r] \arrow[d] & q \arrow[d] \\
p \arrow[r] & r
\end{tikzcd}$$

Рис. 4.42. Пояснение к декартову квадрату в категории предпорядка $(P,\ <)$

$$\begin{tikzcd}
d \arrow[r, "g"] \arrow[d, "f"] & b \arrow[d, "!"] \\
a \arrow[r] & 1
\end{tikzcd}$$

Рис. 4.43. Пояснение к понятию декартова квадрата в произвольной категории с конечным объектом

$$\begin{tikzcd}
e \arrow[r, "i"] \arrow[d, "i"'] & a \arrow[d, "g"] \\
a \arrow[r, "f"'] & b
\end{tikzcd}$$

Рис. 4.44. Пояснение к связи декартова квадрата и уравнителя

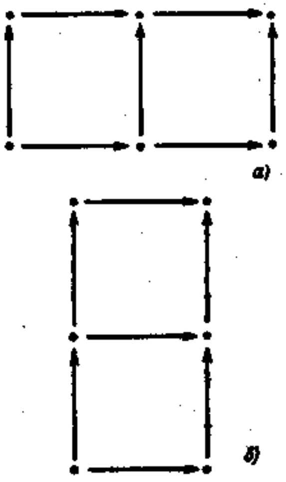


Рис. 4.45. Пояснение к лемме о квадратах

$$\begin{tikzcd}
a \arrow[r, "1_a"] \arrow[d, "1_a"'] & a \arrow[d, "f"] \\
a \arrow[r, "f"'] & b
\end{tikzcd}$$

Рис. 4.46. Пояснение к примеру 4.22

Лемма о квадратах имеет большое значение в теории категорий. Очень часто она используется для диаграмм вида приведенной на рис. 4.45, *б* в случае, когда на этой диаграмме внешний прямоугольник и *нижний* — декартовы. Тогда можно утверждать что *верхний* квадрат декартов.

Пример 4.22. Стрелка $f : a \to b$ в произвольной категории мономорфна тогда и только тогда, когда квадрат на рис. 4.46 декартов.

Амальгамы. Понятие амальгамы является двойственным к понятию обратного образа. Амальгамой пары стрелок $b \xleftarrow{f} a \xrightarrow{g} c$ с общим началом называется копредел диаграммы на рис. 4.47.

В категории Set амальгама получается построением дизъюнктивного объединения $b+c$ и отождествлением $f(x)$ с $g(x)$ для каждого $x \in a$ (т. е. амальгама совпадает с коуравнителем пары композиций:

$$a \xrightarrow{f} b \text{ включается в } b+c,\ a \xrightarrow{g} c \text{ включается в } b+c\Big).$$

Полнота. Категория K называется *полной*, если в ней каждая диаграмма имеет предел. Двойственным образом категория K называется *кополной*, если каждая K-диаграмма имеет

$$\begin{tikzcd}
a \arrow[r, "g"] \arrow[d, "f"'] & c \\
b &
\end{tikzcd}$$

Рис. 4.47. К определению амальгамы

$$\begin{tikzcd}
d \arrow[r] \arrow[d, "p"'] & a \arrow[d, "\langle 1_a, g\rangle"] \\
a \arrow[r, "\langle 1_a, f\rangle"'] & a \times b
\end{tikzcd}$$

Рис. 4.48. Обратные образы и уравнители

копредел. *Биполной* называется категория, которая является одновременно полной и кополной.

Конечной диаграммой называется диаграмма, имеющая конечное число объектов и конечное число стрелок между ними. Категория называется *конечно полной*, если она содержит предел любой конечной диаграммы. *Конечная кополнота и конечная биполнота* определяются аналогичным образом. Имеется важная теорема, которая приводится без доказательства.

Теорема 4.1. Если категория K имеет конечный объект и в ней для любой пары стрелок с общим концом существует обратный образ, то она конечно-полная.

Из этой теоремы следует:

а) при наличии конечного объекта и обратных образов произведение объектов a и b получается как обратный образ стрелок

$$a \to 1 \leftarrow b;$$

б) при наличии обратных образов и произведений для данной пары стрелок $f,\ g : a \rightrightarrows b$ сначала образуем стрелки $a \xrightarrow{\langle 1_a, f\rangle} a \times b$, $a \xrightarrow{\langle 1_a, g\rangle} a \times b$, а затем строим их обратный образ (рис. 4.48). Очевидно, что $p = g$ и p (или g) будет уравнителем f и g.

Экспоненцирование. Для множества A и B можно образовать в Set совокупность B^A всех функций, определенных на A и принимающих значение в B, т. е. $B^A = \{f : f$ — функция из A в $B\}$. Для характеристики B^A в терминах стрелок заметим, что с множеством B^A ассоциируется специальная стрелка

$$ev : B^A \times A \to B,$$

которая задается правилом $ev(<f, x>) = f(x)$. Стрелка ev называется *функцией значения*. Входами такой функции являются пары $<f, x>$, где $f : A \to B$ и $x \in A$. Действие функции ev заключается в применении f к x, т. е. в вычислении f в точке x. На выходе получится $f(x)$. Категориальное описание B^A осноа)

$$\begin{tikzcd}
B^A \times A \arrow[dr, "\hat{e}v(\ )"] & \\
 & B \\
C \times A \arrow[uu, dashed, "\hat{g}\times id_A"] \arrow[ur, "g"'] &
\end{tikzcd}$$

б

$$\begin{tikzcd}
b^a \times a \arrow[dr, "\hat{e}v(\ )"] & \\
 & b \\
c \times a \arrow[uu, dashed, "\hat{g}\times 1a"] \arrow[ur, "g"'] &
\end{tikzcd}$$

Рис. 4.49. К определению экспоненцирования

вано на том факте, что *ev* обладает свойством универсальности среди всех функций вида

$$C \times A \xrightarrow{g} B.$$

Для любой такой функции g существует одна и только одна функция $\hat{g}: C \to B^A$, для которой диаграмма на рис. 4.49, *а* коммутативна. На этой диаграмме $\hat{g} \times \mathrm{id}_A$ — (функторное) произведение функций, рассмотренное ранее. На паре $<c, a> \in C \times A$ значение функции $\hat{g} \times \mathrm{id}_A$ равно $<\hat{g}(c), \mathrm{id}_A(a)> = <\hat{g}(c), a>$.

Такое определение функции $\hat{g}$ вызвано тем обстоятельством, что при каждом фиксированном c функция g определяет отображение $A \to B$, которое получается, если первый элемент упорядоченных пар, являющихся аргументами функции g, положить равным c, в то время как второй будет пробегать множество A. Иначе: определим для данного $c \in C$ $g_c: A \to B$ следующим правилом $g_c(a) = g(<c, a>)$ для каждого $a \in A$. Функция $\hat{g}$ определяется теперь равенством $\hat{g}(c) = gc$ для $c \in C$. Для произвольной пары $<c, a> \in C \times A$ оказываются справедливыми равенства

$$ev(\langle \hat{g}(c), a \rangle) = g_c(a) = g(\langle c, a \rangle).$$

Таким образом, диаграмма на рис. 4.49, *б* коммутативна. Кроме того, требование коммутативности этой диаграммы, т. е. справедливости равенства $ex(<\hat{g}(c), a>) = g(<c, a>)$, означает, что $\hat{g}(c)$ должна быть функцией, которая для данного a дает $g(<c, a>)$, т. е. $\hat{g}(c)$ должна совпадать с определенной выше функцией g_c.

Определение. Считается, что категория K допускает *экспоненционирование*, если в ней существует произведение любых двух объектов и если для любых двух объектов a и b существуют K-объект b^a, называемый экспоненциалом, и K-стрелка $ev: b^a \times a \to b$, называемая *стрелкой значения*, такие, что для любых K-объекта C и K-стрелки $g: c \times a \to b$ существует единственная K-стрелка $\hat{g}: c \to b^a$, для которой диаграмма на рис. 4.49, *б* коммутативна, т. е. $ev \circ (\hat{g} \times 1a) = g$. Функция, ставящая в соответствие стрелке g стрелку $\hat{g}$ устанавливает биекцию

$$k(c \times a,\ b) \cong k(c,\ b^a)$$

между совокупностью K-стрелок из $c \times a$ в b и совокупностью K-стрелок из c в b^a. Докажем ее инъективность.

Если $\hat{g} = \hat{h}$, то $ev \circ (\hat{g} \times 1a) = ev \circ (\hat{h} \times 1a)$, т. е. $g = h$. Для установления ее сюръективности возьмем $h: c \to b^a$ и положим $g = ev \circ (h \times 1a)$. В силу единственности мы должны иметь $\hat{h} = g$. Две стрелки (g и $\hat{g}$), соответствующие друг другу на основании установленной биекции, называются *экспоненциально присоединенными* друг к другу.

Конечно-полная категория, допускающая *экспоненционирование, называется декартово-замкнутой категорией* (ДЗК).

Пример 4.23. Если A и B — конечные множества с m и n элементами соответственно, то множество B^A также конечно и имеет n^m элементов. В выражении n^m число m называется *экспонентой*. Это объясняет выбранную выше терминологию.

Категория Finord декартово-замкнута, и объект n^m совпадает с числом n^m.

Пример 4.24. Частично упорядоченное множество $P = (P, \leqslant)$, являющееся линейно упорядоченным (т. е. $p < q$ или $q < p$ для любых $p, q \in P$), называется *цепью*.

Пусть P — цепь с конечным объектом.

Положим

$$q^p = \begin{cases} 1, & \text{если } p \leqslant q; \\ q, & \text{если } q < p \ (\text{т. е. } q \leqslant p \text{ и } q \neq p). \end{cases}$$

Всякая цепь обладает произведением (наибольшая нижняя грань):

$$p \times q = \begin{cases} p, & \text{если } p \leqslant q; \\ q, & \text{если } q \leqslant p. \end{cases}$$

Таким образом, рассматривая *ev*, получаем два случая:

а) $p < q$. Тогда $q^p \times p = 1 \times p = p < q$;

б) $q < p$. Тогда $q^p \times p = q \times p = q$.

В обоих случаях $q^p \times p < q$, так что *ev* будет единственной имеющейся в P стрелкой $q^p \times p \to q$. Можно показать, что это определение наделяет P экспоненцированием.

* **Теорема 4.2.** Пусть K — декартово-замкнутая категория с начальным объектом O. Тогда:

1) $0 \cong 0 \times a$ для каждого объекта a;

2) если существует стрелка $a \to 0$, то $a \cong 0$;

3) если $0 \cong 1$, то категория K вырожденная, т. е. все K-объекты изоморфны между собой;

$$\begin{tikzcd}
 & a \arrow[dl, "f"'] \arrow[d, dashed, "{<f, 1_a>}"] \arrow[dr, "1_a"] & \\
0 & 0\times a \arrow[l, "pr_0"] \arrow[r, "pr_a"'] & a
\end{tikzcd}$$

а)

$$\begin{tikzcd}
 & C \arrow[dl, "f_1"'] \arrow[d, dashed, "{<f_1, f_2>}"] \arrow[dr, "f_2"] & \\
A_1 \arrow[d, "g_1"'] & A_1\times A_2 \arrow[l, "p_1"] \arrow[r, "p_2"'] \arrow[d, dashed, "g_1\times g_2"] & A_2 \arrow[d, "g_2"] \\
B_1 & B_1\times B_2 \arrow[l, "p_1"] \arrow[r, "p_2"'] & B_2
\end{tikzcd}$$

б)

$$\begin{tikzcd}
C\times A \arrow[r, "f"] \arrow[d, dashed, "\Lambda f\times id_A"'] & B \\
B^A\times A \arrow[ur, "ev"'] &
\end{tikzcd}$$

в)

Рис. 4.50. Пояснение к свойствам декартово-замкнутой категории (а), к определению произведения (б) и степени (в) ДЗК

4) любая стрелка $0\to a$ с началом O мономорфна;

5) $a^1\cong a$, $a^0\cong 1$, $1^a\cong 1$.

Доказательство 1. Для любого K-объекта b множество $K(O, b^a)$ имеет только один элемент (так как O — начальный объект). По определению экспоненцирования $K(O, b^a)\cong K(O\times a, b)$. Поэтому последнее множество содержит только один элемент. Таким образом, для любого b существует только одна стрелка $O\times a\to b$. Следовательно, $O\times a$ является начальным объектом. Так как последний единствен с точностью до изоморфизма, то $O\times a\cong 0$.

2. Пусть существует стрелка f: $a\to b$. Покажем, что $a\cong O\times a$. Отсюда в силу (1) будет $a\cong 0$. По определению произведения $pr_a\circ <f, 1_a> = 1_a$ (рис. 4.50). Кроме того, композиция $<f, 1_a>\circ pr_a$ является стрелкой из $O\times a$ в $O\times a$. Но такая стрелка единственная, так как $O\times a$ — начальный объект.

Таким образом, $<f, 1_a>\circ pr_a = 1_{O\times a}$, что дает $<f, 1_a> = pr_a^{-1}$ и $pr_a : O\times a\cong a$.

3. Пусть $O\cong 1$. Так как для произвольного a существует стрелка из a в 1, то в силу изоморфизма $O\cong 1$ имеется стрелка из a в O. В соответствии с (2) $O\cong a$. Таким образом, всякий объект изоморфен O, т. е. все объекты изоморфны друг другу.

4. Пусть f — стрелка из O в a. Предположим, что $f\circ g=f\circ h$, т. е. диаграмма

$$\begin{tikzcd}
b \arrow[r, shift left, "g"] \arrow[r, shift right, "h"'] & O \arrow[r, "f"] & a
\end{tikzcd}$$

коммутативна. В силу (2) $b\cong O$. Поэтому b будет начальным объектом. Следовательно, существует только одна стрелка $b\to O$. Таким образом, $g=h$, т. е. на f можно сокращать слева.

В заключение подведем итоги. Мы показали, что в категории Set мономорфные эпистрелки являются изострелками — свойство, не имеющее места в категории Mon. Однако в категории Grp это свойство справедливо, но Grp не является декартово-замкнутой категорией (это следует из только что доказанной теоремы: Grp — невырожденная категория, но $0\cong 1$). Однако не все декартово-замкнутые категории «подобны» Set. Частично упорядоченным множеством $n=\{0, \ldots, n-1\}$ будет декартово-замкнутая категория (будучи цепью с конечным объектом), но мономорфные эпистрелки в ней не обязательно являются изострелками.

Ввиду важности понятия ДЗК приведем резюме сложному его определению.

Категория K называется декартово-замкнутой категорией (ДЗК), если выполняются следующие три группы условий.

1. В K имеется терминальный объект T, такой, что для любого объекта $A\in K$ существует единственный морфизм $l_A : A\to T$.

2. Для любых объектов A_1, $A_2\in K$ имеется объект $A_1\times A_2$ (их декартово произведение) с морфизмами $p_i : A_1\times A_2\to A_i$ (проекциями), такими, что для любых $f_i : C\to A_i (i=1, 2)$ имеется единственный морфизм $<f_1, f_2> : C\to A_1\times A_2$, такой, что $p_0\circ <f_1, f_2> = f_i$ (рис. 4.50, a). Здесь требуется дать пояснение обозначений:

а) тождественный морфизм объекта $A\in K$ обозначается через id_A;

б) если $g_i : A_i\to B_i$ $(i=1, 2)$, то $g_1\times g_2 = <g_1\circ p_1, g_2\circ p_2> : : A_1\times A_2\to B_1\times B_2$ (рис. 4.50, б).

3. Для A, $B\in K$ имеется объект $B^A\in C$ (степень) с морфизмом $ev=ev_{A,B} : B^A\times A\to B$, такой, что для любого $f : C\times A\to B$ имеется единственный $\Lambda f : C\to B^A$, удовлетворяющий равенству $f=ev\circ(\Lambda f\times id_A)$ (рис. 4.50, в).

4.5. ФУНКТОРЫ

Функтор — это отображение из одной категории в другую, сохраняющее категорную структуру. *Ковариантным функтором* из категории K в категорию L называется отображение $F : K\to L$, удовлетворяющее следующим условиям.

1. Для каждого $A\in O$ в K $F(A)\in O$ в L.

2. Для каждого $\alpha\in Mor_K(A, B)$;

$$F(\alpha)\in Mor_L(F(A),\ F(B)).$$

3. Для всякой единицы $1_A\in Mor_K(F, A)$;

$$F(1_K) = 1_{F(A)}\in Mor_L(F(A),\ F(A)).$$

4. Если $\alpha \in \mathrm{Mor}_K(A, B)$, $\beta \in \mathrm{Mor}_K(B, C)$, то $F(\alpha\beta) = F(\alpha)F(\beta)$.

Ковариантный функтор из категории K^0, двойственной к категории K, в категорию L называется контравариантным из K в L.

Пример 4.25. Построим контравариантный функтор из KG в Set (см. пример 4.1). Каждому элементу из $N \cup \Sigma$ сопоставим множество из Set. Цепочкам из $(N \cup \Sigma)^*$ сопоставим декартово произведение соответствующих множеств, взятых в том же порядке. На выводы (морфизмы) этот функтор действует следующим образом. Каждой продукции из P сопоставляется функция из Set, такая, что областью ее определения является множество, соответствующее правой части продукции. Композиции правил вывода соответствует композиция функций. Этот функтор [49] называется интерпретацией и обозначается I. *Образ вывода* $\alpha \in \mathrm{Mor}_{KG}(S, \omega)$, $\omega \in (N \cup \Sigma)^*$ при функторе I называется семантикой предположения. Под языком будем понимать тройку (G, I, Sem), где G — грамматика, порождающая категорию синтаксиса; I — интерпретация, Sem — категория семантики. Тогда каждой программе компьютера ставится в соответствие ее семантика (или множество в случае неоднозначных грамматик) при помощи функтора I. Множество выводов цепочки в грамматике G есть $\mathrm{Mor}_{KG}(S, \omega)$. Обозначим через $I(\mathrm{Mor}_{KG}(S, \omega))$ множество образов этих выводов при отображении I. Это множество и будет множеством программ-реализаций запроса ω.

На «стрелочном» (топологическом, геометрическом) языке функторы вводятся следующим образом.

Функтором F из категории K в категорию D называется функция, ставящая в соответствие:

1) каждому K-объекту a некоторый D-объект $F(a)$;

2) каждой K-стрелке $f: a \to b$ D-стрелку $F(b)$, $F(a) \to F(b)$, такую что:

а) $F(1_a) = 1_{F(a)}$ для каждого K-объекта a, т. е. единичной стрелке, соответствующей объекту a, сопоставляется единичная стрелка, соответствующая объекту $F(a)$;

б) $F(g \circ f) = F(g) \circ F(f)$ для любых g и f, для которых определена композиция $g \circ f$.

Условие «б» означает, что F-образ композиции двух стрелок является композицией их F-образов, т. е. диаграмма на рис. 4.51, *а* коммутативна в K, а диаграмма на рис. 4.51, *б* коммутативна в D. Для обозначения функтора F из K в D используется обозначение $F: K \to D$ или $K \to D$. Таким образом, функтор — это такое отображение (от слова «образ»), которое сохраняет операции dom и cod, а также единичные стрелки и композиции.

Пример 4.26. *Тождественный функтор* $1_K: K \to K$ определяется правилами $1_K(a) = a$, $1_K(f) = f$. Эти же правила задают функцию *включения* K в D в случае, когда K — подкатегория в D.

Пример 4.27. *Пренебрегающий функтор.* Пусть K — какая-нибудь категория из списка табл. 4.1, к примеру K-Top. Тогда K-объектами служат множества, несущие некоторую дополнительную структуру. Пренебрегающий функтор $U: K \to \mathrm{Set}$ ставит в соответствие каждому K-объекту его несущее множество, а каждой K-стрелке — саму эту стрелку, т. е. U пренебрегает структурой K-объектов, как бы «забывая» ее и «помнит» только, что K-стрелки являются теоретико-множественными структурами.

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[dr, "h"'] & b \arrow[d, dashed, "g"] \\
 & c
\end{tikzcd}
\qquad
\begin{tikzcd}
F(a) & F(b) \arrow[l, "F(f)"'] \arrow[d, dashed, "F(g)"] \\
 & F(c) \arrow[ul, "F(h)"]
\end{tikzcd}$$

а) б)

Рис. 4.51. Пояснение к понятию функтора

Пример 4.28. *Функтор степени* $P: \mathrm{Set} \to \mathrm{Set}$ ставит в соответствие каждому множеству A его множество-степень $P(A)$, а каждой функции $f: A \to B$ — функцию $P(f): P(A) \to P(B)$, переводящую любое множество подмножества $X \subseteq A$ в его f-образ $f(X) \subseteq B$.

Пример 4.29. Пусть P и Q — категории порядка. Тогда функтор $F: P \to Q$ — это просто функция $F: P \to Q$, являющаяся монотонной, т. е. такой, что если $r < q$ в R, то $F(r) < F(q)$ в Q. Частный случай этого примера получается, если рассматривать множества-степени как ЧУМ $(P(A), \subseteq)$. Для данной функции $f: A \to B$ и произвольных множеств X и Y в A из того, что $X \subseteq Y$, следует $f(X) \subseteq f(Y)$. Таким образом, функция $P(f): P(A) \to P(B)$ сама является функтором между соответствующими категориями порядка.

Пример 4.30. *Гомоморфизмы моноидов.* Функтор между моноидами $(M, *, e)$ и $(N, \square, e')$, рассматриваемыми как однообъектные категории, является, по существу, *гомоморфизмом* моноидов, т. е. функцией $F: M \to N$, такой, что

$$F(e) = e';$$
$$F(x * y) = F(x) \square F(y).$$

Пример 4.31. Если категория $\mathscr{G}$ имеет произведения, то каждый объект a определяет функтор $- \times_a: \mathscr{G} \to \mathscr{G}$, сопоставляющий каждому объекту b объект $b \times a$, а каждой стрелке $f: b \to c$ — стрелку $f \times 1_a: b \times a \to c \times a$.

Пример 4.32. *hom*-функторы. Каждый $\mathscr{G}$-объект a определяет функтор $\mathscr{G}(a, -): \mathscr{G} \to \mathrm{Set}$, который ставит в соответствие объекту b множество $\mathscr{G}(a, b)$ всех $\mathscr{G}$-стрелок из a в b, а каждой $\mathscr{G}$-стрелке $f: b \to c$ — функцию $\mathscr{G}(a, b): \mathscr{G}(a, b) \to \mathscr{G}(a, c)$, переводящую g в $f \circ g$ (рис. 4.52). Название hom-функтор происходит от слова homomorphism, которое используется иногда вместо термина «стрелка» (arraw). Совокупность $\mathscr{G}(a, b) = \hom \mathscr{G}(a, b)$ называется hom-*множеством*. hom-функтор будет определен только тогда, когда hom-множества категории $\mathscr{G}$ малые, т. е. настоящие множества, а не собственные классы.

$$\begin{tikzcd}
a \arrow[r, "g"] \arrow[dr, "f \circ g"'] & {} \arrow[d, "f"] \\
& c
\end{tikzcd}$$

Рис. 4.52. Пояснение к понятию hom-функтора

Контравариантные функторы. Ранее рассмотренные примеры относились к ковариантным функторам. Они сохраняли направления стрелок, т. е. начало стрелки переходило в начало, а конец — в конец. Контравариантный функтор меняет направление стрелок, отображая начало в конец и наоборот. Если ковариантный функтор отображает категорию K в категорию L, контравариантный функтор отображает категорию K^0 в категорию L, где K^0 — двойственная к K категория. Напомним, что категория K^0 содержит объекты и морфизмы K $(A, B \ldots, \alpha, \beta \ldots)$, которые обозначает через A^0, B^0 $\ldots$; α^0, $\beta^0 \ldots$, как входящие в K. Для любой пары объектов A^0, $B^0 \in ObK^0$, по определению $Mor_{K^0}(A^0, B^0) = Mor_K(B, A)$. Для любых морфизмов $\alpha^0 : A^0 \to B^0$ и $\beta^0 : B^0 \to C^0$, по определению, $\alpha^0\beta^0 = (\beta\alpha)^0$. Таким образом, *контравариантный функтор* $F : K \to D$ ставит в соответствие стрелке $f : a \to b$ стрелку $F(f) : F(b) \to F(a)$ так, что $F(1_a) = 1_{F(a)}$, как и раньше, но теперь $F(g \circ f) = F(f) \circ F(g)$, т. е. коммутативная диаграмма на рис. 4.53, *а* переходит в другую коммутативную диаграмму на рис. 4.53, *б*.

Пример 4.33. Контравариантный функтор между категориями порядка представляет собой *антитонную* функцию $F : P \to Q$, т. е. такую, что из $p < q$ в P следует $F(q) < F(p)$ в Q.

Пример 4.34. *Контравариантный функтор степени* $p : \mathrm{Set} \to \mathrm{Set}$ ставит в соответствие каждому множеству A его множество-степень $P(A)$, а каждой функции $f : A \to B$ — функцию $\widehat{p}(f) : P(B) \to P(A)$, переводящую подмножество $X \subseteq B$ в его прообраз $f^{-1}(X) \subseteq A$.

Пример 4.35. *Контравариантный hom-функтор* $\mathcal{G}(-, a) : \mathcal{G} \to \mathrm{Set}$, ассоциированный с фиксированным объектом a, ставит в соответствие объекту b множество $\mathcal{G}(b, a)$, а $\mathcal{G}$-стрелке $f : b \to c$ — функцию $\mathcal{G}(f, a) : \mathcal{G}(c, a) \to \mathcal{G}(b, a)$, переводящую g в $g \circ f$ (рис. 4.54).

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[dr, "h"'] & b \arrow[d, "g"] \\
& c
\end{tikzcd}
\qquad
\begin{tikzcd}
F(a) \arrow[r, "F(f)"] \arrow[dr, "F(h)"'] & F(b) \arrow[d, "F(g)"] \\
& F(c)
\end{tikzcd}$$

а) *б)*

Рис. 4.53. Пояснение к понятию ковариантного и контравариантного функтора

$$\begin{tikzcd}
b \arrow[r, dash, "f"] \arrow[dr, dash, "g \circ f"'] & c \arrow[d, dash, "g"] \\
& a
\end{tikzcd}$$

Рис. 4.54. Пояснение к понятию контравариантного hom-функтора

$$\begin{tikzcd}
d \arrow[r, tail] \arrow[d] & a \arrow[d, "f"] \\
c \arrow[r, tail, "g"] & b
\end{tikzcd}$$

Рис. 4.55. Пояснение к понятию функтора Sub

Пример 4.36. Функтор $\mathrm{Sub} : \mathcal{G} \to \mathrm{Set}$ ставит в соответствие каждому $\mathcal{G}$-объекту a совокупность $\mathrm{Sub}(a)$ всех его подобъектов, а каждой $\mathcal{G}$-стрелке $f : a \to b$ — функцию $\mathrm{Sub}(f) : \mathrm{Sub}(b) \to \mathrm{Sub}(a)$, которая переводит стрелку $g : c \rightarrowtail b$ в стрелку $h : d \rightarrowtail a$, получающуюся из g подъемом вдоль f (рис. 4.55). Конечно, эта конструкция возможна, только если $\mathcal{G}$ имеет обратные образы. Функтор Sub является обобщением функтора из примера 4.10. Заметим, что в абстрактной категориальной машине (см. гл. 1) одной из основных операций в процессе получения топосов из категории является получение из объектов категории соответствующих подобъектов. В соединения с соответствующими ДЗК они дают топосы (рис. 4.56). В общем случае процесс получения подобъектов из объектов осуществляется с помощью функтора.

Пример 4.37. Под языком будем понимать тройки (G, I, Sem), где G — грамматика, порождающая категорию синтаксиса; I — интерпретация, а Sem — категория семантики. Тогда каждой программе компьютера ставится в соответствие ее семантика (или множество семантик в случае неоднозначных грамматик) при помощи функтора I. Множество выводов цепочки ω в грамматике G есть $Mor_{KG}(S, \omega)$. Обозначим через I $(Mor_{KG}(s, \omega))$ множество всех образов этих выводов при отображении I. Это множество и будет множеством программ-реализаций как образ функтора, определенного при помощи синтаксиса и семантики.

Слово «функтор» будет в дальнейшем всегда обозначать ковариантный функтор. Контравариантный функтор $F : \mathcal{G} \to D$ может быть заменен ковариантным функтором $\overline{F} : \mathcal{G}^{op} \to D$, $\overline{F}(a) = F(a)$ и для стрелки $f^{op} : b \to a$ из $\mathcal{G}^{op}$ (здесь $f : a \to b$ — произвольная стрелка из категории $\mathcal{G}$) $\overline{F}(f^{op}) = F(f) : F(b) \to F(a)$. Для данных функторов $F : \mathcal{G} \to D$ и $G : D \to F$ их композиция снова дает функтор $G \circ F : \mathcal{G} \to F$. Эта операция ассоциативна $H \circ (G \circ F) = (H \circ G) \circ F$. Таким образом, можно рассматривать функторы как стрелки между категориями. Существует категория Cat всех категорий, объектами которой являются категории, а стрелками — функторы. Единичными стрелками в ней являются тождественные функторы из примера 4.26.

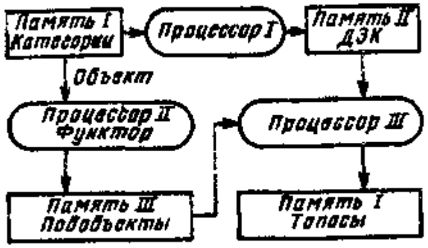


Рис. 4.56. Фрагмент категориальной абстрактной машины (КАМ)

$$\begin{tikzcd}
a \arrow[dd, "f"] & F(a) \arrow[r, dash, "\tau_a"] \arrow[dd, dash, "F(f)"'] & G(a) \arrow[dd, dash, "G(f)"] \\
 & & \\
b & F(b) \arrow[r, dash, "\tau_b"] & G(b)
\end{tikzcd}$$

Рис. 4.57. Пояснение к понятию естественных преобразований

В связи с рассмотрением категории Cat встают фундаментальные проблемы. Категория Set не должна быть элементом класса Cat-объектов (если считается, что Cat-объекты образуют класс), так как Set как совокупность является собственным классом и не является членом никакой совокупности. Вопрос о том, является ли Cat Cat-объектом, подводит нас вплотную к парадоксу Рассела. Вообще категория Cat понимается как категория малых категорий, т. е. таких, для которых совокупность всех стрелок образует множество.

Естественные преобразования. Введя функторы между категориями, первоначально определенными как совокупности объектов и стрелок, мы поднялись на новую ступень абстракции и рассматриваем теперь категории как объекты с функторами между ними в качестве стрелок.

Для данных категорий $\mathcal{G}$ и D построим категорию, обозначаемую Finset$(\mathcal{G}, D)$ или $D^{\mathcal{G}}$, объектами которой являются функторы из $\mathcal{G}$ в D. Дадим определение стрелки из одного функтора в другой. Рассматривая функторы $F:\mathcal{G}\to D$ и $G:\mathcal{G}\to D$, будем считать, что они дают различные «изображения» категории $\mathcal{G}$ в D. Удачное «интуитивное» представление о «преобразовании» F в G можно получить, если вообразить себе попытку наложения или сдвига F-изображения на G-изображение с помощью стрелок категории D. Такой сдвиг можно осуществить, поставив в соответствие каждому $\mathcal{G}$-объекту a D-стрелку, начинающуюся в F-образе объекта a и заканчивающуюся в G-образе объекта a. Обозначим эту стрелку через τ_a. Таким образом, $\tau_a: F_a\to G_a$. Чтобы этот процесс сдвига или преобразования сохранил структуру, потребуем, чтобы диаграмма на рис. 4.57, возникающая для каждой стрелки $f: a\to b$, была коммутативна. Таким образом, τ_a и τ_b представляют категориальный способ преобразования F-изображения стрелки $f: a\to b$ в ее G-изображение. Подводя итоги, можно сказать, что *естественное преобразование функтора* $F:\mathcal{G}\to D$ в функтор $G:\mathcal{G}\to D$ — это функция τ, ставящая в соответствие каждому $\mathcal{G}$-объекту a D-стрелку $\tau_a: F(a)\to G(a)$ так, что для любой $\mathcal{G}$-стрелки $f: a\to b$ приведенная выше диаграмма коммутативна в D, т. е. $\tau_b \circ F(f) = G(f)\circ\tau_a$. Будем использовать обозначения $\tau: F \dot{\to} G$ или $F \overset{\tau}{\dot{\to}} G$, чтобы указать на то, что τ — естественное преобразование F в G. Стрелки τ_a называются *компонентами* преобразования τ.

Если все компоненты τ_a являются изострелками, это можно интерпретировать следующим образом: «F- и G-изображения категории $\mathcal{G}$ в D одинаковы». Будем называть естественное преобразование τ в этом случае *естественным изоморфизмом.* Тогда каждая стрелка $\tau_a: F(a)\to G(a)$ имеет обратную стрелку $\tau_a^{-1}: G(a)\to F(a)$, и эти стрелки τ_a^{-1} можно рассматривать как компоненты естественного изоморфизма $\tau^{-1}: G \dot{\to} F$. Будем обозначать естественный изоморфизм через $\tau: F\cong G$. Подводя итог, формально естественный изоморфизм двух функторов $F, G: \mathcal{G}\to D$ можно определить следующим образом: функция $\tau: \mathrm{Ob}\mathcal{G}\to \mathrm{Mor}\, D$ называется естественным изоморфизмом функторов F и G, если она удовлетворяет следующим условиям.

1. Для каждого $a\in\mathrm{Ob}\mathcal{G}$

$$\tau(a) = \tau_a \in \mathrm{Mor}(F(a),\ G(a)).$$

2. Для любых a, $b\in\mathrm{Ob}\mathcal{G}$ и любого $f: a\to b$ справедлива коммутативная диаграмма на рис. 4.57.

3. Для каждого $a\in\mathrm{Ob}\mathcal{G}$ τ_a — изоморфизм в D.

Пример 4.38. Тождественное естественное преобразование $I_F: F\to F$ сопоставляет каждый объект a с единичной стрелкой $I_{F(a)}: F(a)\to F(a)$. Очевидно, что такая функция является естественным изоморфизмом.

Пример 4.39. Ранее был определен изоморфизм $A\cong A\times 1$ для каждого Set-объекта, который является естественным. Это значит, что если функтор $-\times 1$: Set→Set (см. пример 4.1), то для каждой Set-стрелки $f: A\to B$ диаграмма на рис. 4.58, *а* коммутативна, где $\tau_a(x) = <x, 0>$. Левая сторона квадрата представляет собой образ стрелки f при тождественном функторе. Таким образом, биекции τ_a являются компонентами естественного изоморфизма τ-функтора 1_{Set} в функтор $-\times 1$.

Пример 4.40. В категории Set имеем изоморфизм $A\times B\cong B\times A$, даваемый отображением $t_{wB}: A\times B\to B\times A$, где $t_{wB}(<x, y>) = <y, x>$. Для каждого Set-объекта A аналогично тому, как определялся функтор левого умножения $-\times A$: Set→Set, можно определить функтор левого умножения $A\times -$: Set→Set, ставящий в соответствие множеству B множество $A\times B$ и функции $f: B\to C$ функцию $1_A\times f: A\times B\to A\times C$. Тогда для произвольной Set-стрелки $f: B\to C$ диаграмма на рис. 4.58, *б* коммутативна, это доказывает, что биекции t_{wB} являются компонентами естественного изоморфизма функтора $A\times -$ в функтор $-\times A$.

Пример 4.41. Пусть KG есть категория грамматики G, а L — категория семантики; F и G — возможные интерпретации. Определим естественный изо-

$$\begin{tikzcd}
A \arrow[d] & A \arrow[r] \arrow[d, "f"'] & A \times 1 \arrow[d, "f \times id"] \\
B & B \arrow[r, "\tau_B"'] & B \times 1
\end{tikzcd}$$

а)

$$\begin{tikzcd}
B \arrow[d] & A \times B \arrow[r, "t_{W_B}"] \arrow[d, "1_A \times f"'] & B \times A \arrow[d, "f \times 1_A"] \\
C & A \times C \arrow[r, "t_{W_C}"] & C \times A
\end{tikzcd}$$

б)

Рис. 4.58. Пояснение к примерам 4.39, 4.40

$$\begin{tikzcd}
X \arrow[r, "\tau_x"] \arrow[d, "f"'] & F(G(X)) \arrow[d, "F(G(f))"] \\
Y \arrow[r, "\tau_y"] & F(G(Y))
\end{tikzcd}$$

Рис. 4.59. Пояснение к примеру на эквивалентность категорий

морфизм $\varphi : F \to G$ следующим образом: каждую цепочку $\alpha \in (N \cup \Sigma)^*$ сопоставим с парой $(F(\alpha), G(\alpha))$, т. е. функцией φ_α, отображающей множество, соответствующее α при интерпретации G. Пусть теперь даны две цепочки α и $\beta \in (N \cup \Sigma)^*$ и $\theta : \alpha \to \beta$ — вывод β из α. Тогда $F(\theta)$ есть функция, соответствующая этому выводу при интерпретации F, а $G(\theta)$ — при интерпретации G. Коммутативность диаграммы означает, что композиция функций φ_α и $G(\theta)$ совпадает с композицией функции $F(\theta)$ и φ_β. Поэтому можем записать, что $F(\theta) = \varphi_\alpha^{-1} G(\theta) \varphi_\beta$. Например, если $\theta : S \to \omega$, то семантика ω при интерпретации F может быть получена из семантики этой же цепочки при естественно изоморфной интерпретации G.

Эквивалентность категорий позволяет отвечать на вопросы одинаковости категорий. Один вариант сравнения может быть предложен с помощью категорий как объектов в Cat. Будем называть функтор $F \to \mathscr{G} \to D$ изофунктором, если он имеет обратный функтор $G : D \to \mathscr{G}$, что $G \circ F = I_{\mathscr{G}}$ и $F \circ G = I_D$. Будем называть категории $\mathscr{G}$ и D *изоморфными* и писать $\mathscr{G} \cong D$, если существует изофунктор $F : \mathscr{G} \to D$. Можно утверждать, что приведенное определение «одинаковости» строже, чем это необходимо. Если F имеет обратный функтор G, то для $\mathscr{G}$-объекта a справедливо равенство $a = G(F(a))$, а для D-объекта b — равенство $b = F(G(b))$. С точки зрения основного категориального принципа неразличимости изоморфных вещей можно считать $\mathscr{G}$ и D «в сущности одинаковыми», если $a \cong G(F(a))$ в $\mathscr{G}$ и $b \cong F(G(b))$ в D. Иначе, $\mathscr{G}$ и D должны быть категориально эквивалентными, если они неразлучны «с точностью до изоморфизма». Это означает, что изоморфизмы $a \to G(F(a))$ $b \to F(G(b))$ являются естественными. Таким образом, функтор $F : \mathscr{G} \to D$ называется *эквивалентностью категорий*, если существуют функтор $G : D \to \mathscr{G}$ и естественные изоморфизмы $\tau : 1_{\mathscr{G}} \cong G \circ F$ и $\sigma : 1_D \cong F \circ G$ тождественного функтора на D и функтор $F \circ G$. Категории называются *эквивалентными* $\mathscr{G} \cong D$, если существует эквивалентность $F : \mathscr{G} \to D$.

Пример 4.42. Finord$\cong$Finset. Пусть F — функция включения Finord в Finset — функтор включения. Для каждого конечного множества X определим $G(X) = n$, где n — число элементов в X. Для каждого X пусть τx обозначает биекцию множества X на $G(X)$, а если X — ординал, то τx — тождественная функция на X. Для произвольной функции $f : X \to Y$ положим $G(f) = \tau y \circ f \circ \tau x^{-1}$. Тогда G является функтором из Finset в Finord. Так как диаграмма на рис. 4.59 коммутативна (по определению G), то отображения τx являются компонентами естественного изоморфизма $\tau : 1 \to F \circ G$. Кроме того, функтор $G \circ F$ — тождественный функтор на Finord.

Понятие эквивалентности категорий можно пояснить с помощью скелетальных категорий, в которых изоморфные объекты равны, т. е. если $a \cong b$, то $a = b$. Категория Finord скелетарна, так как изоморфные конечные множества имеют одно и то же число элементов. *Скелетом* категории $\mathscr{G}$ называется полная подкатегория $\mathscr{G}_0$ категории $\mathscr{G}$, являющаяся скелетарной, и такая, что для каждого $\mathscr{G}$-объекта существует изоморфный ему $\mathscr{G}_0$-объект. Категория Finord является скелетом категории Finset. Можно считать, что скелет $\mathscr{G}_0$ категории $\mathscr{G}$ отображает всю существенную категориальную структуру категории $\mathscr{G}$. Категории $\mathscr{G}_0$ и $\mathscr{G}$ эквивалентны. Эквивалентность обеспечивается функтором включения $\mathscr{G}_0$ в $\mathscr{G}$, в чем можно убедиться с помощью метода, использованного в последнем примере.

Всякая категория имеет скелет. Отношение изоморфизма разделяет $\mathscr{G}$-объекты на классы эквивалентности. Выберем по одному объекту каждого класса эквивалентности и положим, что категория $\mathscr{G}_0$ равна полной подкатегории категории $\mathscr{G}$, классом объектов которой является совокупность всех выбранных представителей; $\mathscr{G}_0$ — скелет категории $\mathscr{G}$.

Эквивалентность категорий можно определить следующим образом: категории $\mathscr{G}$ и D эквивалентны тогда и только тогда, когда они имеют изоморфные скелеты ($\mathscr{G} \cong D$, $\mathscr{G}_0 \cong D_0$).

Таким образом, категориальные эквивалентные категории — в сущности, одинаковые категории. Заметим, что эквивалентные категории не обязаны находиться в биектном соответствии друг

$$\begin{tikzcd}
a \arrow[d, "f"'] & F(a) \arrow[r, "\tau_a"] \arrow[d, "F(f)"'] & G(a) \arrow[r, "\sigma_a"] \arrow[d, "G(f)"'] & H(a) \arrow[d, "H(f)"'] \\
b & F(b) \arrow[r, "\tau_b"] & G(b) \arrow[r, "\sigma_b"] & H(b)
\end{tikzcd}$$

Рис. 4.60. Пояснение к понятию категории функторов

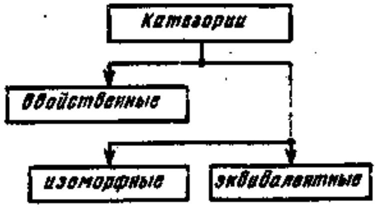


Рис. 4.61. Разновидности категорий

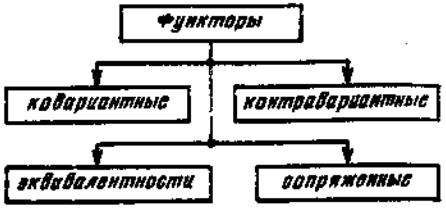


Рис. 4.62. Разновидности функторов

с другом и вообще быть сравнимыми по величине. Совокупность конечных ординалов — малая категория, т. е. множество, отождествляемое с множеством натуральных чисел, в то время как объекты категории Finset образуют собственный класс (так как он, например, содержит $\{x\}$ для каждого множества X).

Категории функторов. Продолжим рассмотрение категории $D^{\mathscr{G}}$ всех функторов из $\mathscr{G}$ в D, которое было начато ранее. Пусть F, G, M — функторы из $\mathscr{G}$ в D и $\tau : F \to G$, $\sigma : G \dot{\to} H$ — естественные преобразования функторов. Тогда для произвольной $\mathscr{G}$-стрелки $f : a \to b$ можно нарисовать диаграмму (рис. 4.60). Наша цель — определить естественное преобразование, являющееся композицией $\sigma \circ \tau$ преобразований τ и σ. Диаграмма на рис. 4.60 дает способ для этого. Для каждого a положим $(\sigma \circ \tau)_a = \sigma_a \circ \tau_a$. Так как оба квадрата диаграммы коммутативны, то внешний прямоугольник коммутативен, т. е. $(\sigma \circ \tau)_b \circ F(b) = H(f) \circ (\sigma \circ \tau)_a$. Поэтому стрелки $(\sigma \circ \tau)_a$ служат компонентами естественного преобразования $\sigma \circ \tau : F \dot{\to} H$. Так определяется операция композиции в категории $D^{\mathscr{G}}$ функторов. Для каждого функтора $F : \mathscr{G} \to D$ тождественное преобразование $1_F : F \to F$ является единичной стрелкой на $\mathscr{G}$-объекты F. Основные разновидности категорий приведены на рис. 4.61, функторов — на рис. 4.62.

4.6. ТОПОСЫ

Для введения понятия топосов рассмотрим класс конечных множеств как категорию с произвольными отображениями множеств как категорию с произвольными отображениями в качестве морфизмов. В этом классе существует одноэлементное множество, декартово произведение двух множеств, прообраз множества при любом отображении и множество всех подмножеств данного множества. Эти свойства класса выразимы в определенном смысле в категории конечных множеств на категориальном языке, т. е. с использованием переменных как по объектам, так и по морфизмам, символа операции композиции морфизмов и символа отношения равенства между морфизмами. Эти четыре категориальных

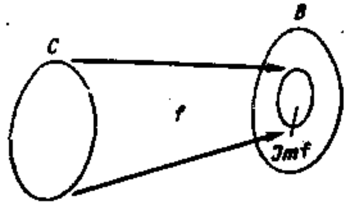


Рис. 4.63. К определению топоса

$$\begin{tikzcd}
B & \\
 & D \arrow[ul] \\
A \arrow[uu, hook] \arrow[ur, hook] &
\end{tikzcd}$$

Рис. 4.64. Пояснение к понятию подобъекта

предложения образуют аксиоматику топосов. Таким образом, получается, что категория всех конечных множеств является топосом.

Элементарным топосом называется категория K, такая что:

1) K конечно полна;
2) K конечно кополна;
3) K допускает экспоненцирование;
4) K имеет классификатор подобъектов.

Ниже показано, что условия 2 и 3 соответствуют определению *декартово-замкнутой категории* (ДЗК). Кроме того, условие 1 может быть заменено условием: K обладает конечным объектом и обратными образами. Условие 2 может быть двойственным образом заменено на условие: K имеет начальный элемент и амальгамы.

Далее можно доказать, что условие 2 следует из остальных. *Топос* может быть определен как ДЗК с классификатором подобъектов.

Ниже дано еще одно определение топоса, основанное на категориальной характеризации множества-степени.

Слово «элементарный» относится к деталям теории топосов и в дальнейшем опущено. Как было отмечено, топос формируется из ДЗК и подобъектов. Поэтому прежде всего займемся подобъектами.

Подобъекты. Если множество A является подмножеством множества B, то функция включения A и B инъективна и поэтому мономорфна. С другой стороны, произвольная мономорфная функция $f : C \rightarrowtail B$ определяет подмножество множества B, а именно $\mathrm{Im} f = \{f(x) : x \in C\}$. Легко убедиться, что f индуцирует биекцию между C и $\mathrm{Im} f$. Следовательно, $C \cong \mathrm{Im} f$ (рис. 4.63).

Таким образом, область определения мономорфной функции изоморфна некоторому подмножеству области значений этой функции. Другими словами, область определения является с точностью до изоморфизма подмножеством области значений. Итак, приходим к категориальному аналогу понятия подмножества, т. е. к понятию подобъекта: *подобъектом* $\mathscr{G}$-объекта d или *подобъектом* в d называется мономорфная $\mathscr{G}$-стрелка $f: a \rightarrowtail d$ с концом d.

Далее, если D — произвольное множество, то множество всех его подмножеств, обозначаемое через $\mathscr{P}(D)$, называется *множеством-степенью* множества D. Таким образом, $\mathscr{P}(D) = \{A : A$ — подмножество $D\}$. Отношение теоретико-множественного включения является частичным порядком на множестве-степени $\mathscr{P}(D)$, т. е. пара $(\mathscr{P}(D), \subseteq)$ образует ЧУМ, которое можно рассматривать как категорию, в которой стрелка $A \to B$ имеется тогда и только тогда, когда $A \subseteq B$. При наличии стрелки $A \to B$ диаграмма на рис. 4.64 коммутативна. Отсюда получается определение *отношения включения* между подобъектами объекта d. Для заданных подобъектов $f: a \rightarrowtail d$ и $g: b \rightarrowtail d$ положим $f \subseteq g$ тогда и только тогда, когда существует $\mathscr{G}$-стрелка $h: a \to b$, такая, что диаграмма на рис. 4.65 коммутативна, т. е. $f = g \circ h$. Такая стрелка h будет мономорфна, т. е. будет подобъектом объекта b. Таким образом, $f \subseteq g$ тогда и только тогда, когда f пропускается через g.

Отношение включения над подобъектах:

1) *рефлексивно* $f \subseteq f$, так как $f = f \circ 1a$ (рис. 4.66);

2) *транзитивно*: если $f \subseteq g$ и $g \subseteq k$, то $f \subseteq k$, так как из $f = g \circ h$ и $g = k \circ i$ следует $f = k \circ (i \circ h)$ (рис. 4.67).

Если $f \subseteq g$ и $g \subseteq f$, то f и g пропускаются друг через друга, в соответствии с диаграммой на рис. 4.67 $f = g \circ h$, $g = f \circ i$. В этом случае $h: a \to b$ будет изострелкой, а i — обратной к ней.

Таким образом, если $f \subseteq g$ и $g \subseteq f$, то стрелки f и g имеют изоморфные начала. Поэтому будем их называть *изоморфными подобъектами* и писать $f \approx g$. Для того чтобы отношение $\subseteq$ было антисимметричным, требуется справедливость равенства $f = g$ всякий раз, когда $f \approx g$. В действительности, однако, это может быть и не так, поскольку a может быть не равно b. Таким образом, знак $\subseteq$ есть, вообще говоря, отношение предпорядка на подобъектах (при данном выше определении подобъекта), а не отношение частичного порядка. Если оставим это положение без изменения, то в дальнейшем встретимся с определенными трудностями. Поэтому желательно иметь возможность считать отношение $\subseteq$ антисимметричным. Механизм, который позволяет это сделать, рассмотрен ранее. Отношение $\simeq$ является отношением эквивалентности. Каждая стрелка $f: a \rightarrowtail d$ определяет класс эквивалентности

$$[f] = \{g : f \simeq g\}.$$

$$\begin{tikzcd}
b \arrow[rd, tail, "g"] & \\
 & d \\
a \arrow[uu, tail, "h"] \arrow[ru, tail, "f"'] &
\end{tikzcd}$$

Рис. 4.65. Пояснение к отношению включения

$$\begin{tikzcd}
a \arrow[rd, tail, "f"] & \\
 & d \\
a \arrow[uu, tail, "1_a"] \arrow[ru, tail, "f"'] &
\end{tikzcd}$$

Рис. 4.66. Рефлективность отношения включения

Рис. 4.67. Пояснение к свойству транзитивности отношения включения

$$\begin{tikzcd}
c \arrow[rd, tail, "k"] & \\
b \arrow[u, "i"] \arrow[r, tail, "g"] & d \\
a \arrow[u, "h"] \arrow[ru, tail, "f"'] \arrow[uu, bend left=60, "i \circ h"] &
\end{tikzcd}$$

Образуем множество

$\mathrm{Sub}(d) = \{[f] : f$ есть монострелка и $\mathrm{cod} f = d\}$.

Будем считать элементы множества $\mathrm{Sub}(d)$ подобъектами, т. е. переопределим понятия подобъекта; подобъектами объекта будем называть классы эквивалентности монострелок с концом в d. Чтобы определить отношение включения на введенных объектах, положим (используя тот же самый символ, что и раньше) $[f] \subseteq [g]$ тогда и только тогда, когда $f \subseteq g$. Здесь можно задать вопрос, является ли определение, даваемое с помощью представителей из классов эквивалентности, независимым от выбора представителей. Ответ: да, является (рис. 4.68).

Если $[f] = [f']$ и $[g] = [g']$, то $f \subseteq g$, если и только если $f' \subseteq g'$, т. е. отношение $\subseteq$ устойчиво относительно $\simeq$.

Нашей целью было сделать отношение $\subseteq$ антисимметричным. Действительно, если $[f] \subseteq [g]$ и $[g] \subseteq [f]$, то $f \subseteq g$ и $g \subseteq f$, следовательно, $f \simeq g$, откуда $[f] = [g]$. Таким образом, подобъекты (в смысле нового определения) объекта d образуют ЧУМ $(\mathrm{Sub}(d), \subseteq)$.

Далее, стираем различие между классом эквивалентности и представителем этого класса. Будем использовать термин «подобъект f» вместо «подобъект $[f]$» и писать $f \subseteq g$, когда строго говоря, надо записывать $[f] \subseteq [g]$ и т. д. Все рассматриваемые свойства подобъектов и построения, использующие подобъекты, будут однако устойчивы относительно эквивалентности (на самом деле будут категориями, они определены лишь с точностью до изоморфизма). Такая вольность в обозначениях облегчает изложение материала, однако вопрос о равенстве требует пунктуальности. В этой связи выражение $f \simeq g$ будет использоваться, когда подразумевается, что f и g являются одним и тем же под-

$$\begin{tikzcd}
b \arrow[dd, bend left, "i"] \arrow[dr, tail] & \\
& d \\
a \arrow[uu, bend left, "h"] \arrow[ur, tail] &
\end{tikzcd}$$

Рис. 4.68. Процедура образования подобъектов из объектов

$$\begin{tikzcd}
b^a \times a \arrow[dr, "ev"] & \\
& b \\
1 \times a \arrow[uu, "\ulcorner f \urcorner \times 1_a"] \arrow[ur, "f \circ pr_a"'] &
\end{tikzcd}$$

Рис. 4.69. Пояснение к понятию «имена стрелок»

объектом, т. е. $[f]=[g]$, в то время как запись $f=g$ будет обозначать, что f и g в действительности являются одной и той же стрелкой.

Элементы. Элемент x множества $A (x \subseteq A)$ можно отождествить с одноэлементным подмножеством $\{x\}$ множества A и, следовательно, со стрелкой, обозначающей включение x в A из конечного объекта $\{x\}$ в A. Наоборот, функция $f: 1 \to A$ определяет в категории Set элемент из A, именно f-образ единственного элемента конечного объекта 1. Итак, если категория $\mathcal{G}$ имеет конечный объект 1, то *элементом* $\mathcal{G}$-объекта a называется всякая $\mathcal{G}$-стрелка $x: 1 \to a$. Заметим, что $x: 1 \to a$ всегда монострелка.

Имена для стрелок. Функция $f: A \to B$ в множестве B является *элементом* множества B^A, т. е. $f \in B^A$. Поэтому возникает функция $\ulcorner f \urcorner : \{0\} \to B^A$, такая, что $\ulcorner f \urcorner (0) = f$. Далее, если x — элемент из A, то имеется категориальный «элемент» $\bar{x}: \{0\} \to A$, определенный равенством $\bar{x}(0) = x$. Так как $ev(<f, x>) = f(x)$, то $ev \circ < \ulcorner f \urcorner, \bar{x} > (0) = ev(\ulcorner f \urcorner (0), \bar{x}(0)) = f(x) = f(x(0))$, и имеем равенство

$$ev \circ \langle \overline{f}, \bar{x} \rangle = f \circ \bar{x}.$$

Эта ситуация может быть распространена на произвольную категорию, допускающую экспоненцирование. Для данной $\mathcal{G}$-стрелки $f: a \to b$ построим стрелку $f \circ pr_a: 1 \times a \to b$, являющуюся композицией $f \circ pr_a: 1 \times a \to a \to b$. Тогда *именем* f называется стрелка $\ulcorner f \urcorner : 1 \to b^a$, являющаяся экспоненциально присоединенной к стрелке $f \circ pr_a: 1 \times a \to b$. Таким образом, $\ulcorner f \urcorner$ — это единственная стрелка, для которой диаграмма на рис. 4.69 коммутативна.

Для произвольного $\mathcal{G}$-элемента $x: 1 \to a$ объекта a имеет место равенство

$$ev \circ \langle \overline{f}, x \rangle = f \circ x.$$

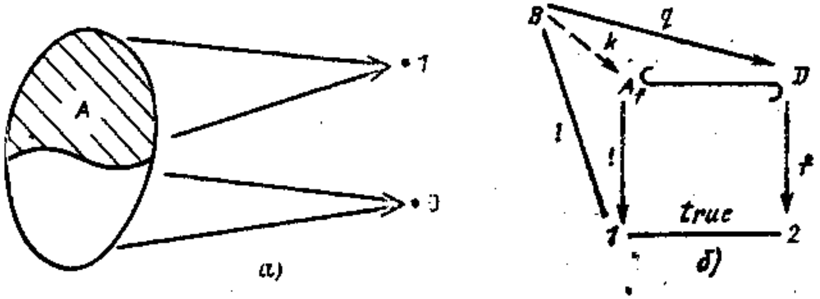


Рис. 4.70. Пояснение к понятию характеристической функции

Классификация подобъектов. В теории множеств множество-степень $\mathcal{P}(D)$ обозначается через 2^D. Последний символ согласно нашим обозначениям определяет совокупность всех функций из D в $2 = \{0, 1\}$.

Оправдание такого обозначения заключается в наличии изоморфизма $\mathcal{P}(D) \cong 2^D$, т. е. взаимно однозначного соответствия между подмножествами множества D и фукциями $D \to 2$. Этот изоморфизм устанавливается следующим образом. Для данного подмножества $A \subseteq D$ определим функцию $\chi_A: D \to 2$, называемую *характеристической функцией* множества A, с помощью правила: для элементов из D, принадлежащих A, значение этой функции равно 1, а для элементов, не принадлежащих A, оно равно 0, т. е.

$$\chi_A(x) = \begin{cases} 1, & \text{если} \quad x \in A; \\ 0, & \text{если} \quad x \notin A. \end{cases}$$

Отображение (рис. 4.70, *а*), которое ставит в соответствие множеству A функцию χ_A, является инъектным отображением из $\mathcal{P}(D)$ в 2^D, т. е. если $\chi_A = \chi_B$, то $A = B$. Оно также сюръективно. Действительно, если $f \in 2^D$, то $f = \chi_A$, где

$$A_f = \{x : x \in D \quad \text{и} \quad f(x) = 1\}.$$

Это соответствие между подмножествами и характеристическими функциями может быть выражено категориально с помощью понятия обратного образа. Множество A_f, определенное выше, является прообразом при отображении f $\{1\}$ множества $\{0, 1\}$, т. е.

$$A_f = f^{-1}(\{1\}).$$

В соответствии с ранее изложенным диаграмма на рис. 4.71 является декартовым квадратом, т. е. A_f получается подъемом стрелки, обозначающей включение $\{1\}$ в 2, вдоль 1. Модифицируем эту диаграмму. Нижнюю стрелку на рис. 4.71 заменим

$$\begin{tikzcd}
A_f \arrow[r, hook] \arrow[d, "!"'] & B \arrow[d, "f"] \\
\{1\} \arrow[r, hook] & 2
\end{tikzcd}$$

Рис. 4.71. Характеристические функции и обратный образ

$$\begin{tikzcd}
A \arrow[r, hook] \arrow[d, dash, "!"'] & D \arrow[d, dash, "\chi_A"] \\
1 \arrow[r, dash, "true"] & 2
\end{tikzcd}$$

Рис. 4.72. Пояснение к переходу от множеств к характеристической функции

функцией из $1=\{0\}$ в $2=\{0, 1\}$, принимающей значение 1. Обозначим эту функцию через true (истина). Таким образом, true $(0)=1$ и внутренний квадрат в диаграмме рис. 4.70, *б* декартов. Действительно, допустим, что внешний «квадрат» коммутативен. Тогда если $b\in B$ и $f(g(b))=\text{true}\ (!(b))=1$, то $g(b)\in A_f$. Поэтому при $k: B\to A_f$, определенном равенством $k(b)=g(b)$, вся диаграмма будет коммутативна, и ясно, что она коммутативна только при одном k. Следовательно, если $A\subseteq D$, то квадрат на рис. 4.72 декартов, так как подъем true вдоль χ_A дает множество $\{x:\chi_A(x)=1\}$, равное A. Более того, отсюда следует, что χ_A можно определить как единственную функцию из D в 2, превращающую диаграмму на рис. 4.72 в декартов квадрат, т. е. как единственную функцию, подъем вдоль которой функции true дает A. Действительно, если для некоторой функции f внутренний квадрат в диаграмме на рис. 4.73 декартов, то для $x\in$A будет $f(x)=1$, так что $x\in A_f$. Поэтому $A\subseteq A_f$. Так как внешний «квадрат» коммутативен (в действительности, как мы уже видели выше, он является декартовым), то существует единственная функция k, удовлетворяющая равенству $i\circ k=j$. Так как функции i и j являются включениями, k должна быть включением. Таким образом, $A_f\subseteq A$, что вместе с предыдущим дает $A_f=A$. Но f является характеристической функцией множества A_f. Поэтому $f=x_A$.

$$\begin{tikzcd}
A_f \arrow[rrd, hook, "f"] \arrow[rd, dashed, "k"] \arrow[rdd, "!"'] & & \\
& A \arrow[r, hook, "i"] \arrow[d, "!"] & D \arrow[d, "f"] \\
& 1 \arrow[r, dash, "true"] & 2
\end{tikzcd}$$

Рис. 4.73. Характеристические функции и множества

$$\begin{tikzcd}
a \arrow[r, tail, "f"] \arrow[d, "!"'] & d \arrow[d, "\chi_f"] \\
1 \arrow[r, "true"] & \Omega
\end{tikzcd}$$

Рис. 4.74. Пояснение к понятию классификатора подобъектов

Итак, множество, состоящее из одного элемента 2, вместе с функцией true : $1\to 2$ играет особую роль в переходе от подмножеств к характеристическим функциям. Эта роль может быть выражена на категориальном языке, что приводит к следующему определению.

Определение. Пусть $\mathcal{G}$ — категория с конечным объектом 1. *Классификатором подобъектов* для $\mathcal{G}$ называется $\mathcal{G}$-объект Ω вместе с $\mathcal{G}$-стрелкой true : $1\to\Omega$, такой, что выполняется следующая аксиома.

Ω-аксиома.Для каждой моностреяки $f: a\rightarrowtail d$ существует одна и только одна $\mathcal{G}$-стрелка $x_f: d\to 0$, для которой диаграмма на рис. 4.74 является декартовым квадратом.

Стрелка x_f называется *характеристической стрелкой*, или *характером* мoнострелки f (подобъекта объекта d). Стрелка true будет обозначаться как T. Сам объект Ω называется классифицирующим объектом.

Классификатор подобъектов, если он существует, *единствен с точностью до изоморфизма*. Пусть $T: 1\to\Omega$ и $T: 1\to\Omega$ являются классификаторами подобъектов. Рассмотрим диаграмму на рис. 4.75, *а*. Верхний ее квадрат декартов и определяет характеристическую стрелку χ'_T стрелки T на основании того, что T' является классификатором подобъектов (напомним, что каждая стрелка, имеющая dom, равный конечному объекту, будет мономорфна). Нижний квадрат также декартов и определяет характеристическую стрелку для T', используя T как классификатор.

В соответствии с леммой внешний прямоугольник на рис. 4.75, *б* декартов. В силу Ω-аксиомы имеется только одна стрелка $\Omega\to\Omega$, для которой этот квадрат декартов. Так как для стрелки $1_\Omega:\Omega\to\Omega$, указанный квадрат декартов $\chi_{T'}\circ\chi'_T=1_\Omega$. Меняя ролями T и T' в этом рассуждении, получаем

$$\chi'_T\circ\chi_{T'}=1_{\Omega'}.$$

$$\begin{tikzcd}
1 \arrow[r, tail, "T"] \arrow[d] & \Omega \arrow[d, "\chi_{T'}"] \\
1 \arrow[r, tail, "T"] \arrow[d] & \Omega' \arrow[d, "\chi_{T'}"] \\
1 \arrow[r, tail, "T"] & \Omega
\end{tikzcd}
\qquad
\begin{tikzcd}
1 \arrow[r, "T"] \arrow[d] & {} \arrow[d, "\chi_p\circ\chi'_T"] \\
1 \arrow[r, "T"] & {}
\end{tikzcd}$$

а) б)

Рис. 4.75. Пояснение к понятию единственности классификатора подобъектов

$$\begin{tikzcd}
b \arrow[dr, dashed, "k"] \arrow[drr, "g"] \arrow[ddr] & & \\
& a \arrow[r, tail, "f"] \arrow[d] & d \arrow[d, "\chi_f"] \\
& 1 \arrow[r, "true"] & \Omega
\end{tikzcd}
\qquad
\begin{tikzcd}
a \arrow[r, "f"] \arrow[d] & d \arrow[d, "h"] \\
1 \arrow[r, "true"] & \Omega
\end{tikzcd}$$

a) б)

Рис. 4.76. Характеристическая стрелка и эквивалентность

Таким образом, $\chi_{T'}:\Omega'\cong\Omega$.

Из предыдущего следует, что для любых двух классификаторов подобъектов один из них получается из другого композицией с некоторой изострелкой, соединяющей их концы. Переход от f к χ_f устанавливает взаимно однозначное соответствие между подобъектами объекта d и стрелками $d\to\Omega$, как показывает следующая теорема.

Теорема 4.3. Для любых $f:a\rightarrowtail d$ и $g:b\rightarrowtail d$ $f\simeq g$ тогда и только тогда, когда $\chi_f=\chi_g$.

Доказательство. Допустим сначала, что $\chi_f=\chi_g$. Рассмотрим диаграмму на рис. 4.76, *а*. Так как $\chi_f=\chi_g$, то внешний квадрат коммутативен (в действительности он декартов). В силу свойства универсальности внутреннего декартова квадрата существует стрелка k, пропускающая g через f. Поэтому $g\subseteq f$. Меняя местами f и g в этой диаграмме, получаем включение $f\subseteq g$. Следовательно, $f\simeq g$.

Наоборот, пусть $f\simeq g$ и внутренний квадрат — декартов. Тогда существует изострелка $k:b\to a$, для которой верхний треугольник коммутативен. Используя это обстоятельство, можно показать, что внешний квадрат тоже декартов. По Ω-аксиоме, примененной к моностреле g, получаем $\chi_f=\chi_g$.

Функция, сопоставляющая χ_f моностреле f (точнее, подобъекту $[f]$), вкладывает Sub (d) в $\mathscr{G}(d,\Omega)$. Кроме того, если для произвольной стрелки $h:d\to\Omega$ поднять true вдоль h (рис. 4.76, *б*), то возникающая при этом стрелка f будет мономорфна (так как true-монострелка и обратный образ монострелки сам является монострелкой). Поэтому h должна совпадать с χ_f. Таким образом, в категории, обладающей классификатором подобъектов,

$$\mathrm{Sub}(d)\cong\mathscr{G}(d,\Omega).$$

Пример 4.43. Set-топос. Это простой пример, который служит обоснованием введения понятия топоса.

$$\begin{tikzcd}
P \arrow[r, "j"] \arrow[d, "i"'] & C \arrow[d, "h"] \\
A \arrow[r, "f"] & E
\end{tikzcd}
\qquad
\begin{tikzcd}
Q \arrow[r, "v"] \arrow[d, "u"'] & D \arrow[d, "k"] \\
B \arrow[r, "q"] & F
\end{tikzcd}$$

Рис. 4.77. Пояснение к примеру Set²-топос

$$\begin{tikzcd}
\langle P,Q\rangle \arrow[r, "\langle j,v\rangle"] \arrow[d, "\langle i,u\rangle"'] & \langle C,D\rangle \arrow[d, "\langle h,k\rangle"] \\
\langle A,B\rangle \arrow[r, "\langle f,q\rangle"'] & \langle E,F\rangle
\end{tikzcd}$$

Рис. 4.78. Декартов квадрат Set³

Пример 4.44. Finset-топос. Пределы, экспонирования и классификатор подобъектов $T:1\to\Omega$ точно такие же, как в Set.

Пример 4.45. Finord-топос. Каждое конечное множество изоморфно некоторому конечному ординалу ($A\cong n$, если A имеет n элементов). Поэтому все категориальные построения в Finest переносятся в Finord (как мы уже видели, это для произведений и экспоненциалов). Классификатор подобъектов в Finord представляет собой ту же самую функцию true : $\{0\}\to\{0,\ 1\}$, что и в Finset и в Set.

Пример 4.46. Set$^{\to}$, т. е. категория функций, являтся топосом. Все конструкции получаются «удвоением» соответствующих конструкций в Set.

Конечным объектом служит пара $<\{0\},\ \{0\}>$ двух стрелок $<f,\ g>$:

$$\langle A,\ B\rangle\to\langle\mathscr{G},\ F\rangle,\ \langle n,\ k\rangle:\langle C,\ D\rangle\to\langle E,\ F\rangle \text{ в } \mathrm{Set}^3$$

с общим концом. Построим в категории Set декартовы квадраты (рис. 4.77).

Тогда квадрат, изображенный на рис. 4.78, будет декартовым в Set².

Экспоненциалом $<C,\ D>^{<A,\ B>}$ служит пара $<C^A,\ D^B>$, а стрелкой значения из объекта $<C,\ D>^{<A,\ B>}\times<A,B>=<C^A\times A,\ D^B\times B>$ в объект $<C,\ D>$ — пара соответствующих стрелок значения $e:C^A\times A\to C$ к $f:D^B\times B\to D$ в Set.

Пара $<T,\ T>:<\{0\},\ \{0\}>\to<2,2>$ является классификатором подобъектов. Категория Set не играет здесь никакой особой роли. Если $\mathscr{G}_1$ и $\mathscr{G}_2$ — топосы, то произведение категорий $\mathscr{G}_1\times\ \mathscr{G}_2$ также будет топосом.

Пример 4.47. Set$^{\to}$-топос, т. е. категория функций. Конечным объектом служит тождественная функция $\mathrm{id}_{\{0\}}$ из $\{0\}$ в $\{0\}$.

Обратный образ. Рассмотрим «куб» на рис. 4.79, *а*, где f, g, h — данные Set$^{\to}$-объекты; $<i,\ j>$ есть Set$^{\to}$-стрелка из f в g и

$$\begin{tikzcd}
P \arrow[rr, "r"] \arrow[dr, "k"] \arrow[dd, "u"'] & & C \arrow[dr, "h"] \arrow[dd, "p"] & \\
& Q \arrow[rr, "s"] \arrow[dd] & & D \arrow[dd, "g"] \\
A \arrow[rr, "i"] \arrow[dr, "f"'] & & E \arrow[dr, "g"] & \\
& B \arrow[rr, "j"'] & & F
\end{tikzcd}$$

а)

$$\begin{tikzcd}
Q \arrow[r, "s"] \arrow[d, "v"'] & D \arrow[d, "g"] \\
B \arrow[r, "j"'] & F
\end{tikzcd}
\qquad
\begin{tikzcd}
P \arrow[r, "r"] \arrow[d, "u"'] & C \arrow[d, "p"] \\
A \arrow[r, "i"'] & E
\end{tikzcd}$$

б)

Рис. 4.79. Пояснение к понятию обратного образа

$<p, g>$ есть Set-стрелка из h и g. Оставшаяся часть диаграммы получается образованием в Set декартовых квадратов (рис. 4.79, *б*), а стрелка k возникает в силу свойства универсальности обратного образа стрелок j и g. Тогда $\mathrm{Set}^{\rightarrow}$-стрелки $<u, v>$ и $<r, s>$ составляют обратный образ стрелок $<i, j>$ и $<\varphi, q>$.

Классификатор. Если $f: A \rightarrow B$ — подобъект $g: C \rightarrow D$ в $\mathrm{Set}^{\rightarrow}$ и пара $<i, j>$ является $\mathrm{Set}^{\rightarrow}$-моностредкой из f в g, то ее компоненты будут монострелками в Set и имеет место коммутативная диаграмма (рис. 4.80, *а*). Можно выбрать монострелки i и j действительными включениями так, что $A \subseteq C$, $B \subseteq D$ и f является ограничением g, т. е. $f(x)=g(x)$ для всех $x \in A$ (рис. 4.80, *б*). Элементы из множества можно разделять на следующие три класса:

(1) $x \in A$;

(2) $x \notin A$, где $g(x) \in B$;

(3) $x \notin A$, где $g(x) \notin B$.

Введем трехэлементное множество $\{0, 1/2, 1\}$ и определим $\psi: C \rightarrow \{0, 1/2, 1\}$ условиями

$$\begin{tikzcd}
A \arrow[r, tail, "i"] \arrow[d, "f"'] & C \arrow[d, "g"] \\
B \arrow[r, tail, "j"'] & D
\end{tikzcd}$$

а)

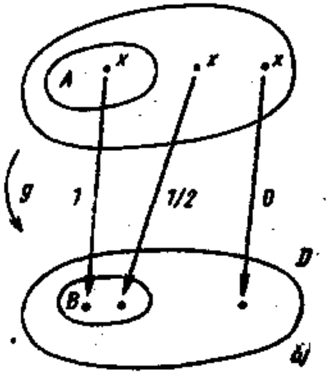


Рис. 4.80. Пояснение к понятию классификатора

$$\begin{tikzcd}
A \arrow[rr, hook, "i"] \arrow[dr, "f"] \arrow[dd] & & C \arrow[dr, "g"] \arrow[dd, "\psi"] & \\
& B \arrow[rr, hook, "j"] \arrow[dd] & & D \arrow[dd] \\
\{0\} \arrow[rr, "t'"] \arrow[dr, "id_{\{0\}}"'] & & \{0,1/2,1\} \arrow[dr, "t"] & \\
& \{0\} \arrow[rr, "true"'] & & \{0,1\}
\end{tikzcd}$$

Рис. 4.81. Пояснение к разделению множества C на три класса

$$\psi(x)=\begin{cases} 1, & \text{если имеет место (1);} \\ 1/2, & \text{если имеет место (2);} \\ 0, & \text{если имеет место (3).} \end{cases}$$

Построим куб (рис. 4.81), где $\mathrm{true}(0)=t'(0)=1$ и $t(0)=0$, $t(1)=t(1/2)=1$. Через χ_B обозначена характеристическая функция множества B.

Нижняя грань куба представляет классификатор подобъектов $T: 1 \rightarrow \Omega$ в $\mathrm{Set}^{\rightarrow}$. Пара $<t', \mathrm{true}>$ является $\mathrm{Set}^{\rightarrow}$-стрелкой из $1=\mathrm{id}_{\{0\}}$ в $\Omega=t:\{0, 1/2, 1\} \rightarrow \{0, 1\}$. Передняя и задняя грани куба являются декартовыми квадратами в Set. Вся диаграмма показывает, что пара $<\psi, \chi_B>$ является характеристической $\mathrm{Set}^{\rightarrow}$-стрелкой монострелки $<i, j>$.

Экспоненцирование. Пусть $f: A \rightarrow B$, $g: C \rightarrow D$ — два $\mathrm{Set}^{\rightarrow}$-объекта. Определим $\mathrm{Set}^{\rightarrow}$-объект $g^f: E \rightarrow F$. Пусть $F=D^B$ (экспоненциал в Set), а E — совокупность всех $\mathrm{Set}^{\rightarrow}$-стрелок из f в g, т. е.

$$E=\{\langle h, k\rangle : \text{диаграмма на рис. 4.82, } a \text{ коммутативна}\}$$

и

$$g^f[\langle h, k\rangle]=K.$$

$$\begin{tikzcd}
A \arrow[r, "h"] \arrow[d, "f"'] & C \arrow[d, "k"] \\
B \arrow[r, "k"'] & D
\end{tikzcd}
\qquad
\begin{tikzcd}
E \times A \arrow[r, "u"] \arrow[d, "g^f \times f"'] & C \arrow[d, "g"] \\
F \times B \arrow[r, "v"'] & D
\end{tikzcd}$$

а) *б)*

Рис. 4.82. Пояснение к механизму экспоненцирования в примере 4.47

Произведение в Set$^{\to}$-объектов g^f и f равно их функторному произведению $g^f\times f: E\times A\to F\times B$ в Set, а пара $<u, v>$ является стрелкой значения из $g^f\times f$ в g (диаграмма на рис. 4.82, б), где v — обычная стрелка значения в Set, а u преобразует пару $\ll h, k>, x>$ в $h(x)$.

Приведенные конструкции для классификатора $T:1\to\Omega$ и экспоненциала g^f являются частным случаем более общего определения, дающего целый класс топосов.

4.7. ФУНКТОРЫ И ТОПОСЫ

Различные топосы, рассмотренные в предыдущем параграфе, могут быть построены как категории «Set-значных функторов» следующим образом (рис. 4.83).

1. Set^2. Множество $2=\{0, 1\}$ является дискретной категорией. Функтор $F:2\to$Set сопоставляет множество F_0 объекту 0 и множество F_1 — объекту 1. Так как F как функтор должен сохранять единичные стрелки, то можно отождествить F с парой $<F_0, F_1>$. Таким образом, функторы $2\to$Set являются, в сущности, объектами категории Set^2 пар множеств. Далее, для двух функторов F и G, отождествленных с парами $<F_0, F_1>$ и $<G_0, G_1>$ соответственно, естественное преобразование $\tau:F_0\to G_0$ имеет два компонента: $\tau_0:F_0$ и $\tau_1:F_1\to G_1$. Можно отождествить τ с парой $<\tau_0, \tau_1>$, являющейся ничем другим, как Set^{2}стрелкой из $<F_0, F_1>$ в $<G_0, G_1>$.

2. Set$^{\to}$. Рассмотрим категорию $2=\{0, 1\}$ с одной неединичной стрелкой $0\to1$. Функтор $F:2\to$Set определяется множеством F_0, F_1 и функцией $f:F_0\to F_1$. Таким образом, F — это, в сущности, стрелка f категории Set, т. е. объект из категории Set$^{\to}$. Если G — другой такой функтор, определяемый функцией $g:G_0\to G_1$, то естественное преобразование $\tau:F\to G$ имеет такие компоненты τ_0 и τ_1, что диаграмма на рис. 4.84 коммутативна. Таким образом, τ, отождественная с парой $<\tau_0; \tau_1>$, является стрелкой из f в g категории Set$^{\to}$. Поэтому последняя категория представляет собой категорию Set^2 функторов из 2 в Set.

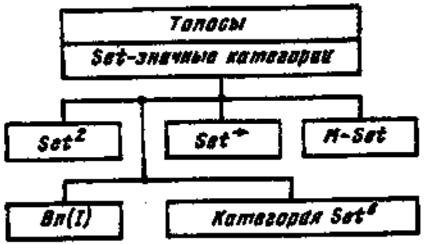
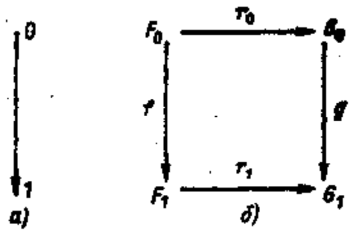


Рис. 4.83. Функторы и топосы. Категории и топосы


$$\begin{tikzcd}
0 \arrow[d] \\
1
\end{tikzcd}
\qquad
\begin{tikzcd}
F_0 \arrow[r, "\tau_0"] \arrow[d, "f"'] & G_0 \arrow[d, "g"] \\
F_1 \arrow[r, "\tau_1"] & G_1
\end{tikzcd}$$

а) б)

Рис. 4.84. Пояснение к понятию топоса как категории M—Set-значного функтора

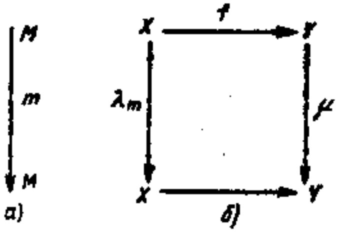
3. M — Set. Пусть $M=(M, \ast, e)$ — произвольный моноид; M — множество, представляющее собой пару (X, λ), где X — множество, а λ — функция, сопоставляющая каждому $m\in M$ функцию $\lambda_m:X\to X$ так, что выполняются следующие условия:

$$(1)\ \lambda_e=\mathrm{id}_x;$$

$$(2)\ \lambda_m\circ\lambda_p=\lambda_{m\ast p}.$$

Моноид M определяет категорию с одним объектом, скажем M, стрелками служат элементы $m_э$ множества M; операция $\ast$ является операцией композиции, а e — единичной стрелкой. Тогда функцию λ можно рассматривать как функтор $\lambda: M\to\to$Set, такой, что $\lambda(M)=X$ для единственного объекта категории M, и $\lambda(m)=\lambda_m$ для каждой стрелки m. Тогда условия (1) и (2) выражают тот факт, что λ — функтор. Для любого другого функтора $\mu:M\to$Set, такого, что $\mu(M)=Y$, естественное преобразование $\tau:\lambda\to\mu$ ставит в соответствие объекту M функцию $f:x\to Y$, для которого диаграмма на рис. 4.85 коммутативна при каждом $m\in M$. Но это в точности означает, что f — инвариантное отображение из (X, λ) в (Y, μ). Таким образом, категорию M—Set$^{\mu}$ можно отождествить с категорией Set^M функторов из M в Set.

4. Вп (I). Если рассматривать множество I как дискретную категорию, то функтор $F:I\to$Set, сопоставляющий каждому $i\in I$ множество F_i, можно отождествить с семейством $\{F_i: i\in I\}$ множеств, заиндексированных множеством I. Объект (X, f) из Вп (I) (т. е. функция $\bar f:X\to I$) определяет функтор $\bar f:I\to$Set, такой, что $\bar f(i)$ равно f^{-1} $(\{i\})$, т. е. слою над i. Произвольная стрелка $h:(X, f)\to(Y, g)$ является функцией, отображающей f-слой над i и g-слой над i. Поэтому она определяет функцию $h_i:\bar f(i)\to\bar g(i)$, являющуюся сужением h на множество $\bar f(i)$. Эти функции hi служат компонентами естественного преобразования $\bar h:\bar f\to\bar g$. Таким образом, каждое расслоение над I определяет функтор из I в Set. Обращение этой конструкции возможно, если все множества F_i попарно не пересекаются. Поэтому для данного функтора $F:I\to$Set определим новый функтор $\bar F:I\to$Set, полагая $\bar F(i)=$ $=F(i)\times\{i\}$, а затем преобразуем семейство $\{\bar F(i):i\in I\}$ в расслое-


$$\begin{tikzcd}
M \arrow[d, "m"] \\
M
\end{tikzcd}
\qquad
\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "\lambda_m"'] & Y \arrow[d, "\mu"] \\
X \arrow[r] & Y
\end{tikzcd}$$

а) б)

Рис. 4.85. Пояснения к категориям функторов

$$\begin{tikzcd}
K(a) \arrow[r, "\mu_a"] \arrow[d, "\lambda_a"'] & G(a) \arrow[d, "\sigma_a"] \\
F(a) \arrow[r, "\tau_a"] & H(a)
\end{tikzcd}$$

а)

$$\begin{tikzcd}
K(a) \arrow[rr, "\mu_a"] \arrow[dr, dashed] \arrow[dd, "\lambda_a"'] & & G(a) \arrow[dr, "G(f)"] \arrow[dd, "\sigma(a)"' near start] & \\
& K(b) \arrow[rr] \arrow[dd, "\lambda_b" near end] & & G(b) \arrow[dd, "\sigma_b"] \\
F(a) \arrow[rr, "\tau_a" near start] \arrow[dr, "F(f)"'] & & H(a) \arrow[dr, "H(f)"'] & \\
& F(b) \arrow[rr, "\tau_b"'] & & H(b)
\end{tikzcd}$$

б)

$$\begin{tikzcd}
K \arrow[r, "\mu"] \arrow[d, "\lambda"'] & G \arrow[d, "\sigma"] \\
F \arrow[r, "\tau"] & H
\end{tikzcd}$$

в)

Рис. 4.86. Пояснение к понятию обратного образа

ние над I. Так как $F(i)\cong F(i)x\{i\}$, то функторы F и $\overline{F}$, естественно, изоморфны. В результате получается, что переход от (x, f) к $\overline{f}$ является эквивалентностью категорий. Категория Bn (I) расслоений над I эквивалентна категории Set^I Set-значных функторов, определенных на I. Приведенные четыре примера поясняют конструкцию, снабжающую нас самыми разнообразными топосами, и позволяют сформулировать следующее положение: *для произвольно малой категории $\mathscr{G}$ категория $\mathrm{Set}^{\mathscr{G}}$ функторов является топосом.*

Конечный объект. В $\mathrm{Set}^{\mathscr{G}}$ им является постоянный функтор $1:\mathscr{G}\to\mathrm{Set}$, ставящий в соответствие каждому $\mathscr{G}$-объекту одноэлементное множество $\{0\}$, а каждой $\mathscr{G}$-стрелке — тождественное отображение на $\{0\}$. Для произвольного функтора $F:\mathscr{G}\to\to\mathrm{Set}$ существует единственная $\mathrm{Set}^{\mathscr{G}}$-стрелка $F\dot{\to}1$, состоящая из единственных функций: $F(a)\to\{0\}$ для каждого $\mathscr{G}$-объекта a.

Обратный образ. Обратный образ определяется «покомпонентно», как и все пределы и копределы в $\mathrm{Set}^{\mathscr{G}}$. Пусть $\tau:G\dot{\to}H$ и $\sigma:G\dot{\to}H$ — произвольные естественные преобразования. Для каждого $\mathscr{G}$-объекта a построим обратный образ (рис. 4.86,*а*) в Set компонентов τ_a и σ_a. Сопоставляя объекту a объект $K\ (a)$, получаем функтор $K:\mathscr{G}\to\mathrm{Set}$, ставящий в соответствие $\mathscr{G}$-стрелке $f:a\to b$ единственную стрелку $K(f):K(a)\to\to K(b)$, получаемую по свойству универсальности переднего декартова квадрата «кубической» диаграммы (рис. 4.86,*б*). Стрелки λ_a и μ_a являются компонентами естественных преобразований $\lambda:K\dot{\to}F$ и $\mu:K\dot{\to}G$, а диаграмма на рис. 4.86,*в* будет декартовым квадратом в $\mathrm{Set}^{\mathscr{G}}$. *Классификатор подобъектов.* Для его определения нам понадобится новое понятие. Для данного $\mathscr{G}$-объекта a обозначим через S_a совокупность всех $\mathscr{G}$-стрелок, начинающихся в a, т. е.

$$S_a=\{f:\mathrm{dom}\,f=a\}$$

(S_a является классом объектов относительной категории $\mathscr{G}\uparrow a$ объектов над a). Множество S_a замкнуто относительно левого умножения, т. е. если $f:a\to b$, $f\in S_a$, $g:b\to c$ — произвольная $\mathscr{G}$-стрелка, то $g\circ f\in S_a$ (рис. 4.87).

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[dr, "g\circ f"'] & b \arrow[d, "g"] \\
& {}
\end{tikzcd}$$

Рис. 4.87. Пояснение к понятию классификатора объектов

Корешетом на a, или a-корешетом, называется всякое подмножество S множества S_a, замкнутое относительно левого умножения, т. е. такое, что если $f\in S$, то $g\circ f\in S$ для любой S-стрелки. Для всякого объекта a существуют, по крайней мере, два a-корешета: S_a и $\varnothing$ (пустое корешето).

Пример 4.48. В дискретной категории имеем $S_a=\{1_a\}$; множества S_a и $\varnothing$ являются единственными a-корешетами.

Пример 4.49. В категории 2 с единственной неединичной стрелкой $f:0\to1$ существуют три корешета на 0, а именно $\varnothing$, $S_0=\{1_0, f\}$ и $\{f\}$.

Пример 4.50. В однообъектной категории (моноиде) M-корешета на M — это в точности левые идеалы в M.

Определим теперь классифицирующий объект $\Omega:\mathscr{G}\to\mathrm{Set}$. Положим $\Omega(a)=\{S:S$ есть a-корешето$\}$. Для произвольной $\mathscr{G}$-стрелки $f:a\to b$ пусть $\Omega(f):\Omega(a)\to\Omega(b)$ обозначает функцию, сопоставляющую a-корешету S b-корешето $\left\{b\xrightarrow{g}c:g\circ f\in S\right\}$. Для категории Set^M множество $\Omega(M)$ равно множеству L_M левых идеалов в M, а для стрелки $m:M\to M$ функция $\Omega(m):L_M\to\to L_M$ ставит в соответствие левому идеалу S' множество $\{n:n\ast m\in S\}=\omega(m, S)$. Таким образом, Ω оказывается действием ω моноида M на L_M. Определим $\mathrm{Set}^{\mathscr{G}}$ стрелку $T:1\to\Omega$ как единственное преобразование с компонентами $T_a:\{0\}\to\to\Omega(a)$, задаваемыми равенством $T_a(0)=S_a$, т. е. $T_a\ (0)$ равно наибольшему корешету на a. *Так, определенная стрелка T является классификатором в $\mathrm{Set}^{\mathscr{G}}$*. Докажем это. Предположим, что

$$\begin{tikzcd}
F(a) \arrow[r, hook, "\tau_a"] \arrow[d, "F(f)"'] & G(a) \arrow[d, "G(f)"] \\
F(b) \arrow[r, hook, "\tau_b"] & G(b)
\end{tikzcd}$$

Рис. 4.88. Пояснение к классификатору подобъектов

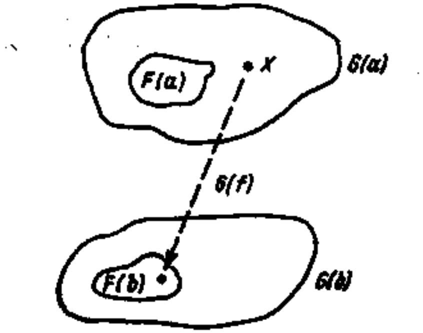


Рис. 4.89. Пояснение к утверждению, что стрелка T является классификатором

$F \to G$ — монострелка в $\mathrm{Set}^{\mathscr{G}}$. Тогда для каждого $\mathscr{G}$-объекта a компонент $\tau_a : F(a) \to G(a)$ является мономорфной в Set стрелкой. Будем считать, что этот компонент на самом деле является включением F_a в $G(a)$. Характеристическая стрелка $\chi_\tau : G \to \Omega$ подобъекта τ должна быть естественным преобразованием, каждый компонент $(\chi_\tau)_a$ которого является теоретико-множественной функцией из $G(a)$ в $\Omega(a)$. Функция $(\chi_\tau)_a$ сопоставляет $x \in G(a)$ некоторое a-корешето $(\chi_\tau)_a(x)$. Надо определить, какие стрелки $f : a \to b$, начинающиеся из объекта a, принадлежат множеству $(\chi_\tau)_a(x)$. Для каждой стрелки $f : a \to b$ имеем коммутативную диаграмму (рис. 4.88), так что $F(f)$ является ограничением $G(f)$ на $F(a)$. Зачислим f в множество $(\chi_\tau)_a(x)$, если и только если $G(f)$ переводит x в $F(b)$ (рис. 4.89). Таким образом, $(\chi_\tau)_a(x) = \{f : a \to b;\ G(f)(x) \in F(b)\}$. В более общем случае, когда τ_a не обязательно является функцией включения, положим

$$(\chi_\tau)_a(x) = \left\{a \xrightarrow{f} b : G(f)\ (x) \in \tau_b(F(b))\right\} =$$

$$= \left\{a \xrightarrow{f} b : \text{существует } y \in F(b), \text{ такое, что } G(f)(x) = \tau_b(y)\right\}.$$

Понятием, двойственным к корешету, является *a-решето*. Таким образом, a-решето — это совокупность стрелок с концом в a, замкнутая относительно правого умножения. Решета используются для доказательства того, что категория контравариантных функторов из $\mathscr{G}$ в Set является топосом. Для нас корешета важны, в частности, для изложения их с помощью семантики Крипке [46].

Экспоненцирование в $\mathrm{Set}^{\mathscr{G}}$. Пусть $F : \mathscr{G} \to \mathrm{Set}$ — произвольный функтор. Для каждого $\mathscr{G}$-объекта a определим пренебрегающий функтор $F_a : \mathscr{G}{\uparrow}a \to \mathrm{Set}$, ставящий в соответствие стрелке $f : a \to b$ множество $F(b)$, а стрелке $h : f \to g$, такой, что диаграмма на рис. 4.90 коммутативна,— функцию $F(h)$.

$$\begin{tikzcd}
& a \arrow[dl, "f"'] \arrow[dr, "g"] & \\
b \arrow[rr, "h"] & & c
\end{tikzcd}$$

Рис. 4.90. Пояснение к понятию пренебрегающего функтора

$$\begin{tikzcd}
a \arrow[r, "k"] \arrow[dr, "f \circ k"'] & d \arrow[d, "f"] \\
& b
\end{tikzcd}$$

Рис. 4.91 Пояснение к экспоненцированию в $\mathrm{Set}^{\mathscr{G}}$

Для данных функторов $F, G : \mathscr{G} \to \mathrm{Set}$ определим функтор $G^F : \mathscr{G} \to \mathrm{Set}$. Положим

$$G^F(a) = \mathrm{Nat}\,[F_a,\ G_a],$$

где $\mathrm{Nat}\,[F_a,\ G_a]$ — совокупность всех естественных преобразований функтора F_a в функтор G_a. Для стрелки $k : a \to d$ пусть $G^f(k)$ — функция из $\mathrm{Nat}\,[F_a,\ G_a]$ в $\mathrm{Nat}\,[F_d,\ G_d]$, сопоставляющая естественному преобразованию $\tau : F_a \dot{\to} G_a$ преобразование $\tau' : F_d \dot{\to} G_d$, компонент τ'_f которого для произвольного $\mathscr{G}{\uparrow}d$-объекта f определяется равенством $\tau'_f = \tau_{f \circ k}$ (рис. 4.91).

Пример 4.51. Пусть F и G — функторы вида $2 \to \mathrm{Set}$, т. е. $\mathrm{Set}^{\to}$ — объекты $f : A \to B$ и $g : C \to D$. Так как $2{\uparrow}1$ — дискретная однообъектная категория, то F_1 можно отождествить с $F(1) = B$ и подобным же образом G_1 — с D. Поэтому

$$G^F(1) \simeq D^B,$$

где D^B — множество всех функций $B \to D$. Так как категория $2{\uparrow}0$ изоморфна самой категории 2, то можно считать, что F_0 и G_0 совпадают с F и G соответственно. Тогда

$$G^E(0) = \mathrm{Nat}\,[F,\ G],$$

т. е. $G^F(0)$ равно множеству E всех $\mathrm{Set}^{\to}$-стрелок из f в g. Наконец, G^f ставит в соответствие стрелке $1 : 0 \to 1$ функцию $g^f : E \to D^B$, определенную следующим образом. Для данного $\tau : F \dot{\to} G$, соответствующего $\mathrm{Set}^{\to}$-стрелке $<\tau_0,\ \tau_1>$ из f в g, естественное преобразование $G^F(\tau) : F_1 \dot{\to} G_1$ имеет единственный компонент, равный τ_1, так как категория $2{\uparrow}1$ имеет один объект 1_1 (рис. 4.92). Таким образом $g^f(<\tau_0,\ \tau_1>) = \tau_1$ и эта запутанная конструкция приводит к построению экспоненциала в $\mathrm{Set}^{\to}$.

Мы должны еще определить стрелку значения $ev : G^F \times F \dot{\to} G$ в категории $\mathrm{Set}^{\mathscr{G}}$. Каждый ее компонент $ev_a : G^F(a) \times F(a) \to G(a)$ вычисляется по формуле $ev_a(<\tau,\ x>) \simeq \tau_{1a}(x)$, где $x \in F(a)$ — произвольный элемент множества $F(a)$, а $\tau \in G^F(a)$ — произвольное преобразование $F_a \dot{\to} G_a$ [заметим, что компонент τ_{1a}, определяемый $\mathscr{G}{\uparrow}a$-объектом 1_a, является функцией из $F(a)$ в $G(a)$]. Далее, для каждой $\mathrm{Set}^{\mathscr{G}}$-стрелки $\tau : H \times F \to G$ компоненты экспоненциально присоединенного преобразования $\hat{\tau}_a : H \to G^F$ являются функциями вида

$$\hat{\tau}_a : H(a) \to G^F(a).$$

Для каждого $y \in H(a)$ $\hat{\tau}_a(y)$ обозначает естественное преобразование $F_a \to G_a$. Для каждого $\mathscr{G}{\uparrow}a$-объекта $f : a \to b$ компонент естественного преобразования $\hat{\tau}_a(y)$,

$$\begin{tikzcd}
0 \arrow[r, "!"] \arrow[dr, "1_1 \circ !"'] & 1 \arrow[d] \\
& 1
\end{tikzcd}$$

Рис. 4.92. Пояснение к построению дискретной однообъектной категории $2{\uparrow}1$

соответствующий f, является функцией из $F(a)$ в $G(b)$, переводящей $x\in F(b)$ в $\tau_b(<H(f)(y),\ x>)$ [заметим, что $\tau_b: H_b\times \times F_b\to G_b$ и $H(f): H(a)\to H(b)$]. Особенно будет нужно описание объектов-степеней в специальном топосе для моделей Крипке, в частности классификатор подобъектов в категориях Set-значных функторов [46].

4.8. СОПРЯЖЕННОСТЬ В ТЕОРИИ КАТЕГОРИЙ

Исходными данными *ситуации сопряжения,* или просто *сопряжения,* служит пара категорий $\mathcal{G}$ и D и функторов F и G между ними в обоих направлениях

$$\mathcal{G}\underset{G}{\overset{F}{\rightleftarrows}}D,$$

представляющих возможность взаимно заменять их объекты и стрелки. Для данных $\mathcal{G}$-объекта a и D-объекта b получаем объект $G(b)$ в $\mathcal{G}$ и объект $F(a)$ в D. Сопряженность возникает тогда, когда имеется точное соответствие между стрелками для этих объектов, направленных так, как показано на рис. 4.93, так что переход из a в $G(b)$ в категории $\mathcal{G}$ однозначно определяется переходом из $F(a)$ в b категории D. Иначе: требуется, чтобы для каждых объектов a и b категорий $\mathcal{G}$ и D соответственно существовала биекция

$$\theta_{ab}: D(F(a),\ b)\cong\mathcal{G}(a,\ G(b)) \tag{4.1}$$

между множеством D-стрелок вида $F(a)\to b$ и множеством $\mathcal{G}$-стрелок вида $a\to G(b)$. Более того, сопоставление биекции каждой паре a, b должно быть естественным по a и b. Это означает, что оно сохраняет категориальную структуру при варьировании как a, так и b. Функция, ставящая в соответствие паре $\langle a,\ b\rangle$ «hom-множество» $D(F(a),\ b)$, порождает функтор из произведения $\mathcal{G}^{op}\times D$ в Set, а функция, переводящая $\langle a, b\rangle$ в $\mathcal{G}\ (a, G(b))$, порождает другой такой же функтор. Требуется, чтобы отображения θ_{ab} являлись компонентами естественного преобразования θ между этими функторами, т. е. требуется, чтобы для любых объектов a', b' и стрелок $\alpha: a'\to a$, $\beta: b\to b'$ из категорий $\mathcal{G}$ и D соответственно и любой D-стрелки $f: F(a)\to b$ имело место равенство

$$G(\beta)\circ\theta_{ab}(b)\circ\alpha=\theta_{a'b'}(\beta\circ f\circ F(a)). \tag{4.2}$$

Когда такое θ существует, называем тройку $\langle F,\ G,\ \theta\rangle$ *сопряжением* от $\mathcal{G}$ к D. Функтор F называется *сопряженным слева* к G, что обозначается через $F\dashv G$, а функтор G — *сопряжением справа* к F (обозначается через $G\vdash F$). Отношение между F и G, задаваемое θ в (4.1), представляется схематически в виде

$$\frac{a\to G(b)}{F(a)\to b},$$

делающим наглядным различные «слева—справа».

Ситуация сопряжения выразима в терминах поведения специальных стрелок, ассоциированных с каждым объектом из $\mathcal{G}$ и из D. Пусть a — произвольный $\mathcal{G}$-объект, и положим в (4.1) $b=F(a)$. Применяя θ (т. е. соответствующий компонент) к единичной стрелке для $F(a)$, получаем $\mathcal{G}$-стрелку $h_a=\theta(1_{F(a)})$, называемую *единицей над a*. Тогда мы знаем, что для произвольного b из D всякой стрелке $g: a\to G(b)$ однозначно соответствует стрелка $f: F(a)\to b$, переходящая при отображении θ_{ab} в g. Пользуясь естественностью θ по a и b, обнаруживаем, что η_a обладает некоторым свойством коуниверсальности [46]. Именно для любой такой g существует, и притом только одна, стрелка $f: F(a)\to b$, для которой диаграмма на рис. 4.94 коммутативна. Действительно, $g=\theta_{ab}(f)$, и поэтому

$$\theta_{ab}(f)=G(f)\circ h_a. \tag{4.3}$$

В частности, это равенство получается из (4.2), если вместо a', b и b' подставить соответственно a, $F(a)$ и b, а вместо α, β и f — соответственно 1_a, f и $1_{F(a)}$. Из естественности θ следует также, что диаграмма на рис. 4.95 коммутативна для любой $\mathcal{G}$-стрелки k. С учетом равенства, выражающего θ через G и h,

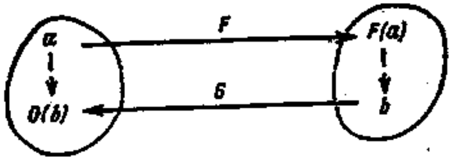


Рис. 4.93. Пояснение к понятию сопряженности

$$\begin{tikzcd}
a \arrow[r, "\eta_a"] \arrow[dr, "g"'] & G(F(a)) \arrow[d, dashed, "G(f)"] & F(a) \arrow[d, dashed, "f"'] \\
 & G(b) & b
\end{tikzcd}$$

Рис. 4.94. Пояснение к свойству коуниверсальности η_a

$$\begin{tikzcd}
a \arrow[r, "h_a"] \arrow[d, "k"'] & G(F(a)) \arrow[d, "G(F(k))"] \\
a' & G(F(a')) \arrow[l, "h_{a'}"']
\end{tikzcd}$$

Рис. 4.95. Пояснение к определению преобразования θ

и в силу естественности θ по a и $b=\eta_{a'}\circ k$ по определению η имеем

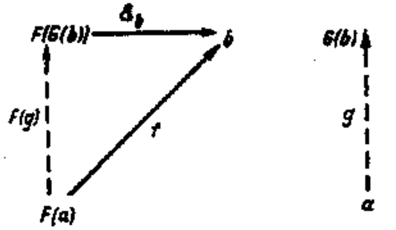
$$G(F(k))\circ\eta_a=\theta_{aF(a')}(F(k))=\theta_{aF(a')}(1_{F(a')}\circ 1_{F(a')}\circ F(k))=$$
$$=G(1_{Fa'})\circ\theta_{a'Fa'}(1F_{(a')})\circ k.$$

Поэтому стрелки η_a образуют компоненты некоторого естественного преобразования $\eta: 1_{\mathscr{G}}\to G\circ F$, называемого *единицей* сопряжения. Действенным образом для произвольного D-объекта b положим в (4.1) $a=G(b)$. Пусть τ обозначает естественный изоморфизм, обратный к $\theta(\tau_{ab}=\theta_{ab}{}^{-1})$. Тогда, применяя τ к единичной стрелке для $G(b)$, получаем коединицу $\varepsilon_b=\tau(1_{G(a)})$ над b. Стрелка ε_b обладает свойством универсальности, состоящим в том, что для произвольной D-стрелки $f: F(a)\to b$ существует только одна $\mathscr{G}$-стрелка $g: a\to\mathscr{G}(b)$, для которой диаграмма на рис. 4.96 коммутативна. Так как $f=\tau_{ab}(g)$, то

$$\tau_{ab}(g)=\varepsilon_b\circ F(g) \tag{4.4}$$

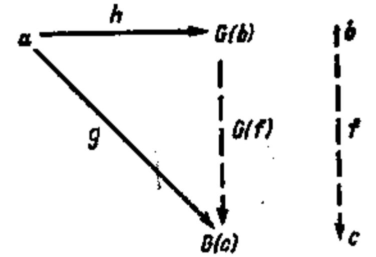
и стрелки $\mathscr{G}_b$ составляют компоненты некоторого естественного преобразования $\mathscr{G}: F\circ G\to 1_D$ — коединицы сопряжения. С другой стороны, для данных естественных преобразований η и ε указанного вида мы могли бы определить естественное преобразование θ и τ, задавая их компоненты формулами (4.3) и (4.4). Если выполняются свойства коуниверсальности и универсальности, представленные диаграммами на рис. 4.94 и 4.96 соответственно, то θ_{ab} и τ_{ab} будут взаимно обратными, т. е. биекциями, определяющими сопряжение θ от $\mathscr{G}$ к D. Таким образом, для указанных выше F и G следующие условия эквивалентны:

a) F является сопряженным слева к G, $F\dashv G$;
b) G является сопряженным справа к F, $G\vdash F$;
c) существует сопряжение $\langle F, G, Q\rangle$ от $\mathscr{G}$ к D;'



$$\begin{tikzcd}
F(G(b)) \arrow[r, "\varepsilon_b"] & b & & G(b) \\
F(a) \arrow[u, dashed, "F(g)"] \arrow[ur, "f"'] & & & a \arrow[u, dashed, "g"']
\end{tikzcd}$$

Рис. 4.96. Пояснение к универсальности стрелки ε_b

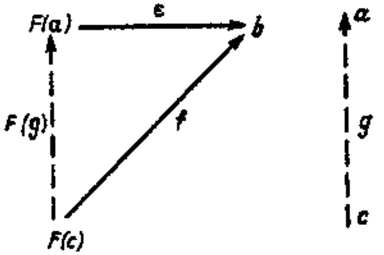


$$\begin{tikzcd}
a \arrow[r, "h"] \arrow[dr, "g"'] & G(b) \arrow[d, dashed, "G(f)"] & b \arrow[d, dashed, "f"] \\
& G(c) & c
\end{tikzcd}$$

Рис. 4.97. Пояснение к понятию пары, свободной над a относительно G

d) существуют естественные преобразования $\eta: 1_{\mathscr{G}}\to G\circ F$ и $\varepsilon: F\circ G\to 1_D$, компоненты которых обладают свойствами коуниверсальности и универсальности, представленными диаграммами на рис. 4.94 и 4.96.

Эти диаграммы относятся к частным случаям более общего факта. Предположим, что $G: D\to\mathscr{G}$ — произвольный функтор и a — некоторый объект из $\mathscr{G}$ Тогда пара $<b, \eta>$, состоящая из D-объекта b и $\mathscr{G}$-стрелки $\eta: a\to G(b)$, называется *свободной над a относительно* G, если для любой $\mathscr{G}$-стрелки вида $g: a\to G(c)$ существует единственная D-стрелка $f: b\to c$, такая, что диаграмма на рис. 4.97 коммутативна. Такая пара $<b, \eta>$ также называется *универсальной стрелкой от* a к G. Таким образом, если $F\dashv G$, то пара $<F(a), \eta_a>$ будет свободна над a относительно G. Двойственным образом для данных функтора $F:\mathscr{G}\to D$ и D-объекта b пара $<a, \varepsilon>$, состоящая из $\mathscr{G}$-объекта a и стрелки $f: F(c)\to b$, существует и притом только одна $g: c\to a$ из $\mathscr{G}$, такая, что диаграмма на рис. 4.98 коммутативна. Такая пара называется также *универсальной стрелкой от* F к b.



$$\begin{tikzcd}
F(a) \arrow[r, "\varepsilon"] & b & a \\
F(c) \arrow[u, dashed, "F(g)"] \arrow[ur, "f"'] & & c \arrow[u, dashed, "g"']
\end{tikzcd}$$

Рис. 4.98. Пояснение к понятию пары, косвободной над b относительно F

Из существования для данного функтора сопряженного к нему вытекают важнейшие следствия о свойствах этого функтора. Например, если $F\dashv G$, то G сохраняет пределы (т. е. отображает предел диаграммы в D в предел G-образа в $\mathscr{G}$, а F сохраняет копределы.

Примеры сопряжения

Начальные объекты. Пусть $\mathscr{G}=1$ — категория с одним объектом, скажем 0, и одним морфизмом 1_0, а G — единственный функтор $D\to 1$. Пусть функтор $F: 1\to D$ является сопряженным слева к G. Тогда

$$\frac{0\to G(b)}{F(0)\to b}.$$

Так как существует ровно одна стрелка $0\to G(b)$, то имеется ровно одна стрелка $F(0)\to b$. Поэтому $F(0)$ является *начальным объектом* в D. Коединицей ε_b служит единственная стрелка $F(0)\to b$.

Произведения. Пусть $\Delta:\mathscr{G}\to\mathscr{G}\times\mathscr{G}$ обозначает диагональный функтор, ставящий в соответствие объекту a пару $\langle a, a\rangle$, а стрелке $f: a\to b$ — пару стрелок $\langle f, f\rangle: \langle a, a\rangle\to\langle b, b\rangle$. Предположим, что Δ имеет правый сопряженный $G:\mathscr{G}\times\mathscr{G}\to\mathscr{G}$. Тогда

$$\begin{tikzcd}
\Delta(G(x)) \arrow[r, "\varepsilon_x"] & x & G(x) \\
\Delta(c) \arrow[u, dash, dashed, "\Delta(h)"] \arrow[ur, "\langle f,g\rangle"'] & & c \arrow[u, dashed, "h"]
\end{tikzcd}$$

а)

$$\begin{tikzcd}
a & G(x) \arrow[l, "p"'] \arrow[r, "g"] & b \\
& c \arrow[ul, "f"] \arrow[u, dashed, "h"] \arrow[ur, "g"'] &
\end{tikzcd}$$

б)

Рис. 4.99. Пояснение к понятию произведения

$$\frac{c\to G(x)}{\langle c,\ c\rangle\to x},$$

где c из $\mathcal{G}$, а $x=\langle a,\ b\rangle$ из $\mathcal{G}\times\mathcal{G}$. Коединица $\varepsilon_x:\Delta(G(x)\to\langle a,\ b\rangle$ есть пара $\mathcal{G}$-стрелок $p:G(x)\to a$ и $q:G(x)\to b$. По свойству косвободности стрелки ε_x для произвольных стрелок $f:c\to a$, $g:c\to b$ существует единственная стрелка $h:c\to G(x)$, такая, что диаграмма на рис. 4.99,a коммутативна, следовательно, коммутативна и диаграмма на рис. 4.99,$б$. Таким образом, $G(x)$ является произведением объектов a и b с парой ε_x ассоциированных с ним проекций. Имеем сопряжение

$$\frac{c\to a\times b}{c\to a,\ c\to b},$$

единицей $\eta_c:c\to c\times c$ которого служит произведение $\langle 1_c,\ 1_c\rangle$. Можно показать, что предел (копредел) диаграммы какого-либо типа в категории $\mathcal{G}$ возникает вследствие существования сопряженного справа и слева к «диагональному» функтору $\mathcal{G}\to\mathcal{G}^J$, где J — каноническая категория, имеющая тот же вид, что и данная диаграмма (для произведения J будет дискретной категорией $\{0,1\}$). Единицей в случае существования сопряженного слева функтора служит универсальный косинус, а коединицей в случае существования сопряженного справа — универсальный конус.

Топология и алгебра. Имеется много важных конструкций, выступающих как сопряженные функторы к пренебрегающим функторам. Функтор, сопряженный слева к пренебрегающему функтору $v:G_{\text{гр}}\to\mathrm{Set}$ из категории групп в категорию множеств, ставит в соответствие каждому множеству свободную группу, порожденную этим множеством («свобода» здесь понимается так же, как и выше при рассмотрении единиц сопряжения). Построение поля частных области целостности дает функтор, сопряженный слева к пренебрегающему функтору из категории полей в категорию областей целостности. Задание дискретной топологии на множестве дает функтор, сопряженный слева к $v:\mathrm{Top}\to\mathrm{Set}$, а введение антидискретной (тривиальной) дает сопряженный справа функтор к v. Пополнение метрического пространства определяет функтор, сопряженный слева к пренебрегающему функтору из полных метрических пространств в метрические пространства.

$$\begin{tikzcd}
b^a\times a \arrow[r, "ev"] \arrow[d, dashed, "f^a\times 1a"'] & b \\
c^a\times a \arrow[ur, "b\circ ev'"'] &
\end{tikzcd}$$

Рис. 4.100. Пояснение к понятиям экспоненцирования и сопряжения

Экспоненцирование. Если $\mathcal{G}$ допускает экспоненцирование, то существует биекция

$$\mathcal{G}(c\times a,\quad b)\cong\mathcal{G}(c,\quad b^a)$$

для любых объектов a, b, c, указывающая на наличие сопряжения. Пусть $F:\mathcal{G}\to\mathcal{G}$ — функтор умножения справа — $\times a$, переводящий произвольное c в $c\times a$. Тогда F имеет сопряженный справа функтор $(\quad)^a:\mathcal{G}\to\mathcal{G}$; переводящий b в b^a, а произвольную стрелку $f:c\to b$ или $f^a:c^a\to b^a$, являющуюся экспоненциально присоединенной к композиции $f\circ ev':c^a\times a\to c\to b$ — в единственную стрелку вида $c^a\to b^a$, для которой диаграмма на рис. 4.100 коммутативна. Коединицей $\varepsilon_b:F(b^a)\to b$ является стрелка значения $ev:b^a\times a\to b$, ее свойство косвободы дает аксиому экспоненцирования из ( ). Ситуация сопряжения имеет вид

$$\frac{c\to b^a}{c\times a\to b}.$$

Таким образом, $\mathcal{G}$ *допускает экспоненцирование тогда и только тогда*, когда функтор $\times a$ имеет сопряженный справа для каждого $\mathcal{G}$-объекта функтор a.

Относительные псевдоподобия. Относительные дополнения являются частным случаем экспоненцирования. В произвольной решетке с относительными псевдодополнениями условие $c\sqcap a\leqslant b$ выполняется тогда и только тогда, когда $c\leqslant a\Rightarrow b$ дает сопряжение

$$\frac{c\to a\Rightarrow b}{c\sqcap a\to b}.$$

Решетка имеет относительные псевдодополнения тогда и только тогда, когда функтор $\sqcap a$, *переводящий* c *в* $c\sqcap a$, *обладает сопряженным справа для каждого объекта функтором* a[1].

Сопряженности в ЧУ множествах. Пусть $(P,\ \leqslant)$ и $(Q,\ \leqslant)$ — ЧУ множества. Функтор из P в Q — это монотонная функция $f:P\to Q$, т. е. если $p\leqslant q$, то $f(p)\leqslant f(q)$. Тогда $g:Q\to P$

---

[1] ЧУМ, в котором для любых двух элементов существует наименьшая верхняя грань (НВГ) и наибольшая низшая грань (ННГ), называется решеткой и полным ЧУМом (см. гл. 2).

будет сопряженным справа к f, $\frac{p \to g(r)}{f(p) \to r}$, если и только если для всех $p \in P$ и $r \in Q$ $p \leqslant g(r)$ тогда и только тогда, когда $f(p) \leqslant r$. С другой стороны, g будет сопряженным слева, к f,

$$\frac{r \to f(p)}{g(r) \to p},$$

если $g(r) \leqslant p$ тогда и только тогда, когда $r \leqslant f(p)$. Рассмотрим пример. Пусть $f: A \to B$ — произвольная функция. Тогда для множеств $x \leqslant A$, $Y \subseteq B$ имеем $X \subseteq f^{-1}(Y)$ тогда и только тогда, когда $f(Y) \subseteq Y$. Поэтому функтор $f^{-1}: \mathscr{P}(B) \to \mathscr{P}(A)$, переводящий $Y \subseteq B$ в $f^{-1}(Y)$, является сопряженным справа к функтору $\mathscr{P}(f): \mathscr{P}(A) \to \mathscr{P}(B)$, переводящему $X \subseteq A$ в его f-образ $f(X) \subseteq B$. Функтор f^{-1} имеет также сопряженный справа $f^{+}: \mathscr{P}(A) \to \mathscr{P}(B)$, задаваемый равенством $f^{+}(X) = \{y \in B : f^{-1}\{y\} \subseteq \subseteq X\}$, где $f^{-1}\{y\} = \{x : f(x) = y\}$ — прообраз $\{y\}$. Соотношение $f^{-1} \dashv f^{+}$ следует из того, что $f^{-1}(Y) \subseteq X$ тогда и только тогда, когда $Y \subseteq f^{+}(X)$.

Классификатор подобъектов. Запись

$$\frac{d \to \Omega}{?\rightarrowtail d},$$

где $? \rightarrowtail d$, обозначает произвольный подобъект в d и указывает на то, что Ω-аксиома выражает свойство, связанное с сопряженностью.

Функтор $\mathrm{Sub}: \mathscr{G} \to \mathbf{Set}$, описанный ранее, ставит в соответствие каждому объекту d совокупность его подобъектов, а каждой стрелке $f: c \to d$ — функцию $\mathrm{Sub}(f): \mathrm{Sub}(d) \to \mathrm{Sub}(c)$, переводящую каждый подобъект в d в его обратный образ относительно f. Так, определенный функтор является контравариантным. Переходом к двойственной категории выражаем Sub как ковариантный функтор

$$\mathrm{Sub}: \mathscr{G}^{op} \to \mathrm{Set}.$$

$$\begin{tikzcd}
a \arrow[r, tail, "g_0"] \arrow[d] & d \arrow[d, "\chi_{g_0}"] \\
1 \arrow[r, "true"] & \Omega
\end{tikzcd}$$

а)

$$\begin{tikzcd}
1 \arrow[r, "h"] \arrow[dr, "g"'] & \mathrm{Sub}(f) \arrow[d, dashed, "\mathrm{Sub}(f)"] & \Omega \arrow[d, dashed, "f"'] \\
& \mathrm{Sub}(d) & d
\end{tikzcd}$$

б)

Рис. 4.101. Пояснение к классификатору подобъектов с использованием сопряжения

Теперь в случае, когда $\mathscr{G} = \mathscr{G}$ (некоторый топос), стрелка $t_{\text{rue}}: 1 \to \Omega$ определяет подобъект в Ω и соответствует поэтому некоторой функции $\eta: 1 = \{0\} \to \mathrm{Sub}(\Omega)$. Рассмотрим диаграмму на рис. 4.101. Показанная на ней функция g выбирает подобъект $g_0: a \rightarrowtail d$ в d, для которого имеем характеристическую стрелку χ_{g_0}, определяемую декартовым квадратом (рис. 4.101, a). Тогда $f = (\chi_{g_0})^{op}$ является $\mathscr{G}^{op}$-стрелкой из Ω в d, и функция $\mathrm{Sub}(b)$ [равная исходной $\mathrm{Sub}(\chi_{g_0})$] переводит true в ее обратный образ относительно χ_{g_0}, т. е. в подобъект g_0. Поэтому рассмотренная выше треугольная диаграмма на рис. 4.101, б коммутативна. Но в силу однозначности характеристической стрелки для g_0 стрелка λ_{g_0} является единственной, подъемом вдоль которой стрелки true получается g_0. Поэтому единственной $\mathscr{G}^{op}$-стрелкой, для которой эта треугольная диаграмма коммутативна, является стрелка $f = (\lambda_{g_0})$. Таким образом, пара $\langle \Omega, h \rangle$, т. е. $\langle \Omega, \mathrm{true}: 1 \to \Omega \rangle$, будет свободной над 1 относительно Sub. Наоборот, из того, что пара $\langle \Omega, \eta \rangle$ свободна, вытекает, что $\eta(0)$ классифицирует подобъекты, и можно сказать, что категория $\mathscr{G}$, допускающая обратные образы, имеет классификатор подобъектов тогда и только тогда, когда существует универсальная стрелка из 1 в $\mathrm{Sub}: \mathscr{G}^{op} \to \mathrm{Set}$. Заметим, что Ω-аксиома утверждает, что

$$\mathrm{Sub}(d) \cong \mathscr{G}(d, \Omega) \cong \mathscr{G}^{op}(\Omega, d)$$

и аналогично

$$\mathrm{Rel}(b, a) \cong \mathscr{G}(b, \Omega^a) \cong \mathscr{G}^{op}(\Omega^a, b),$$

так что ковариантные варианты функторов Sub и $\mathrm{Rel}(-, a)$, естественно изоморфны hom-функторам вида $(\mathscr{G}(c, -))$. Вообще Set-значный функтор, изоморфный hom-функтору, называется *представимым*. Представимые функторы всегда характеризуются своими объектами, свободными над 1 в Set.

Теорема 4.4. Пусть $\mathscr{G}$ — категория, допускающая обратные образы, и $f: a \to b$ — произвольная $\mathscr{G}$-стрелка. Тогда f индуцирует *функтор обратного образа* $f^*: \mathscr{G}\downarrow b \to \mathscr{G}\downarrow a$, обобщающий функтор $f^{-1}: \mathscr{P}(B) \to \mathscr{P}(A)$ из предыдущего раздела. Функтор f^* вводится в рассмотрение в соответствии с диаграммой на рис. 4.102, где k есть $\mathscr{G}\downarrow b$-стрелка

$$\begin{tikzcd}
c \arrow[rr] \arrow[dr, dashed] \arrow[dd, "f^*(g)"'] & & d \arrow[dr, "k"] \arrow[dd] & \\
& m \arrow[rr, "g"'] \arrow[dl, "f^*(h)"] & & h \arrow[dl, "h"] \\
a \arrow[rr, "f"'] & & b &
\end{tikzcd}$$

Рис. 4.102. Пояснение к понятию функтора обратного образа

$$\begin{tikzcd}
c \arrow[rr, "k"] \arrow[dr, "g"'] & & d \arrow[dl, "h"] \\
& a &
\end{tikzcd}$$

а)

$$\begin{tikzcd}
c \arrow[rr, "k"] \arrow[dr, "f\circ g"'] & & d \arrow[dl, "f\circ h"] \\
& b &
\end{tikzcd}$$

б)

Рис. 4.103. Пояснение к функтору «композиция с f»

из g в h; $f^*(g)$ и $f^*(h)$ — обратные образы стрелок g и h относительно f, однозначно определяющие показанную штриховой линией стрелку $c \to m$, которую обозначим через $f^*(k): f^*(g) \to \to f^*(h)$.

Функтор «композиция с f»

$$\Sigma_f : \mathcal{G}{\downarrow}a \to \mathcal{G}{\downarrow}b$$

сопоставляет $\mathcal{G}{\downarrow}a$-объекту $g: c \to a \mathcal{G}{\downarrow}b$ объект $f \circ g : c \to b$, $\mathcal{G}{\downarrow}b$-объекту (рис. 4.103, a) $\mathcal{G}{\downarrow}b$-стрелку (рис. 4.110, $б$). Произвольная стрелка диаграммы на рис. 4.104, a из $\Sigma f(g)$ в $t: b \to d$ категории $\mathcal{G}{\downarrow}b$ соответствует единственной $\mathcal{G}{\downarrow}a$-стрелке k' диаграммы на рис. 4.104, $б$ из g в $f^*(t)$, определяемой свойством универсальности обратного образа $f^*(t)$, и мы имеем поэтому сопряжение

$$\frac{g \to f^*(t)}{\Sigma_f(g) \to t},$$

показывающее, что $\Sigma_f \dashv f^*$.

Для теоретико-множественных функций функтор f^* имеет также функтор, сопряженный справа

$$\Pi_f : \mathrm{Set}{\downarrow}A \to \mathrm{Set}{\downarrow}B.$$

Для данного $g: X \to A$ $\mathrm{Set}{\downarrow}B$ объект $\Pi_f(g)$ является стрелкой вида $k: Z \to B$, которую рассматриваем как расслоение над B.

$$\begin{tikzcd}
c \arrow[rr, "k"] \arrow[dr, "f\circ g"'] & & d \arrow[dl, "t"] \\
& b &
\end{tikzcd}$$

а)

$$\begin{tikzcd}
c \arrow[drr, "k"] \arrow[dr, dashed, "k'"'] \arrow[ddr, "g"'] & & \\
& p \arrow[r] \arrow[d, "f^*(t)"] & d \arrow[d, dash, "t"] \\
& a \arrow[r, "f"] & b
\end{tikzcd}$$

б)

Рис. 4.104. Функтор «композиция с f» и сопряжение

$$\begin{tikzcd}
f^{-1}(b) \arrow[rr, "h"] \arrow[dr, hook] & & X \arrow[dl, "g"] \\
& A &
\end{tikzcd}$$

Рис. 4.105. Пояснение к понятию функции z как множества пар (b, h)

$$\begin{tikzcd}
P \arrow[rr] \arrow[dr, dashed, "t'"] \arrow[dd, "f^*(h)"'] & & Y \arrow[dr, "t"] \arrow[dd, "h"'] & \\
& X \arrow[dl, "g"] & & Z \arrow[dl, "\Pi_f(g)"] \\
A \arrow[rr, "f"] & & B &
\end{tikzcd}$$

Рис. 4.106. Пояснение к понятию $\mathrm{Set}{\downarrow}B$-стрелки

Понимая подобным же образом и g, можем сказать, что слой в Z над $b \in B$, т. е. множество $k^{-1}\{b\}$, определяется как множество всех локальных сечений расслоения g, определенных на $f^{-1}\{b\} \subseteq A$. Формально Z есть множество всех пар (b, h), таких, что h является функцией, определенной на $f^{-1}\{b\}$, для которой диаграмма на рис. 4.105 коммутативна. Функция k является проекцией на B. Если g — включение X в A, то единственным возможным сечением h будет включение $f^{-1}\{b\}$ в X при условии, что $f^{-1}\{b\} \subseteq X$, и имеется один элемент в противном случае. Таким образом, k можно отождествить с включением множества

$$\{b : f^{-1}\{b\} \subseteq X\} = f^+(X)$$

в B, и функтор f^+ оказывается частным случаем функтора Π_f. Для данных стрелок $g: X \to A$ и $h: Y \to B$ рассмотрим диаграмму на рис. 4.106, где t является $\mathrm{Set}{\downarrow}\mathrm{B}$-стрелкой из h в $\Pi_f(g)$. Стрелка $f^*(h)$ — обратный образ h относительно f — является проекцией в A множества

$$P = \{\langle a, y\rangle : f(a) = h(y)\},$$

так что если $\langle a, y\rangle \in P$, то y лежит в слое над $f(a)$ в B и поэтому $t(y)$ лежит в слое над $f(a)$ расслоения $\Pi_f(g)$. Таким образом, $t(y)$ будет сечением S расслоения g над множеством

$$\begin{tikzcd}
a \arrow[r, tail, "\langle f, 1_a\rangle"] \arrow[d, "1_a"'] & b \times a \arrow[d, "k"] \\
a \arrow[r, tail, "\eta_a"] & \tilde{a}
\end{tikzcd}$$

Рис. 4.107. Пояснение к основной теореме теории топосов

$$\begin{tikzcd}
\Pi_f(c) \arrow[r] \arrow[d, "\Pi_f(g)"'] & \tilde{c}^a \arrow[d, "\tilde{g}^a"] \\
b \arrow[r, "h"] & \tilde{a}^a
\end{tikzcd}$$

Рис. 4.108. Пояснение к определению $\Pi_f(g)$

$$\begin{tikzcd}
c \arrow[r, tail, "\eta_c"] \arrow[d, "g"'] & \tilde{c} \arrow[d, "\tilde{g}"] \\
a \arrow[r, tail, "\eta_a"] & \tilde{a}
\end{tikzcd}$$

Рис. 4.109. Пояснение к определению стрелки $\tilde{g}^a$

$$\begin{tikzcd}
f^{-1}(b) \arrow[rr, "t"] \arrow[dr, "f"'] & & Y \arrow[dl, "h"] \\
& B &
\end{tikzcd}$$

Рис. 4.110. Пояснение факта допущения экспоненцирования $\varepsilon\downarrow b$

$f^{-1}\{f(a)\}$, содержащим a. Положим $t'(\langle a, y\rangle)=S(a)$. Тогда t' является стрелкой из $f^*(h)$ в g, принадлежащей категории $\mathrm{Set}\downarrow A$. Этим установлено соответствие

$$\frac{h \xrightarrow{t} \Pi_f(g)}{f^*(h) \xrightarrow{t'} g},$$

приводящее к сопряженности $f^* \dashv \Pi_f$. Полная формулировка *основной теоремы топосов* следующая:

Для любого топоса $\mathcal{G}$ и любого $\mathcal{G}$-объекта b относительная категория $\mathcal{G}\downarrow b$ является топосом, и для произвольной стрелки $f:a\to b$ функтор обратного образа $f^*:\mathcal{G}\downarrow b\to\mathcal{G}\downarrow a$ имеет как сопряженный слева Σf, так и сопряженный справа Π_f. Существование функтора, сопряженного слева, требует только обратные образы. Построение функтора Π_f основано на всех аксиомах топоса и использует классификатор частичных стрелок (локальные сечения являются частичными стрелками). Для данной ~~стрелки $f:a\to b$ пусть k обозначает единственную~~ стрелку $b\times\times a\to\tilde{a}$, для которой квадрат на рис. 4.107 — декартов, где η_a обозначает теперь классификатор частичных стрелок.

Пусть $h:b\to\tilde{a}^a$ — стрелка, экспоненциально присоединенная к k. В Set h переводит $b\in B$ в стрелку, соответствующую частичной функции $f^{-1}\{b\}$, включаемой в A (из A в A). Тогда для произвольной стрелки $g:c\to a$ определим $\Pi_f(g)$ как обратный образ стрелки $\tilde{g}^a$ относительно h с помощью диаграммы на рис. 4.108, где $\tilde{g}$ — единственная стрелка, определяемая декартовым квадратом (рис. 4.109), а $\tilde{g}^a$ — результат применения функтора $(\quad)^a:\mathcal{G}\to\mathcal{G}$ к стрелке $\tilde{g}$. Можно показать, что это определение отражает определение Π_f в Set. Функтор Π_f используется также для проверки того, что $\mathcal{G}\downarrow b$ допускает экспоненцирование. Проиллюстрируем это опять на категории Set. Для данных $\mathrm{Set}\downarrow B$-объектов $f:A\to B$ и $h:Y\to B$ их экспоненциал h^f имеет вид $h^f:E\to B$. В соответствии с теорией слоев [46] слой в E над b состоит из всех пар $\langle b, t\rangle$, таких, что функция $t:f^{-1}\{b\}\to Y$ удовлетворяет условию коммутативности диаграммы (рис.4.110). Если теперь построить обратный образ $f^*(h)$ (рис. 4.111) и определить $t':f^{-1}(b)\to P$ равенством $t'(a)=$ $=\langle a, t(a)\rangle$; где $\langle b, t\rangle\in E$, то, используя определение P, данное выше, можно заметить, что t' является сечением расслоения $f^*(h)$ над $f^{-1}\{b\}$, т. е. элементом из слоя над b расслоения $\Pi_f(f^*(h))$. Кроме того, $t=g\circ t'$, что приводит к изоморфизму между h^f и $\Pi_f(f^*(h))$ в Set. В категории $\mathcal{G}\downarrow b$ для данных $f:a\to b$ и $h:c\to b$ $\mathcal{G}\downarrow b$-объект $\Pi_f f^*(h))$ служит экспоненциалом h^f. Это можно пояснить языком сопряженностей, так как функтор произведения $-\times f:\mathcal{G}\downarrow\beta-\mathcal{G}\downarrow b$ является композицией вида

$$\begin{tikzcd}
& P \arrow[r, "g"] \arrow[d, "f^*(h)"] & Y \arrow[d, dash, "h"] \\
f^{-1}(b) \arrow[ur, "t'"] \arrow[r, hook] & A \arrow[r, "f"] & B
\end{tikzcd}$$

Рис. 4.111. Пояснение к тому, что t' является сечением

$$\begin{tikzcd}
p \arrow[r, "f^*(h)"] & a \arrow[d, "f"] \\
c \arrow[u] \arrow[r, "h"] & b
\end{tikzcd}$$

Рис. 4.112. Пояснение к тому, что произведение $h\times f$ является расслоенным произведением

$$\mathcal{G}\downarrow b \xrightarrow{f^*} B\downarrow a \xrightarrow{\Sigma f} \mathcal{G}\downarrow b.$$

Это объясняется тем, что произведение $h\times f$ в $\mathcal{G}\downarrow b$ объектов h и f есть расслоенное произведение (рис. 4.112) $f\circ f^*(h)=$ $=\Sigma_f(f^*(h))$ в $\mathcal{G}$. Но каждый из функторов f^* и Σ_f имеет сопряженный справа Π_f и f^* соответственно, а их композиция $\Pi_f\circ f^*$ дает сопряженный справа к $-\times f$. Детальное доказательство теоремы можно найти в [38].

4.9. ПРИМЕНЕНИЕ ТЕОРИИ КАТЕГОРИЙ ДЛЯ ОПИСАНИЯ МОДЕЛЕЙ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

А. Категориальная модель знаний

Покажем, как методы теории категорий применяются для описания модели представления знаний в виде фреймов [30, 52]. В частности, дадим категориальные модели представления образца и процесса сопоставления по образцу. Подробное рассмотрение фрейм-моделей представления знаний, механизмов

Рис. 4.113. Пояснение к работе механизма сопоставления

сопоставления по образцу в различных системах было дано в т. 2 «Основ кибернетики» первого издания [30]. Здесь эти понятия даны в общем описательном плане.

Применим теорию категорий к таким понятиям (рис. 4.113), как фрейм-ситуация (на входе системы искусственного интеллекта), фрейм-образец, хранимый в БЗ системы, продукция как модель знаний, механизм сопоставления по образцу.

Будем считать, что продукция состоит из двух частей — *условий применимости* и *указания действий,* которые надо выполнить, если условия соблюдаются. Для проверки выполнимости условий используются модель, называемая образцом, и механизм сопоставления по образцу, который включает в себя означивание переменных, заполнение слотов фрейма и т. д. [52]. Образцы хранятся в памяти машины в БЗ.

Пусть задана произвольная подкатегория C категории множеств. Выделим в ней объект S, который будем называть *базовым.* Образцом назовем пару (X, φ), где X — некоторый объект категории C, а $\varphi \in \mathrm{Mor}\,(X, S)$ — морфизм из X в S (рис. 4.113). Выделим также объект I категории, который будем называть *исходным. Ситуацией* назовем образец (I, α), где α — морфизм из I в S (рис. 4.114).

Будем считать, что ситуация (I, α) сопоставима с образцом (X, φ), если существует хотя бы один морфизм $\psi \in \mathrm{Mor}\,(I, X)$, такой, что диаграмма на рис. 4.115, *а* коммутативна. Следует отметить, что морфизм α определяется неоднозначно и это хорошо с точки зрения теории искусственного интеллекта.

Если даны два образца (X, φ) и (Y, ψ), то будем считать, что первый является частным случаем второго, если существует морфизм $\chi \in \mathrm{Mor}\,(X, Y)$, делающий диаграмму на рис. 4.115, *б* коммутативной. Очевидно, что если один образец является частным случаем другого, то всякая ситуация (входной сигнал), сопоставимая с образцом (из БЗ системы), сопоставима и со вторым, однако обратное утверждение несправедливо. Можно показать [52], что отношение «быть частным случаем» задает на множестве образцов квазипорядок, который необязательно будет порядком. Считается, что два образца эквивалентны, если каждый из них является частным случаем другого. С эквивалентными образцами сопоставимы одни и те же ситуации (рис. 4.116). Ситуация может записываться в виде $\alpha : I \to S$, и можно говорить о ситуации со значением S, при этом можно говорить о ситуации α.

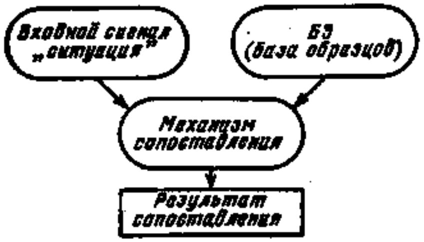


Рис. 4.114. Пояснение к категорной модели знаний

$$\begin{tikzcd} & S & \\ I \arrow[ur, "\alpha"] \arrow[rr, "\psi"] & & X \arrow[ul, "\varphi"'] \end{tikzcd}$$

а)

$$\begin{tikzcd} & S & \\ X \arrow[ur, "\varphi"] \arrow[rr, "\chi"] & & Y \arrow[ul, "\psi"'] \end{tikzcd}$$

б)

Рис. 4.115. Пояснение к представлению знаний с помощью теории категорий

Пусть заданы два образца: (X, φ) и (Y, φ). Образец (Z, χ) вместе с морфизмами $\lambda : Z \to X$ и $\mu : Z \to Y$ назовем *наибольшим частным случаем* (НЧС) пары образцов (X, φ) и (Y, ψ), если диаграмма на рис. 4.116 коммутативна; если (Z', χ') — другой образец, для которого диаграмма на рис. 4.117, *а* коммутативна, то найдется морфизм $\nu : Z' \to Z$, такой, что будет коммутативной и диаграмма на рис. 4. 117, *б*. Нетрудно доказать, что *ситуация сопоставима с каждым из образцов* (X, φ) и (Y, ψ) *в том и только в том случае, когда она сопоставима с их НЧС* (Z, χ), *если он существует* [52].

Это понятие НЧС отличается от такого в теории категорий тем, что существует понятие единственности морфизма χ. Поэтому нельзя утверждать, что НЧС единствен, хотя бы даже с точностью до изоморфизма. Можно доказать, что все НЧС пары образцов являются эквивалентными. Еще раз заметим, что пара образцов может не иметь НЧС.

Можно ввести наряду с НЧС *наименьшее обобщение пары образцов.* Пусть заданы два образца: (X, φ) и (Y, φ). Образец (Z, χ) с морфизмами $\lambda : X \to Z$ и $\mu : Y \to Z$ назовем наименьшим обобщением, если диаграмма на рис.

$$\begin{tikzcd} & S & \\ X \arrow[ur, "\varphi"] & & Y \arrow[ul, "\psi"'] \\ & Z \arrow[ul, "\lambda"] \arrow[ur, "\mu"'] \arrow[uu, "\chi"'] & \end{tikzcd}$$

Рис. 4.116. Пояснение к понятию эквивалентности

$$\begin{tikzcd} & S & \\ X \arrow[ur, "\varphi"] & & Y \arrow[ul, "\psi"'] \\ & Z' \arrow[ul, "\lambda'"] \arrow[ur, "\mu'"'] \arrow[uu, "\chi'"] & \end{tikzcd}$$

а)

$$\begin{tikzcd} & S & \\ Z' \arrow[ur, "\chi'"] \arrow[rr, "\nu"'] & & Z \arrow[ul, "\chi"'] \end{tikzcd}$$

Рис. 4.117. Пояснение к понятию наибольшего частного случая (НЧС)

4.118, *а* коммутативна; если (Z', χ') — другой образец для которого диаграмма на рис. 4.118, *б* коммутативна, то найдется морфизм $\nu: Z \to Z'$, такой, что диаграмма на рис. 4.118, *в* тоже будет коммутативной. Здесь снова не требуется единственности морфизма ν. Пара образцов может не иметь наименьшего обобщения, если она имеет несколько наименьших обобщений, то все они эквивалентны. Аналогично предыдущему, если ситуация сопоставима с одним из образцов (X, φ) и $(Y\ \psi)$, то она сопоставима и с их наименьшим обобщением (Z, χ), если последнее существует (что можно доказать [52]). Но обратное утверждение неверно: ситуация, совместимая с наименьшим обобщением двух образцов, необязательно должна быть сопоставимой с одним из них.

Теория категорий может быть использована для описания баз знаний, основанных на продукциях. Для этого выделим в категории C два базовых объекта S и R и будем говорить об образцах и ситуациях со значениями в S и R. Продукцией из S в R назовем тройку (X, φ, ψ), где X — объект категории; $\varphi \in \mathrm{Mor}\,(X, S)$; $\psi \in \mathrm{Mor}\,(X, R)$. Продукцию поэтому можно рассматривать как пару образцов (X, φ) и (X, ψ). Пусть (I, α), $\alpha \in \mathrm{Mor}\,(I, S)$ — произвольная ситуация со значением в S. Будем считать, что эта продукция (X, φ, ψ) применима к ситуации (I, α), если эта ситуация сопоставима с образцом (X, φ), т. е. если существует морфизм $\beta: I \to X$, такой, что $\alpha = \beta\varphi$. В этом случае будем называть ситуацию $\beta\psi: I \to X$ результатом применения продукции (X, φ, ψ) к ситуации α. Таким образом, результатом применения продукции из S и R к ситуации со значением в S будет ситуация со значением в R. Наиболее интересен случай $R = S$, однако часто бывает удобно различать S и R, полагая ситуацию со значениями в S условиями задачи, а ситуацию со значениями в R решением задачи.

$$\begin{tikzcd} & S & \\ X \arrow[ur, "\varphi"] \arrow[r, "\lambda"'] & Z \arrow[u] & Y \arrow[ul, "\psi"'] \arrow[l, "\mu"] \end{tikzcd}$$

а)

$$\begin{tikzcd} & S & \\ X \arrow[ur, "\varphi"] \arrow[r, "\lambda'"'] & Z' & Y \arrow[ul, "\psi"'] \arrow[l, "\mu'"] \end{tikzcd}$$

б)

$$\begin{tikzcd} & S & \\ Z \arrow[ur, "\chi"] \arrow[rr, "\nu"'] & & Z' \arrow[ul, "\chi"'] \end{tikzcd}$$

в)

Рис. 4.118. Пояснение к понятию наименьшего обобщения пары образцов

Заметим, что морфизм β может определяться условием $\alpha = \beta\varphi$ неоднозначно. Если $\beta': I \to M$ — другой морфизм, для которого выполнено равенство $\alpha = \beta'\varphi$, то возможно, что $\beta'\psi \neq \beta\psi$. Это означает, что результат применения продукции к ситуации не определяется однозначно.

Б. Категориальная модель процесса классификационных построений

В этом разделе под категориальной моделью процесса классификационных построений будем понимать некоторую структуру, предназначенную для представления знаний о задачах исследования, построения и использования классификаций, для описания которой используется аппарат теории категорий. Выделяются три основных компонента процесса классификационных построений: автоматическая классификация, диагностическая классификация и построение систем информативных признаков. Каждому компоненту соответствует своя собственная модель, представляющая совокупность отношений между частями данного компонента. В то же время эти три модели для каждого компонента являются подмоделями общей модели всего процесса классификационных построений в целом. Нами предлагается дополнить общую модель процесса подмоделью, описывающей способы хранения множества классифицируемых объектов в БД и структурные преобразования этого множества, необходимые для проведения классификационных построений.

Введение такой модели обусловлено, на наш взгляд, двумя основными причинами. Во-первых, существующие в исходных данных функциональные зависимости требуется отражать в модели хранения данных, чтобы избежать проблем, связанных с избыточностью и целостностью БД. Следовательно, для хранения исходного материала могут использоваться сетевые, иерархические и реляционные модели данных. Для большинства методов классификационных построений исходный материал должен быть представлен в форме матрицы «объект свойства», поэтому требуются преобразования из структур, в которых исходный материал хранится в БД, в матрицу «объект — свойства». Во-вторых, при описании объектов качественными признаками необходимы преобразования для кодирования и декодирования значений признаков. Кодирование обеспечивает компактное хранение информации в БД и возможность применения математических методов, которые непосредственно с текстовыми значениями работать не могут. Декодирование используется при визуализации содержимого БД для обеспечения информационных потребностей пользователей.

Каждую подмодель процесса квалификационных построений выразим как категорию и затем из этих категорий построим описание всей модели исследования, построения и использования классификаций. При создании модели ограничимся рассмотрением признаков, измеренных в номинальной шкале. Как показывается дальше, такое ограничение вполне оправдано.

Введем аксиоматическое определение категории [52], согласно которому построим категории для каждой подмодели процесса классификационных построений.

Категория включает в себя совокупности: 1) предметов, явлений, которые называются σ-объектами; 2) предметов, явлений, которые называются σ-стрелками (морфизмами); 3) операций, которые ставят каждой σ-стрелке d пару σ-объектов: $\operatorname{dom} d$ — начало стрелки и $\operatorname{cod} d$ — конец стрелки. Пусть $x = \operatorname{dom} d$, $y = \operatorname{cod} d$, тогда σ-стрелку d будем обозначать $d : x \to y$ или $x \to y$; операцию композиции любых двух σ-стрелок d и e, таких, что $\operatorname{dom} e = \operatorname{cod} d$, которая ставит им в соответствие σ-стрелку $e \circ d$ с $\operatorname{dom} (e \circ d) \operatorname{dom} d$ и $\operatorname{cod} (e \circ d) = \operatorname{cod} e$, т. е. $e \circ d : \operatorname{dom} d \to \operatorname{cod} e$. При этом должен выполняться закон ассоциативности: пусть заданы три σ-стрелки $d : x \to y$; $e : y \to z$ и $g : z \to t$. Тогда $g \circ (e \circ d) = (g \circ e) \circ d$. Введение для каждого σ-объекта x единичной или тождественной σ-стрелки $1_x : X \to x$, так, что выполняется закон тождества: для любых двух σ-стрелок $d : x \to y$ и $e : y \to z$ следует, что $1_y \circ d = d$ и $e \circ 1_y = e$.

Описание категории DATA, соответствующей подмодели хранения исходных данных в БД, приводится в табл. 4.2, а диаграмма без единичных стрелок для этой категории — на рис. 4.119.

Необходимо отметить, что функции по модификации содержимого БД (пополнение, удаление, изменение) совмещены в одной DATA-стрелке «загрузка данных в БД...» для каждой модели данных в отдельности. Так как каждая DATA-стрелка

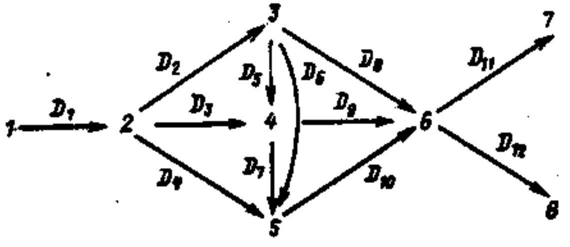


Рис. 4.119. Диаграмма категории DATA

Таблица 4.2

ДАТА-объекты	ДАТА-стрелки	
(1) Исходные данные для классификационных построений	$(1) \xrightarrow{D_1} (2)$	Кодирование исходных данных
(2) Исходные данные, для которых значения качественных признаков заменены кодами	$(2) \xrightarrow{D_2} (3)$ $(2) \xrightarrow{D_3} (4)$ $(2) \xrightarrow{D_4} (5)$	Загрузка данных в БД по сетевой, иерархической и реляционной модели соответственно
(3) Исходные данные, загруженные в БД в соответствии с сетевой моделью данных	$(3) \xrightarrow{D_5} (4)$	Реструктуризация сетевых структур хранения данных в иерархические
	$(3) \xrightarrow{D_6} (5)$	Реструктуризация сетевых структур хранения данных в реляционные
(4) Исходные данные, загруженные в соответствии с иерархической моделью данных		
(5) Исходные данные, загруженные в БД в соответствии с реляционной моделью данных	$(4) \xrightarrow{D_7} (5)$	Реструктуризация иерархических структур хранения данных в реляционные
(6) Множество выбранных исходных данных по заданному набору признаков и заданному набору объектов	$(3) \xrightarrow{D_8} (6)$ $(4) \xrightarrow{D_9} (6)$ $(5) \xrightarrow{D_{10}} (6)$	Выборка данных по заданному набору признаков и заданному набору объектов для сетевой, иерархической и реляционной моделя соответственно
(7) Исходные данные в форме матрицы «объект—свойства»	$(6) \xrightarrow{D_{11}} (7)$	Представление вибрационных данных в виде матрицы «объект—свойства»
(8) Раскодированные исходные данные, представленные в виде распечатки для пользователей	$(7) \xrightarrow{D_{12}} (8)$	Декодирование и визуализация выбранных данных в нужном для пользователя виде

представляет собой некоторую вычислительную функцию, реализующую то или иное преобразование данных, то операцию, которая ставит в соответствие каждой паре DATA-стрелок D_i и D_j с $\operatorname{dom} D_j = \operatorname{cod} D_i$ DATA-стрелку $D_j \circ D_i$, композицию с $\operatorname{dom} (D_j \circ D_i) = \operatorname{dom} D_i$ и $\operatorname{cod} (D_j \circ D_i) = \operatorname{cod} D_j$, определим как функциональную связь между двумя вычислительными функциями D_i и D_j. Тогда композицией можно считать более общую вычислительную информацию, в которой последовательно выполняются сначала функция D_i, а затем функция D_j, при этом очевидно, что выполняется закон ассоциативности, т. е. $D_k \circ (D_j \circ D_i) = (D_k \circ D_j) \circ D_i$ для любых DATA-стрелок D_k, D_j, D_i, таких, что

$$a \xrightarrow{D_i} b \xrightarrow{D_j} c \xrightarrow{D_k} d.$$

В связи с тем что каждый DATA-объект представляет собой множество, то в качестве единичной стрелки $1_a : a \to a$ для любого DATA-объекта a возьмем функцию визуализации элементов этого множества в соответствии с некоторой фиксированной структурой. Тогда для любых двух DATA-стрелок $D_i : a \to b$ и $D_j : b \to c$ выполняется закон тождества $1_b \circ D_i = D_i$, $D_j \circ 1_b = D_j$. Другими словами, коммутативна диаграмма на рис. 4.120.

Аналогичным образом будут введены операции композиции и единичные стрелки для остальных категорий, перечислим для них лишь объекты и стрелки.

В табл. 4.3 приводится описание категории CLASS, соответствующей подмодели построения классификации, и оценки ее качества, и на рис. 4.121 — соответствующая ее диаграмма.

Заметим, что данное определение CLASS-объекта (3) позволяет понимать под ним не только автоматические классификации, т. е. разбиения, полученные с помощью математических методов, но и экспертные классификации — разбиения объектов на классы, проведенные экспертом. Поэтому в оценке и сравнении классификаций могут участвовать и автоматические, и экспертные классификации, что является важной особенностью данной модели.

При определении стрелок категории CLASS мы не указывали конкретных математических методов, которые могут использоваться при их реализации. Это связано с тем, что разработано большое количество различных мер, способов группировки объектов в классы алгоритмов разбиения объектов на классы, сравнения и оценок классификаций, в то время как перед нами стояла задача описать модель процесса построения классификации на глобальном уровне. Поясним только, что CLASS-стрелка (1)→(3) — построение классификаций без использования матрицы сходства между объектами фактически представляет экспертное разбиение исходного материала на классы и подходы, основанные на вычислении статистических характеристик.

$$\begin{tikzcd}
a \arrow[r, "D_i"] \arrow[dr, "D_i"'] & b \arrow[d, dashed, "1_b"] \arrow[dr, "D_j"] & \\
& b \arrow[r, "D_i"'] & c
\end{tikzcd}$$

Рис. 4.120. Пояснение к категориальной модели DATA

Таблица 4.3

CLASS-объекты	CLASS-стрелки	
(1) Матрица объект — свойства	$(1) \xrightarrow{C_1} (2)$	Построение матрицы сходства по определенной выбранной мере сходства между объектами
(2) Матрица сходства между объектами	$(2) \xrightarrow{C_2} (3)$	Разбиение исходного множества на классы с использованием определенного выбранного способа группировки или уровня порога
(3) Множество классов — классификация	$(1) \xrightarrow{C_3} (3)$	Разбиение исходного множества на классы без промежуточного построения матрицы сходства
(4) Множество оценок разбиения исходных данных на классы в соответствии с выбранным критерием качества	$(3) \xrightarrow{C_4} (4)$	Оценивание качества классификации с точки зрения полученной структуры классов
(5) Множество классификаций	$(3) \xrightarrow{C_5} (5)$	Накопление классификаций
(6) Множество наилучших классификаций в соответствии с выбранным критерием	$(3) \xrightarrow{C_6} (7)$	Построение для каждого класса классификации типичного объекта
(7) Множество типичных для каждого класса классификаций объектов	$(5) \xrightarrow{C_7} (6)$	Сравнение различных классификаций

В табл. 4.4 приводится описание категории INFO для подмодели выделения из множества признаков наиболее информативных, а в табл. 4.5 — описание категории DIAGN для подмодели идентификации объектов по заданной классификации. Соответствующие им диаграммы приведены на рис. 4.122 и 4.123.

Определим категорию классификационных построений (КП) для описания соответствующей полной модели процесса исследования, построения и использования классификаций.

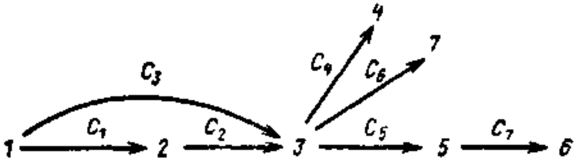


Рис. 4.121. Диаграмма категории CLASS

Таблица 4.4

INFO-объекты	INFO-стрелки	
(1) Матрица «объект—свойства» [$m * n$], где m — число объектов, n — число признаков	$(1) \xrightarrow{I_1} (2)$	Построение матрицы сходства по заданной мере сходства между объектами
(2) Матрица сходства [$m * m$]	$(1) \xrightarrow{I_2} (3)$	Построение подмножества информативных признаков по заданной матрице «объект—свойства»
(3) Множество информативных признаков	$(2) \xrightarrow{I_3} (3)$	Многомерное шкалирование

КП-объекты — это объекты категорий DATA, CLASS, INFO и DIAGN. В случае, если $O^{i}_{\text{DATA}}=O^{j}_{\text{CLASS}}$, то в категории КП будет единственный объект $O^{k}=O^{i}_{\text{DATA}}=O^{j}_{\text{CLASS}}$. Это утверждение справедливо для любой пары категорий DATA, CLASS, INFO и DIAGN.

Очевидно, что категории DATA, CLASS, INFO и DIAGN являются полными подкатегориями категории КП. Это следует непосредственно из определения подкатегории [52] и из того, как была введена категория КП.

Как отмечалось ранее, стрелками подкатегорий являются вычислительные функции. Тогда объекты подкатегорий DATA, CLASS, INFO и DIAGN можно рассматривать как входные и выходные параметры этих вычислительных функций. В силу того что DATA, CLASS, INFO и DIAGN — полные подкатегории КП, то, следовательно, КП-объекты — входные и выходные параметры, а КП-стрелки — вычислительные функции. Пусть $\varphi:(i)\rightarrow(j)$, тогда $\operatorname{dom}\varphi=(i)$ — входной параметр, $\operatorname{cod}\varphi=(j)$ — выходной параметр вычислительной функции φ.

Таблица 4.5

DIAGN-объекты	DIAGN-стрелки	
(1) Классификация, у которой классы заданы как множество объектов	$(6) \xrightarrow{A_1} (1)$	Отнесение неизвестного объекта к классу
(2) Классификация, у которой классы заданы типичными объектами	$(6) \xrightarrow{A_2} (2)$	То же
(3) Совокупность решающих правил	$(6) \xrightarrow{A_3} (3)$	»
(4) Грамматика распознавания	$(6) \xrightarrow{A_4} (4)$	»
(5) Экспертные правила (правила Modus Ponens)	$(6) \xrightarrow{A_5} (5)$	»
(6) Множество неизвестных объектов, описанных значениями признаков		

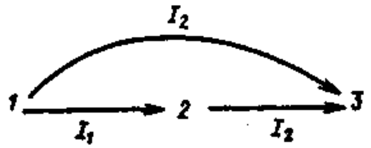


Рис. 4.122. Диаграмма категории INFO

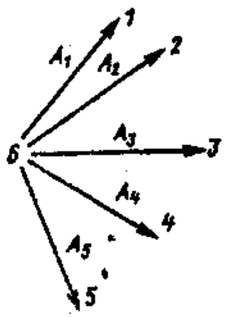


Рис. 4.123. Диаграмма категории DIAGN

Таким образом, на множество КП-объектов определено отношение функциональной зависимости R. Более того, из определения КП-стрелок взаимосвязями между КП-объектами следует, что отношение R является частным случаем отношения предпорядка, так как любые два КП-объекта соединены не более чем одной стрелкой и для любых КП-объектов справедливы свойства рефлексивности и транзитивности. Следовательно, если КП-категория является категорией предпорядка, а категория предпорядка, как показано в [52], есть топос, то КП-категория — топос. В итоге можно сделать вывод, что категориальная модель процесса классификационных построений есть топос.

Заметим, что полученный результат может иметь важное значение при разработке интеллектуальных вычислительных систем специальной архитектуры — так называемых топос-машин. В основе этих машин лежит понятие топоса. В топос-машинах программно-аппаратно реализуются действия в некото-

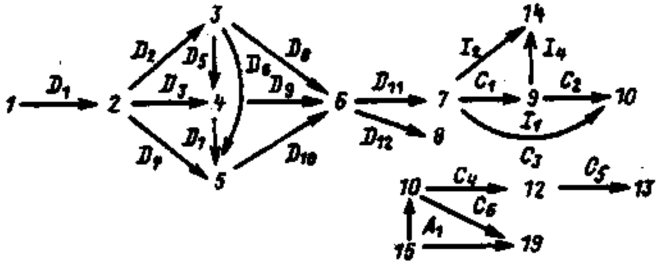


Рис. 4.124. Категориальная модель процесса классификационных построений

рой предметной области, модель знаний о которой представляет собой топос. Таким образом, то, что модель классификационных построений является топосом, служит основой для проектирования и реализации в дальнейшем специализированных топос-машин, ориентированных на эффективное выполнение классификационных построений.

Если каждую вычислительную функцию трактовать как некоторый программный модуль, то в топосе фактически выражаются значения о задачах классификационных построений и способах их автоматизированного решения. Топос выступает как описание семантической модели процесса исследований, построения и использования классификаций в классификационной системе.

Для использования модели классификационных построений в автоматизированной системе необходимо описать ее на некотором языке представления знаний (ЯПЗ).

Заметим, что в топосе отражена наиболее общая модель процесса классификационных построений, тогда как в ЯПЗ должны быть предусмотрены возможности для представления таких важных особенностей предметной области построения, анализа и использования классификаций, как наличие разных методов, реализующих одни и те же вычислительные функции. На рис. 4.124 приведена категориальная модель процесса классификационных построений, на которой для иллюстрации вышесказанного выделены некоторые морфизмы: D_i — морфизмы подкатегорий DATA, C_i — морфизмы подкатегорий CLASS, I_i — морфизмы подкатегорий INFO, A_i — морфизмы подкатегорий DIAGN. Данные морфизмы реализуются несколькими методами, которые включают меру сходства $C_1=${манхэттенское расстояние, евклидово расстояние, расстояние Махалонобиса, метрика Хемминга и т. д.}, способ группировки $C_2=${«ближайший сосед», «дальний сосед», центроидный метод, семейство алгоритмов FOREL и т. д.}, сравнение классификаций $C_7=${метод Чупрова, метод Сокала—Рольфа, метод Крустала—Гудмана и т. д.}, диагностику по решающим правилам $A_3=${метод линейных дискриминантных функций, метод квадратичных дискриминантных функций, метод логических функций и т. д.}.

Как видно, с помощью данной модели процесса классификационных построений описаны внешние взаимосвязи и рассмотрены этапы этого процесса в деталях. Следующим по важности вопросом является вопрос выбора методов и способов классификационных построений, который связан в основном с инженерной реализацией и практической направленностью системы.

4.10. ВЕКТОРЫ, ФРЕЙМЫ И МАТРИЦЫ В КАТЕГОРИИ

В данном параграфе приводится категориальное описание векторов (некоторого пространства), далее рассматриваются функторы — фреймы для описания семантики моделей предметной области. После этого рассматривается категориальное описание матриц реляционной алгебры. В заключение излагается категориальная модель полинейной алгебры и тензоров (рис. 4.125).

4.10.1. Векторы в категории

Проанализируем несколько важных свойств категории vect, объектами которой являются векторные пространства, а морфизмами — линейные отображения [36, 37, 52].

Пример 4.52. Пусть $F : R^3 \to R^2$ — линейное отображение, соответствующее матрице 2×3, элементами которой являются линейные отображения $R \to R$, тогда

$$(f_i^j) = \begin{pmatrix} f_1^1 & f_2^1 & f_3^1 \\ f_1^2 & f_2^2 & f_3^2 \end{pmatrix}.$$

Поскольку $R^3 = R+R+R$, то f соответствует (f_1, f_2, f_3) для $f_i = f \circ in_i : R \to R^2$, $i=1, 2, 3$. Заметим, что каждое f_i соответствует $\begin{pmatrix} f_i^1 \\ f_i^2 \end{pmatrix}$, где $f_i^j = \pi_j \circ f_i : R \to R$, $i=1, 2, 3$; $R^2 = R \times R$; $j=1, 2$; in_i — инъективное отображение; $\circ$ — оператор композиции; π_j — проекция на j-компонент. Из теории категорий следует, что

$$f(x, y, z) = f_1(x) + f_2(y) + f_3(z) = \begin{pmatrix} f_1^1(x) + f_2^1(y) + f_3^1(z) \\ f_1^2(x) + f_2^2(y) + f_3^2(z) \end{pmatrix}.$$

Одно из важнейших свойств vect состоит в том, что копроизведение (раздельное объединение) $A_1 + A_2 \dots + A_n$ конечного числа объектов как объект изоморфно объекту произведения $A_1 \times A_2 \times \dots \times A_n$. Следовательно, R^n — произведение в vect с проекциями $\pi_j : R^n \to R$, $(x_1 \dots x_n) \to x_j$ и копроизведение в vect с инъекциями $\mathrm{in}_i : R \to R^n$, $x \to \begin{pmatrix} 0, \dots, X, 0, \dots, 0 \\ 1, \dots, i, i+1, \dots, n \end{pmatrix}$. Сформулируем следующее утверждение, которое проиллюстрируем примером.

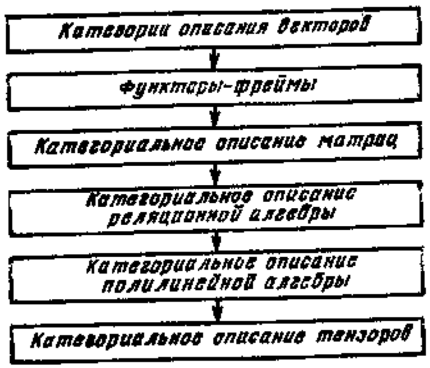


Рис. 4.125. Последовательность категориального описания тензоров

Предложение. Если A — векторное пространство, то переход от $g: R \to A$ к $g(\lambda)$ в A устанавливает биекцию между линейными преобразованиями из R в A и элементами в A. Обратное отображение приписывает каждому $a \in A$ линейное преобразование $\lambda \to \lambda a$.

Пример 4.53. Рассматривая преобразование $F: R^3 \to R^2$, заключаем, что каждое отображение $f_j{}^i: R \to R$, как и в предыдущем примере, соответствует скаляру $\lambda_i{}^j$ в R. Поэтому f соответствует матрице, содержащей скаляры

$$(\lambda_i^j), \quad i = 1, 2, 3; \quad j = 1, 2,$$

т. е.

$$f(x, y, z) = \begin{pmatrix} \lambda_1^1 x + \lambda_2^1 y + \lambda_3^1 z \\ \lambda_1^2 x + \lambda_2^2 y + \lambda_3^2 z \end{pmatrix},$$

что по правилам линейной алгебры записывается в виде

$$\begin{pmatrix} \lambda_1^1 & \lambda_2^1 & \lambda_3^1 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Рассмотрим представление векторов и матриц в категории. Разработка способа категориального представления векторов и матриц требует уточнения применяемой терминологии. Для K-объекта, т. е. объекта в категории K, A, и целого числа $n \geqslant 1$ назовем n-арным коразложением объекта A диаграмму копроизведения вида $(in_i: A_i \to A \mid i = 1, \ldots, n)$. В силу универсального свойства имеется биективное соответствие (рис. 4.126, *а*) морфизмов $f: A \to B$ и наборов $(f_i: A_i \to B \mid i = 1, \ldots, n)$, которые будем записывать в виде $(f_1, \ldots, f_n)$ (вектор-строка). Данное соответствие означает, что $f \circ in_i = f_i$.

Для K-объекта A и целого числа $m \geqslant 1$ назовем m-арным разложением объекта А диаграмму произведения вида

$$(\pi_j: A \to B_j, \quad j = 1, \ldots, m).$$

Соответствие (рис. 4.126, *б*) морфизмов $g: B \to A$ и наборов $(g^j: B \to B_i \mid j = 1, \ldots, m)$, которое задается диаграммой на рис. 4.125, запишем в виде

$$\begin{pmatrix} g_1 \\ \cdots\cdots \\ g_m \end{pmatrix}$$

(вектор-столбец).

$$\begin{tikzcd}
A_i \arrow[r, "in_i"] \arrow[dr, "f_i"'] & A \arrow[d, "f"] \\
 & B
\end{tikzcd}
\qquad
\begin{tikzcd}
A \arrow[r, "\pi_j"] & B_j \\
B \arrow[u, "g"] \arrow[ur, "g^j"'] &
\end{tikzcd}$$

а) *б)*

Рис. 4.126. К представлению векторов как категорий

В общем случае при выполнении n-арного разложения A и m-арного коразложения B производятся следующие действия:

1) морфизм $f: A \to B$ коразлагается в вектор-строку (f_i);
2) каждый f_i разлагается в вектор-столбец $(f_i)^j$.

Можно поступить иначе:

1) сначала разложить f;
2) затем коразложить f^j.

Оба случая отражены в диаграмме на рис. 4.127.

Докажем, что $(f_i)^j = (f^j)_i$ для всех i и j.

Действительно, пусть $(in_i: A_i \to A)$ — n-арное коразложение A, а $(\pi_j: B \to B_j)$ — m-арное разложение B. Тогда для каждого семейства

$$f_i^j: A_i \to B_j, \quad i = 1, \ldots, n; \quad j = 1, \ldots, m,$$

существует единственное отображение $f: A \to B$, такое, что справедлива диаграмма на рис. 4.128, т. е. $f_i{}^j = \pi_j \circ f \circ in_i$ для всех i, j (рис. 4.129). Последнее утверждение проверяется путем анализа случаев.

1. Зафиксируем i; существует единственное f_i, для которого $\pi_j \circ f_i = f_i{}^j$, $j = 1, \ldots, m$.

2. Изменяя i, получаем морфизмы f_i, для которых существует единственное f с $f \circ in_i = f_i$. Поэтому $\pi_j \circ f \circ in_i = \pi_j \circ f_i = f_i{}^j$. Если также $\pi_j \circ g \circ in_i = f_i{}^j$, то, фиксируя i и изменяя j, получаем $g \circ in_i = f_i$. Тогда $g = f$, что и требовалось доказать. Ясно, что в vect существует биективное соответствие между базисами пространства R и разложениями вида $\pi_j: R^n \to R$. Это следует из определения для базиса $x_1, \ldots, x_n$ отображения со свойствами $\pi_j(\Sigma \lambda_i x_i) = \lambda_j$.

$$\begin{tikzcd}
A_i \arrow[r, "(f_i)^j"] \arrow[d, "in_i"'] \arrow[dr, "f_i"] & B_j \\
A \arrow[r, "f"] & B \arrow[u, "\pi_j"']
\end{tikzcd}
\qquad
\begin{tikzcd}
A_i \arrow[r, "(f^j)_i"] \arrow[d, "in_j"'] & B_j \\
A \arrow[r, "f"] \arrow[ur, "f^j"] & B \arrow[u, "\pi_j"']
\end{tikzcd}$$

а) *б)*

Рис. 4.127. Пояснение к n-арному разложению A (*а*) и m-арному B (*б*) в категориях

$$\begin{tikzcd}
A_i \arrow[r, "f_i^j"] \arrow[d, "in_i"'] & B_j \\
A \arrow[r, dashed, "f"] & B \arrow[u, "\pi_j"']
\end{tikzcd}$$

Рис. 4.128. Пояснение к доказательству соотношения $(f_i)^j = (f^j)_i$

$$\begin{tikzcd}
A_i \arrow[r, "f_i^j"] \arrow[d, "in_i"'] \arrow[dr, dashed, "f_i"] & B_j \\
A \arrow[r, "f"] & B \arrow[u, "\pi_j"']
\end{tikzcd}$$

Рис. 4.129. Диаграмма соотношения $f_i{}^j = \pi_j \circ f \circ in_i$

Векторное пространство vect. Пусть vect в качестве объектов имеет векторные пространства и для каждой пары объектов A, B vect (A, B) является множеством линейных отображений из A в B.

Определим следующие свойства:

композиция $A \xrightarrow{f} B \xrightarrow{g} C : a \rightarrow (g \circ f)(a) = g(f(a))$;

тождественное отображение $A \xrightarrow{\mathrm{id}_A} A : a \rightarrow a$.

Покажем, что если g и f — линейные отображения, то $g \circ f$ и id_A линейны. Действительно, $T = (g \circ f)(\lambda \circ a + \lambda' \circ a') = g(f(\lambda \circ a + \lambda \circ a')) = g(\lambda \circ f(a) + \lambda' \circ f(a'))$ в силу линейности f; $T = \lambda \circ g(f(a)) + \lambda' \circ g(f(a'))$ в силу линейности g; $T = \lambda \circ (g \circ f)(a) + \lambda' \circ (g \circ f)(a')$, доказывает линейность $(g \circ f)$. Линейность каждого из id_A очевидна. Можно доказать, что единичные морфизмы и композиции удовлетворяют условию ассоциативности и аксиомам единичных морфизмов. Отсюда следует, что vect — категория.

Векторные пространства можно пояснить на примере R^n, т. е. декартова произведения n экземпляров действительной прямой.

Соглашение: для $x_i \in R$, $i = 1, \ldots, n$, вектор-столбец записываем в виде

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}.$$

Сложение векторов записываем так:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} x_1' \\ x_2' \\ \cdots \\ x_n' \end{pmatrix} = \begin{pmatrix} x_1 + x_1' \\ x_2 + x_2' \\ \cdots \\ x_n + x_n' \end{pmatrix},$$

или в сокращенном виде $(x + x')_j = x_j + x_j'$, $j = 1, \ldots, n$, для всех $x, x' \in R^n$.

Умножение вектора на скаляр покомпонентно записывает как

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \cdots \\ \lambda x_n \end{pmatrix},$$

или в сокращенном виде $(\lambda x)_j = \lambda x_j$, $j = 1, \ldots, n$, для всех $x \in R^n$ и $\lambda \in R$.

Символ O обозначает число $O \in R$, а вектор $O \in R^n$; $-x$ — вектор с $(-x_j) = -(x_j)$ для $j = 1, 2, \ldots, n$.

Для всех x, x', $x'' \in R$ из свойства действительных чисел имеем следующие:

1) $(x + x') + x'' = x + (x' + x'')$;
2) $x + x' = x' + x$;
3) $x + 0 = x$;
4) $x + (-x) = 0$.

Умножение на скаляр имеет следующие свойства:

5) $1 \cdot x = x$;
6) $\lambda(x + x') = \lambda x + \lambda x'$;
7) $(\lambda + \lambda') \cdot x = \lambda x + \lambda' x$;
8) $\lambda(\lambda' x) = (\lambda\lambda') x$.

Любое пространство, имеющее эти свойства, считаем векторным пространством.

Сделаем некоторые обобщения.

Действительное векторное пространство есть набор X с двумя функциями:

а) сложения — $x \times x \rightarrow x : (x, x') \rightarrow x + x'$;

б) умножения на скаляр — $R \times x \rightarrow x : (\lambda, x) \rightarrow \lambda x$, удовлетворяющими свойствам 1—4 и 5—8 соответственно для 0, x, $x' \in x$ и λ, $\lambda' \in R$. Для произвольного (возможно, бесконечного) множества I рассмотрим набор функций: $R^I = \{f / f : I \rightarrow R\}$.

Если $I = \{1, \ldots, n\}$, то этот набор функций есть в точности R^n [дополнительно отождествим каждую функцию f с вектор-столбцом X, для которого $x_j = f(j)$].

Подкомпонентные (подкоординатное) сложение и умножение на скаляр для $f, f' \in R^I$, $\lambda \in R$, $i \in I$ имеют вид

$$(f + f')(i) = f(i) + f'(i);$$
$$(\lambda f)(i) = \lambda \circ f(i).$$

Проверка свойств 1—8 трудностей не вызывает, поэтому R^I (со сложением и умножением на скаляр) есть векторное пространство.

Для представления с помощью категорий семантики предметной области вводятся основные соглашения:

1) модель предметной области состоит из объектов;

2) между объектами существуют отношения, т. е. морфизмы (отображения).

Заметим, что из соглашений 1, 2 можно заключить, что модель предметной области образует категорию. Какими именно свойствами наделена эта категория, т. е. ее объекты и морфизмы, зависит от предмета исследования. Один из самых общих приемов комбинирования объектов в отношения состоит в составлении произведения объектов. Это можно сделать двояким образом:

1) определяя произведения;

2) определяя копроизведения, сохраняющие все индивидуальные особенности участвующих в них объектов.

Из свойств произведений и копроизведений следует связь отношений частичного порядка E или $\leqslant$ в категории, с одной стороны, и ISA-отношением для понятий или фреймов базы знаний, с другой стороны. При планировании вычислений особую роль приобретает построение механизма вызова (по имени) процедур, которые в свою очередь, приводят к вызову других процедур и т. д. В общем виде эта задача носит название «проблемы присвоения имен». Ограничимся ее частным случаем и перейдем к анализу вопроса, как по имени объекта предметной области (категории) вычислить его значение. Представим общий категориальный метод, позволяющий для произвольного объекта категории, который может оказаться морфизмом (фреймом, отношением, функцией, функтором и т. д.), установить связь его имени и значения.

4.10.2. Вычисление значений в категории

Для множеств A и B образуем совокупность в B^A всех функций из A в B, т. е. $B^A = \{f : f$ — функция из A в $B\}$.

Будем ассоциировать с множеством B^A стрелку $\widehat{ev} : B^A \times A \to B$, определенную правилом $\widehat{ev}(\langle f, x\rangle) = f(x)$ или, что эквивалентно, $\widehat{ev}\langle f, x\rangle = f(x)$.

Стрелку $\widehat{ev}$ назовем функцией значения. Входами этой функции служат пары $\langle f, x\rangle$, где $f : A \to B$, $x \in A$, а ее действие состоит в вычислении f в точке x, т. е. $f(x)$. Представим категориальное описание B^A. Отметим, что для любой функции $g : C \times A \to B$ существует одна и только одна функция $\hat{g} : C \to B^A$, для которой диаграмма на рис. 4.130 коммутативна. На паре $\langle c, a\rangle \in C \times A$ значение функций $\hat{g} \times \mathrm{id}_A$ — это $\langle \hat{g}(c), \mathrm{id}_A(a)\rangle = \langle \hat{g}(c), a\rangle$, т. е. при каждом фиксированном c функция g задает отображение $A \to B$, когда и a пробегает множество A.

$$\begin{tikzcd}
B^{A}\times A \arrow[rd, "ev"] & \\
 & B \\
C\times A \arrow[uu, dashed, "\hat{g}\times id_A"] \arrow[ru, "g"'] &
\end{tikzcd}$$

а)

Рис. 4.130. Пояснение к категориальному описанию B^A

$$\begin{tikzcd}
b^{a}\times a \arrow[rd, "ev"] & \\
 & b \\
c\times a \arrow[uu, dashed, "\hat{g}\times 1_a"] \arrow[ru, "g"'] &
\end{tikzcd}$$

б)

Рис. 4.131. Пояснение к операции экспоненцирования

Для данного $c \in C$ определим $g_c : A \to B$ как $g_c(a) = g(\langle c, a\rangle)$ для каждого $a \in A$.

Теперь полагаем $\hat{g}(c) = g_c$, и для произвольной

$$\langle c, a\rangle \in c \times A$$

$$\widehat{ev}\langle \hat{g}(c), a\rangle = g(\langle c, a\rangle).$$

Коммутативность диаграммы теперь становится очевидной, так как

$$\widehat{ev}\langle \hat{g}(c), a\rangle = \widehat{ev}\langle g_c, a\rangle = g_c(a) = g(\langle c, a\rangle).$$

Будем считать, что в категории K определено экспонирование (вычисление значения), если в ней существует произведение двух любых объектов a и b, для которых существуют K-объект b^a (экспоненциал) и K-стрелка $\widehat{ev} : b^a \times a \to b$ (стрелка значения), такие, что для любых K-объекта c и K-стрелки $g : c \times a \to b$ существует единственная стрелка $\hat{g} : c \to b^a$, для которой диаграмма на рис. 4.131 коммутативна, т. е. $\widehat{ev}(\ \ ) \circ (\hat{g} \times 1a) = g$. Отметим, что функция, ставящая в соответствие стрелке g стрелку $\hat{g}$, определяет изоморфизм $K(c \times a, b) \cong K(c, b^a)$. Последнее соответствие в терминах множеств A, B и C означает, что

$$A \times B \to C \cong A \to (B \to C).$$

Вычисление значения по имени. Элементом множества B^A является функция $f : A \to B$, т. е. $f \in B^A$. Это приводит к функции $\ulcorner f \urcorner : \{0\} \to B^A$, такой, что $\ulcorner f \urcorner(0) = f$, т. е. $\ulcorner f \urcorner$ — имя функции f.

Если $x \in A$, то категориальный «элемент» $\bar{x} : \{0\} \to A$ определяется равенством $\bar{x}(0) = x$. Из $\widehat{ev}\langle f, x\rangle = f(x)$ следует, что $\widehat{ev}(\ \ ) \circ \langle \ulcorner f \urcorner, \bar{x}\rangle(0) = \widehat{ev}(\ulcorner f \urcorner(0), \bar{x}(0)) = f(x) = f(\bar{x}(0))$, т. е. получим $\widehat{ev}(\ \ ) \circ \langle \ulcorner f \urcorner \bar{x}\rangle = f \circ \bar{x}$.

В общем случае по K-стрелке $f:a\to b$ найдем стрелку $f\circ pr_a:1\times a\to b$, которая является композицией:

$$f\circ pr_a:1\times a\to a\to b.$$

Именем f назовем стрелку $\ulcorner f\urcorner:1\to b^a$, которая экспоненциально присоединена к стрелке $f\circ pr_a:1\times a\to b$. Поэтому $\ulcorner f\urcorner$ — единственная стрелка, для которой диаграмма на рис. 4.132 коммутативна. Для любого K-элемента $x:1\to a$ объекта a справедливо $\widehat{ev}(\ \ )\circ\langle\ulcorner f\urcorner, x\rangle=f\circ x$.

При последующем определении фреймов как основных конструкций базы знаний полезным оказывается понятие функтора (функционала). Приведем необходимые определения, не решая детально задачи вычисления значений функторов.

В дальнейшем функтор отождествим с фреймом и представим все детали вычисления его значения. В силу важности этого результата приведем подробные определения функторов из теории категорий.

Напомним определение функторов. Ковариантным функтором F из категории K в категорию D называется функция, ставящая в соответствие:

1) каждому K-объекту a некоторый D-объект $F(a)$;

2) каждой K-стрелке $f:a\to b$ D-стрелку $F(f):F(a)\to F(b)$, такую, что:

а) $F(1a)=1_{F_a}$ для каждого K-объекта a, т. е. единичной стрелки, соответствующей объекту a, сопоставляется единичная стрелка, соответствующая объекту $F(a)$;

б) $F(g\circ f)=F(g)\circ F(f)$ для любых g и f, для которых определена композиция $g\circ f$.

$$\begin{tikzcd}
b^a\times a \arrow[rd, "\hat{e}v(\ )"] & \\
 & b \\
1\times a \arrow[uu, "\ulcorner f\urcorner\times 1a"] \arrow[ru, "f\circ pr_a"'] &
\end{tikzcd}$$

Рис. 4.132. Пояснение к вычислению значения по имени

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[rd, "h"'] & b \arrow[d, "g"] \\
 & c
\end{tikzcd}
\qquad
\begin{tikzcd}
F(a) & F(b) \arrow[l, "F(f)"'] \\
 & F(c) \arrow[u, "F(g)"'] \arrow[lu, "F(h)"]
\end{tikzcd}$$

а) б)

Рис. 4.133. Пояснение к определению ковариантного функтора

$$\begin{tikzcd}
a \arrow[r, "f"] \arrow[rd, "h"'] & b \arrow[d, "g"] \\
 & б
\end{tikzcd}
\qquad
\begin{tikzcd}
F(a) \arrow[r, "F(f)"] \arrow[rd, "F(h)"'] & F(b) \arrow[d, "F(g)"] \\
 & F(c)
\end{tikzcd}$$

а) б)

Рис. 4.134. Пояснение к определению контравариантного функтора

Из определения функторов следует, что коммутативность диаграммы на рис. 4.133, *а* в K влечет коммутативность диаграммы на рис. 4.133, *б* в D.

Контравариантный функтор $F:K\to D$ ставит в соответствие стрелке $f:a\to b$ стрелку $F(f):F(b)\to F(a)$, так, что $F(1a)=1$ и $F(g\circ f)=F(f)\circ F(g)$, т. е. коммутативная программа в K, представленная на рис. 4.134, *а*, переходит в коммутативную диаграмму на рис. 4.134, *б*.

Категориальная модель языка фреймов. Начнем с фиксации алфавита:

1) v_1, v_2, ... — бесконечный список индивидных переменных;

2) &, $\vee$, $\neg$, $\Rrightarrow$ — пропозициональные связки;

3) $\forall$, $\exists$, $\approx$,), (— вспомогательные символы.

Знак $\approx$ вводится следующим образом. Предположение $\exists v(v\approx c)$ равносильно утверждению, что имеется актуально существующий, т. е. всюду определенный, элемент, равный c. Иначе говоря,

$$E(c)\equiv\exists c(v\approx c),$$

где $\equiv$ — знак равносильности, которая определяется фразой «если и только если». Выражение $\varphi\equiv\psi$ вводится для сокращения формулы

$$(\varphi\supset\psi)\wedge(\psi\supset\varphi).$$

В частности, в формуле

$$\widehat{ev}(f\approx g)\sqsubseteq\mathrm{dom}\,f\sqcap\mathrm{dom}\,g$$

эта стрелка означает, что начало стрелки объединено с концом стрелки.

Элементарный семантический язык зададим в виде пары

$$L=\{\bar{R},\ \bar{C}\},$$

где $\bar{R}=\{R_1,\ R_2\ldots\}$ — множество двухместных предикатов-символов; $\bar{C}=\{C_1\ldots\}$ — множество констант. Он позволяет сформулировать все принципиальные особенности фреймов.

Под термами (функциональными фреймами) понимаются выражения, обозначающие индивиды (в L — это индивидные переменные v). Атомарные формулы соответствуют простым фреймам. В L — это $t\approx u$, tR_ju, где t и u — термы.

Формулы соответствуют фреймам и определяются по индукции:

1) атомарная формула (простой фрейм) есть формула (фрейм);

2) если φ и ψ — формулы (фреймы), то $(\varphi\,\&\,\psi)$, $(\varphi\vee\psi)$, $(\varphi\wedge\psi)$, $(\neg\,\varphi)$ — формулы (фреймы);

3) если φ — формула (фрейм), v — индивидная переменная, то и $(\forall v)\varphi$ и $(\exists v)\varphi$ — формулы (фреймы).

Кванторы $\forall$, $\exists$ связывают переменные. Если переменные не находятся в сфере действия кванторов, они считаются свободными.

Предложением (замкнутым фреймом) считается формула (фрейм), все переменные которой связаны. Фрейм формула), имеющий, по крайней мере, одно свободное вхождение переменной, называется *открытым* (открытой). В выражении $\varphi(v)$ считаем, что переменная v имеет свободное вхождение. В общем случае указываются все (или некоторые) свободные переменные.

Моделью языка фреймов, или L-моделью, назовем тройку $A=(A, \overline{R}, \overline{\tau})$, состоящую из:

непустого множества A;

отношений $R_i\subseteq A\times A$, $R_i\in\overline{R}$;

индивидов $C_j\in A$, $C_j\in\overline{C}$.

Метод интерпретации переменных в модели A состоит в следующем. Пусть x — произвольная функция x: $n\to x_n$, где $x_n\in A$. Назовем эту функцию A-оценкой и представим ее в виде последовательности $x=\langle x_1, x_2, \ldots, x_i\ldots\rangle$, i-й член которой является значением переменной v_i, которое дается оценкой x. Через $x(a/i)=\langle x_1, x_2, \ldots, x_{i-1}, a, x_{i+1}\ldots\rangle$ обозначим оценку, получаемую заменой x_i элементом $a\in A$. Тот факт, что формула (фрейм) выполняется в структуре A при оценке x, будем записывать в виде $A\models\varphi[x]$ либо в виде $A\models\varphi[x_1, \ldots, x_n]$. Последнее выражение понимается так, что φ выполняется при значениях x_j для переменных v_j, $j=1, \ldots, n$.

Если все переменные фрейма φ встречаются в списке v_1, $v_2, \ldots, v_m$, $m\geqslant 1$, то число m назовем подходящим для фрейма φ.

Положим $\varphi^m=\{\langle x_1 \ \ldots, \ x_m\rangle : A\models\varphi[x_1\ldots x_m]\}$, т. е. φ-множество всех m-членных последовательностей, на которых в модели A выполняется фрейм φ.

Алгебра фреймов легко формулируется на языке m-членных последовательностей.

1. Дополнением к фрейму φ^m, т. е. множеством m-последовательностей, не удовлетворяющих φ, является множество m-последовательностей, удовлетворяющих $\neg\,\varphi : (\neg\,\varphi)^m=-\varphi^m$.

$$\begin{tikzcd}
 & x^m \arrow[dl] \arrow[d, dashed] \arrow[dr, "\widehat{ev}(\psi)^m"] & \\
\Omega & \Omega\times\Omega \arrow[l] \arrow[r] \arrow[d, "\cap"] & \Omega \\
 & \Omega &
\end{tikzcd}$$

а)

$$\begin{tikzcd}
 & x^m \arrow[dl, dash, "\hat{ev}(\psi)^m"'] \arrow[d, dashed] \arrow[dr, "\hat{ev}(\psi)^m"] & \\
\Omega & \Omega\times\Omega \arrow[l, dash] \arrow[r] \arrow[d] & \Omega \\
 & \Omega &
\end{tikzcd}$$

б)

$$\begin{tikzcd}
x^m \arrow[r, "ev(\varphi)^m"] \arrow[dr, "\hat{ev}(\neg\varphi)^m"'] & \Omega \arrow[d, "\neg"] \\
 & \Omega
\end{tikzcd}$$

в)

Рис. 4.135. Пояснение к операциям над фреймами:
а — конъюнкция; *б* — дизъюнкция; *в* — отрицание

2. Конъюнкцией фреймов φ^m и ψ^m, т. е. множеством m-последовательностей, удовлетворяющих φ^m и ψ^m, является пересечение множеств φ^m и ψ^m: $(\varphi\&\psi)^m=\varphi^m\cap\psi^m$.

3. Дизъюнкцией фреймов φ^m и ψ^m, т. е. множеством m-последовательностей, удовлетворяющих φ^m или ψ^m, является объединение множеств φ^m и ψ^m: $(\varphi\vee\psi)^m=\varphi^m\cup\psi^m$.

С другой стороны, выразим операторы алгебры фреймов в терминах значений фреймов, определив характеристическую функцию

$$\widehat{ev}(\varphi)^m : A^m\to\{1, \ 0\},$$

где 1 соответствует истине, а 0 — лжи, т. е.

$$\widehat{ev}(\varphi)^m(\langle x_1, \ \ldots\ldots, x_m\rangle)=\begin{cases} 1, & \text{если } A\models\varphi[x_1, \ \ldots\ldots, x_m]; \\ 0 & \text{в противном случае.}\end{cases}$$

Тогда легко получить следующие равенства:

1) дополнение $\widehat{ev}(\neg\,\varphi)^m=\neg\circ\widehat{ev}(\varphi)^m$;

2) конъюнкцию $\widehat{ev}(\varphi\&\psi)^m=\widehat{ev}(\varphi)^m\cap\widehat{ev}(\psi)^m=\cap\circ\langle\widehat{ev}(\varphi)^m$, $\widehat{ev}(\psi)^m\rangle$;

3) дизъюнкцию $\widehat{ev}(\varphi\vee\psi)^m=\widehat{ev}(\varphi)^m\cup\widehat{ev}(\psi)^m=\cup\circ\langle\widehat{ev}(\varphi)^m$, $\widehat{ev}(\psi)^m\rangle$,

где $\neg$, $\cap$, $\cup$ — классические истинностные функции.

Правилам вычисления значений операций над фреймами соответствуют диаграммы на рис. 4.135, на которых $\Omega=\{1, 0\}$.

Отображения $\neg$, $\cap$, $\cup$ вычисляются в соответствии с таблицами:

	$\neg$
I	0
0	I

$\cap$	I	0
I	I	0
0	0	0

$\cup$	I	0
I	I	I
0	I	0

Выразим значения предикатных символов R, входящих во фреймы. Пусть $r: A^2 \to \{1, 0\}$ — характеристическая функция множества $R \subseteq A \times A$. Тогда диаграмма на рис. 4.136, *а* коммутативна.

Обозначим

$$\rho_t^m(\langle x_1, \quad . \ . \ . \ . \ , \ x_m\rangle) = x_t;$$

$$\rho_u^m(\langle x_1, \quad . \ . \ . \ . \ , \ x_m\rangle) = x_u;$$

$$\widehat{ev}(tR_u)^m(\langle x_1, \quad . \ . \ . \ . \ , \ x_m\rangle) = \begin{cases} 1, & \text{если истинно;} \\ 0 & \text{в противном случае.} \end{cases}$$

В частности, можно положить, что R совпадает с равенством по смыслу ($\approx$). Тогда справедлива коммутативная диаграмма на рис. 4.136, *б*, где для $x, y \in A$

$$S_A(\langle x, y\rangle) = \begin{cases} 1, & \text{если} \quad x = y; \\ 0, & \text{если} \quad x \neq y, \end{cases}$$

т. е. S_A — символ Кронекера.

Вопрос о выработке методов графического представления фреймов для проектирования базы знаний представляет большой интерес. Он предполагает установление связи алгебры фреймов с реляционной алгеброй. Эту связь можно установить, пользуясь матричным методом, а именно характеристическими матрицами. Их можно применить для решения задачи матричного представления базы данных, служащей расширением фреймовой базы знаний (см. гл. 8).

$$\begin{tikzcd} A^m \arrow[r, "\langle \rho_t^m, \rho_u^m \rangle"] \arrow[dr, "tR_u^m"'] & A^2 \arrow[d, "r"] \\ & \{1, 0\} \end{tikzcd}$$

$$\begin{tikzcd} A^m \arrow[r] \arrow[dr, "\widehat{ev}(t \approx u)^m"'] & A^2 \arrow[d, "\delta_A"] \\ & \{1, 0\} \end{tikzcd}$$

б)

Рис. 4.136. Пояснение к предикатному фрейму

4.11. КАТЕГОРИАЛЬНАЯ МОДЕЛЬ СЕМАНТИЧЕСКОЙ КВАНТОВОЙ ИНФОРМАЦИИ ПО В. Г. ТОЛСТОВУ

Материал данного параграфа следует рассматривать как еще одну попытку построения теории и модели семантической информации на базе теории категорий, которая необходима кибернетике, так как модель Шеннона является синтаксической и предназначена для систем связи.

Прежде чем перейти к изложению материала параграфа, учитывая известную неустойчивость в философском толковании понятия информации, уточним свою позицию в этом вопросе.

Во-первых, под информацией будем понимать особую форму взаимосвязи объектов, обусловленную всеобщим законом отражения материи. Такое понимание предполагает объективный характер существования информации как атрибута материи.

Во-вторых, при формализации понятия информации будем исходить из того, что сложившаяся на сегодняшний день практика математического описания информационных свойств предметных областей знаний, отражаемых в базах данных и базах знаний автоматизированных информационных систем, не учитывает, как правило, всей сложности системного строения предметных областей знаний. По-видимому, учесть эту сложность в полном объеме, оставаясь в рамках традиционно используемого для описания свойств математических объектов теоретико-множественного подхода, вообще не представляется возможным.

В качестве альтернативы по отношению к теоретико-множественному подходу может быть предложен подход, опирающийся на концепции математической теории категорий [47, 53].

4.11.1. Концептуально-методологические основы категориальной теории квантовой информатики

Сформулируем главные принципы, образующие концептуальную основу теории квантовой информатики. Изложение принципов будем по мере необходимости сопровождать содержательным комментарием (рис. 4.137).

Принцип 1 (информативности). Всякий объект образует предметную область знаний (предмет познания) некоторого субъекта и является для этого субъекта носителем информации — информационным объектом. Информационный объект определяет все свойства объекта как предметной области знаний субъекта.

Рис. 4.137. Концептуальные принципы категориальной теории квантовой информации

Этот принцип подчеркивает информационный характер всякого процесса познания и определяет знания субъекта через информацию о свойствах объекта.

Принцип 2 (представления). Для всякого информационного объекта существует потенциально бесконечное множество различных представлений — образов (моделей) объекта, отраженных в субъекте. Каждое представление однозначно определяет совокупность соответствующих этому представлению свойств информационного объекта.

Постулируемое здесь утверждение следует из материалистического учения о ступеньках познания, сменяющих друг друга в познавательном процессе. Представления, т. е. образы, отраженные в субъекте (например, в сознании отдельных людей, социальных групп или обществе в целом) и отнесенные к различным фиксированным отрезкам времени, являются аналогом указанных ступенек. Принцип позволяет свести динамику процесса познания свойств информационного объекта субъектом к получению и накоплению информации об этом объекте на отдельных ступеньках познавательного процесса.

Принцип 3 (формализации). Всякое представление информационного объекта допускает формализацию объектом подходящей (приемлемой с точки зрения содержательного описания представления) категории, именуемой категорией представления объекта.

Данный принцип требует более подробного разъяснения. Прежде всего, надо заметить, что термин «категория» понимается здесь в смысле определения, принятого в математической теории категорий [46, 47]. Согласно этой теории категория K — это пара $\langle obK, \mathrm{Mor}K\rangle$, где obK — класс, элементы которого называются объектами категории K, а $\mathrm{Mor}\,K$ — класс, элементы которого именуются морфизмами категории K. По определению объекты и морфизмы категории связаны между собой следующими условиями:

1) каждой упорядоченной паре a, $b \in obK$ сопоставлено некоторое множество $\mathrm{Mor}_{(a,b)}K$ морфизмов категории K;

2) каждый морфизм категории K принадлежит одному и только одному из множеств $\mathrm{Mor}_{(a,b)}K$;

3) в классе $\mathrm{Mor}K$ определена ассоциативная операция произведения: $\forall a, b, c, d \in obK$, $\forall \alpha \in \mathrm{Mor}_{(a,b)}K$, $\forall \beta \in \mathrm{Mor}_{(c,d)}K$, произведение $\alpha\beta$ определено тогда и только тогда, когда $b = c$, и в этом случае $\alpha\beta \in \mathrm{Mor}_{(a,d)}K$;

4) в каждом множестве $\mathrm{Mor}_{(a,a)}K$ содержится единственный морфизм $1a$, определенный условиями $\forall x, y \in obK; \forall \alpha \in \mathrm{Mor}_{(x,a)}K$; $\forall \beta \in \mathrm{Mor}_{(a,y)}K$; $\alpha 1_a = \alpha$ и $1_a \beta = \beta$.

Если $\alpha \in \mathrm{Mor}_{(a,b)}K$, то пишут $\alpha : a \to b$ и говорят, что α — морфизм a в b с началом в a и концом в b.

Примером категории может служить класс всех возможных множеств вместе со всеми возможными отображениями этих множеств друг в друге (категория множеств).

На первый взгляд может показаться, что допущение, постулируемое принципом формализации, является неоправданно жестким. Однако формализация представления информационного объекта в итоге (т. е. когда отдельные ее этапы, например, этап структуризации, мыслятся пройденными) предполагает возможность математического описания представления. С этой точки зрения категориальное представление влечет минимум искажений, неизбежно порождаемых процессом формализации, и обладает наибольшей общностью. В методологическом отношении полезно также обратить внимание на следующее. Несмотря на внешнее различие философского и математического определений категории, сам по себе принцип категориального представления информационного объекта предопределяет много общего в понимании внутреннего механизма формирования представления о предмете познания с пониманием, развиваемым в теории познания. В данном случае, как и в теории познания, формирование представления информационного объекта осуществляется одновременно с формированием категории представления объекта субъектом, при этом для фиксированного представления все его свойства оказываются однозначно определенными категорией представления.

Для практических приложений в информатике достаточно удобным средством формализации представлений информационных объектов являются различные подкатегории категории моделей подходящей сигнатуры [53—58]. Как известно, категория L называется подкатегорией категории K, если: а) $\forall a\ a \in obL \Rightarrow a \in obK$; б) $\forall \alpha\ \alpha \in \mathrm{Mor}L \Rightarrow \alpha \in \mathrm{Mor}K$; в) единичные морфизмы L категории L являются единичными морфизмами категории K; г) произведение $\alpha\beta$ морфизмов в категории L совпадает с произведением этих же морфизмов в категории K.

В категории $K=\langle obK,\ \mathrm{Mor}K\rangle$ моделей сигнатуры Σ класс obK образован всеми возможными системами с отношениями (реляционными системами) вида $M_A=\langle A,\ \theta_A\rangle$, где A — основное множество системы; $\theta_A=\{r_i^{(l_i)}\}_{i\in I}$ — множество l_i-арных отношений на A; I — индексирующее множество, при этом каждой системе определено множество $\Sigma=\{P_i^{(l_i)}\}_{i\in I}$ имен l_i-местных предикатов $P_i^{(l_i)}(x_1,\ \ldots,\ x_{l_i})$, называемое сигнатурой M_A, и взаимно однозначное отображение $\varphi_A:\Sigma\to\theta_A$, для которого выполняются условия

$$\forall i\in I\ \varphi_A\left(P_i^{(l_i)}\right)=r_i^{(l_i)};$$

$$\forall i\in I,\qquad \forall x_1,\ .\ .\ .\ .\ ,\ x_{l_i}\in A;\qquad P_i^{(l_i)}(x_1,\ .\ .\ .\ .\ ,\ x_{l_i})=1\Leftrightarrow$$

$$\Leftrightarrow\langle x_1,\ .\ .\ .\ .\ ,\ x_{l_i}\rangle\in r_i^{(l_i)}.$$

Указанная система M_A носит название модели сигнатуры Σ.

Класс $\mathrm{Mor}K$ образован всевозможными частичными гомоморфизмами таких моделей друг в друга, при этом каждый частичный гомоморфизм $f\in\mathrm{Mor}K$ модели M_A в модель $M_B=\langle B,\ \theta_B\rangle$, $\theta_B=\{\rho_i^{(l_i)}\}_{i\in I}$ той же сигнатуры есть гомоморфизм некоторой ее подмодели $\widehat{M}_A=\langle\widehat{A},\ \widehat{\theta}_A\rangle$, где $\widehat{A}\subseteq A$ и $\widehat{\theta}_A=\{\widehat{r}_i^{(l_i)}\}_{i\in I}$ — множество ограничений отношений из θ_A на множестве $\widehat{A}$, определенное отображением $f:\widehat{A}\to B$, удовлетворяющим условию

$$\forall i\in I,\qquad \forall x_1,\ .\ .\ .\ .\ ,\ x_{l_i}\in\widehat{A}\langle x_1,\ .\ .\ .\ .\ ,\ x_{l_i}\rangle\in$$

$$\in r_i^{(l_i)}\Rightarrow\langle f(x_1),\ .\ .\ .\ .\ ,\ f(x_{l_i})\rangle\in\rho_i^{(l_i)}.$$

Подмодель $\widehat{M}_A$ модели M_A называется областью определения частичного гомоморфизма f.

Практическое удобство представлений информационных объектов объектами различных подкатегорий категории моделей подходящей сигнатуры заключается в том, что, с одной стороны, эти представления имеют однозначно определенную структуру (структуру реляционной системы), с другой — имеются языковые средства, позволяющие описывать свойства структуры с помощью истинных формул логики предикатов сигнатуры модели. Вместе с тем сохраняется возможность перехода к представлению информационного объекта в любой другой категории с помощью соответствующих функторов [46, 47].

Рассмотрим пример. Пусть информационным объектом является группа лиц, подвергающихся анкетированию. Предположим, что анкетные данные заносятся в таблицу, строки которой соответствуют конкретным лицам, а столбцы — характеристикам этих лиц, взятым из словаря признаков $Q=\{q_\xi,\ \ \xi=\overline{1,\ n}\}$. Признаками могут быть фамилия, имя, отчество лица, дата его рождения, место работы и т. д. Заполнение таблицы осуществляется данными, образующими в совокупности словарь данных (значений признаков) $D=\bigcup\limits_{\xi=1}^{n}D_\xi,\ \ D_\xi=\{d_{\xi j},\ \ j=\overline{1,\ m_\xi}\}$, где $d_{\xi j}$ — j-е значение ξ-го признака. Так, например, значениями признака «фамилия» являются фамилии конкретных людей: Иванов, Петров и т. д.

Заполненная таблица образует некоторое представление информационного объекта, в данном случае — представление об анкетируемой группе лиц. Оно может быть описано формально моделью $M_A=\langle A,\ \ \theta_A\rangle$, где A — множество лиц указанной группы; $\theta_A=\{r_{\xi j},\ \xi=\overline{1,\ n},\ \ j=\overline{1,\ m_\xi}\}$ — множество отношений на A (в данном примере — свойств элементов из A), где $r_{\xi j}=\{x, g_\xi(x)=d_{\xi j}\}_{x\in A}$ и $g_\xi(x):A\to D_\xi$ — отображение, определяемое таблицей. Сигнатура данной модели описывается множеством $\Sigma=\{P_{\xi j},\ \xi=\overline{1,\ n},\ j=\overline{1,\ m_\xi}\}$ имен одноместных предикатов $P_{\xi j}(x)$, определенных условием

$$\forall x\in A,\qquad \forall\xi=\overline{1,\ n},\qquad \forall j=\overline{1,\ \ m_\xi},$$

$$P_{\xi j}(x)=1\Leftrightarrow x\in r_{\xi j}.$$

Примером такого предиката может служить предикат «Фамилия лица X есть Иванов», принимающий значение единицы (истинно) для тех и только тех лиц, фамилия которых Иванов.

Описанное представление информационного объекта однозначно для указанных в примере условий. Однако могут быть и другие представления. Так, например, расширение словарей признаков и данных изменяет представление.

Прежде чем перейти к рассмотрению следующего принципа, введем понятие уровня представления информационного объекта. Напомним, что в любой категории морфизм $\alpha:a\to b$ называется изоморфизмом, если существует морфизм $\beta:b\to a$, для которого $\alpha\beta=1_a$ и $\beta\alpha=1_b$. Всякие два объекта, для которых существует изоморфизм, называются изоморфными. Применительно к категории моделей изоморфизм модели $M_A=\langle A,\ \theta_A\rangle$, $\theta_A=\{r_i^{(l_i)}\}_{i\in I}$ на модель $M_B=\langle B,\ \theta_B\rangle$, $\theta_B=\{\rho_i^{(l_i)}\}_{i\in I}$ той же сигнатуры является частным случаем гомоморфизма и определяется взаимно однозначным отображением $f:A\to B$, удовлетворяющим условию

$$\forall i\in I,\qquad \forall x_1,\ .\ .\ .\ .\ ,\ x_{l_i}\in A\langle x_1,\ .\ .\ .\ .\ ,\ x_{l_i}\rangle\in r_i^{(l_i)}\Leftrightarrow$$

$$\Leftrightarrow\langle f(x_1),\ .\ .\ .\ .\ ,\ f(x_{l_i})\rangle\in\rho_i^{(l_i)}.$$

Известно [55], что в любой категории изоморфные объекты образуют непересекающиеся классы.

Пусть дано представление информационного объекта объектом подходящей категории. Тогда класс изоморфных объектов, содержащий это представление, будем именовать уровнем представления информационного объекта.

Принцип 4 (неразличимости). Категория представления информационного объекта однозначно определяет уровень этого представления. Любые два представления одного и того же уровня информационно неразличимы.

Этот принцип является следствием категориального подхода к описанию свойств информационного объекта. С позиции теории категорий изоморфные объекты неразличимы по свойствам.

В качестве иллюстрации обратимся к рассмотренному выше примеру. Легко видеть, что любые модели, получаемые из исходной переупорядочиванием строк или столбцов таблицы, изоморфны и, следовательно, определяют один и тот же уровень представления об анкетируемой группе лиц.

Полезно отметить, что естественное понятие уровня представления информационного объекта имеет формальную теоретико-групповую интерпретацию. Рассмотрим ее на примере моделей.

Пусть M — множество всех моделей сигнатуры Σ с одним и тем же основным множеством A и Γ — множество всех изоморфизмов указанных моделей друг на друга. Тогда Γ образует группу, единицей которой является тождественное отображение множества A на себя [53]. Обозначим через $g(M_A)$, $g\in\Gamma$, $M_A\in M$ образ модели M_A при изоморфизме g (ясно, что образом M_A является некоторая модель $M_A'\in M$). В соответствии с терминологией, принятой в теории групп, множество $Y=\{g(M_A),\ g\in\Gamma\}$ называется орбитой M_A (в группе Γ). Так как Y — в точности класс моделей, изоморфных M, то понятие уровня представления информационного объекта моделью M_A совпадает с понятием орбиты M_A.

Изоморфизм модели на себя принято называть [55] ее автоморфизмом. Множество всех автоморфизмов модели M_A также образует группу. Обозначим ее через Γ_A. Эта группа называется группой автоморфизмов M_A и является подгруппой группы Γ. Теория групп позволяет вычислить мощность орбиты по известной формуле [44]: $n_\nu=n_\Gamma/n_A$, где n_ν — мощность орбиты и n_Γ, n_A — порядки групп Γ и Γ_A соответственно.

Принцип 5 (системности). Всякое представление информационного объекта имеет системное строение, однозначно определенное категорией этого представления.

Для разъяснения этого принципа напомним сначала, что независимо от категории представления информационного объекта всякий объект категории имеет системное строение, однозначно определенное структурой его подобъектов. Понятие подобъекта в теории категорий вводится следующим образом [47]. Пусть дана категория $K=\langle obK,\ \mathrm{Mor}K\rangle$ и $a,\ p\in obK$. По определению морфизм $\mu:p\to a$, $\mu\in\mathrm{Mor}K$ называется мономорфизмом, если для любых двух морфизмов $\alpha,\ \beta:c\to p$ категории K выполняется условие $\alpha\mu=\beta\mu\Rightarrow\alpha=\beta$. (Применительно к категории моделей мономорфизмом μ является всякий инъективный, т. е. удовлетворяющий условию $\forall x,\ y\in A\ \ \mu(x)=\mu(y)\Rightarrow x=y$, гомоморфизм $\mu:A\to B$ модели M_A в модель M_B.) Рассмотрим класс всех пар $(p,\ \mu)$, где $p\in obK$ и $\mu:p\to a$ — мономорфизм категории K. Говорят, что пары $(p_1,\ \mu_1)$ и $(p_2,\ \mu_2)$ указанного вида эквивалентны, если $\mu_1=\varepsilon\mu_2$ для некоторого изоморфизма ε. Легко видеть, что отношение, установленное между указанными парами, является отношением эквивалентности. Поэтому класс всех указанных пар разбивается на непересекающиеся классы эквивалентных пар. Каждый класс эквивалентных пар вида $(p,\ \mu)$ обозначается $[p,\ \mu]$ и называется подобъектом объекта a. Считают, что подобъект $[g,\ \nu]$ содержится в подобъекте $[p,\ \mu]$ (меньше $[p,\ \mu]$), и пишут $[g,\ \nu]\leqslant[p,\ \mu]$, если $\nu=\tau\mu$ для некоторого морфизма τ. Доказано, что отношение $\leqslant$, именуемое «включением», является отношением частичного порядка.

Класс всех объектов p из $[p,\ \mu]$ можно по аналогии с приведенным выше определением уровня представления информационного объекта назвать уровнем представления некоторого подобъекта информационного объекта. Однако при этом надо иметь в виду, что использование термина «подобъект» применительно к информационному объекту является условным. Определение подобъекта информационного объекта вообще вызывает методологические трудности, так как сам информационный объект (вне зависимости от его представления) определен не более как носитель информации. Тем не менее каждый представитель $(p,\ \mu)$ из $[p,\ \mu]$ дает однозначное представление p о некоторой части информационного объекта, если информационный объект в целом представлен объектом a. Например, если информационный объект представлен моделью $M_A=\langle A,\ \theta_A\rangle$, то представлением его подобъекта может служить подмодель $\hat M_A=\langle\hat A,\ \hat\theta_A\rangle$ модели M_A, где $\hat A\subseteq A$, $\hat\theta_A=\{\hat r_i^{(l_i)}\}_{i\in I}$ и $\hat r_i^{(l_i)}$ — ограничения $r_i^{(l_i)}$ из $\theta_A=\{r_i^{(l_i)}\}_{i\in I}$ на множестве $\hat A$. Множество всех моделей с основным множеством $\hat A$ изоморфных $\hat M_A$ образует уровень представления указанного подобъекта. В частности, если в качестве информационного объекта рассматривается описанная выше анкетируемая группа лиц, представленная моделью M_A, то модель $\hat M_A$ соответствует представлению о части информационного объекта определенной группой лиц, входящих в $\hat A$.

Для характеристики системных свойств информационного объекта введем аксиоматическое понятие системного типа этого объекта.

Аксиома (о системном типе информационного объекта). Всякое представление информационного объекта объектом подходящей категории имеет единственный, определенный для этого представления, системный тип. Любые два представления имеют один и тот же системный тип тогда и только тогда, когда они изоморфны.

Эта аксиома является обобщением известной аксиомы теории множеств о реляционном типе системы с отношениями [44]. Из аксиомы следует, что системный тип информационного объекта однозначно определен категорией его представления. Заметим, что понятия системного типа и уровня представления информационного объекта не совпадают, хотя между ними имеется взаимно однозначное соответствие.

В теории квантовой информатики проблема описания системного типа информационного объекта является важной и вместе с тем достаточно сложной проблемой, так как она сводится по-существу к известной проблеме распознавания изоморфных объектов.

Применительно к категориям конечных моделей (т. е. моделей с конечным основным множеством) это описание может быть получено путем измерения объектов заданной категории в подходящей шкале [59]. Можно доказать, что если $K=\langle obK, \mathrm{Mor}K$ — такая категория, то приемлемой шкалой измерения является множество S всех кортежей натуральных чисел вида $\langle n_1, \ldots, n_l\rangle$ подходящей длины l, в которое определено отображение T: $obK\rightarrow S$, удовлетворяющее условию $\forall a$, $b\in obK$ $T(a)=T(b)\Leftrightarrow a\sim b$, где $\sim$ — обозначение изоморфизма объектов (равенство кортежей определяется покоординатно). Обоснование этого утверждения весьма громоздко и не может быть проведено в данной работе. Однако суть указанного подхода можно проиллюстрировать на простом примере.

Пусть дано множество M всех конечных моделей вида $M_A=\langle A, r\rangle$ одной и той же сигнатуры $\Sigma=\{p\}$, где p — имя одноместного предиката $p(x)$ и $\forall x\in Ax\in r\Leftrightarrow p(x)=1$, а также множество S всех упорядоченных пар натуральных чисёл. Построим отображение $T:M\rightarrow S$, положив по определению

$$\forall \langle n_1,\ n_2\rangle \in ST\ (M_A)=\langle n_1,\ n_2\rangle \Leftrightarrow n_1=\mathrm{Card}\,A\ \&\ n_2=\mathrm{Card}\,r,$$

где $\mathrm{Card}\,A$, $\mathrm{Card}\,r$ — мощности множества A и r соответственно и $\&$ — обозначение операции конъюнкции. Легко убедиться, что $\forall M_A$, $M_B\in MT(M_A)=T(M_B)\leftrightarrow M_A\sim M_B$. Следовательно, для всякой M_A из M $T(M_A)=\langle n_1, n_2\rangle$ описывает системный тип модели M_A.

Заметим, что использованная в примере шкала измерения относится к классу так называемых номинальных шкал. Известно [59], что такие шкалы определены с точностью до взаимно однозначного отображения. Поэтому и всякое описание системного типа объекта в номинальной шкале определено с точностью до взаимно однозначного отображения. Так, например, системный тип рассмотренной в примере модели M_A с равным успехом может быть описан в шкале, определенной отображением $T':M\rightarrow S$, где $\forall\langle n_1,\ n_2\rangle\in S$, $T'(M_A)=\langle n_1,\ n_2\rangle\Leftrightarrow n_1=\mathrm{Card}\,A\ \&\ n_2=\mathrm{Card}\,A\backslash r$ ($A\backslash r$ обозначает дополнение r до A). Ясно, что при сравнении системных типов объектов их описания должны осуществляться в одной и той же шкале.

В квантовой информатике отображение $T:ob\,K\rightarrow S$ называется системной характеристикой категории представления объекта в шкале S, кортежи $\langle n_1, \ldots, n_2\rangle$ — называются спецификациями объектов, а элементы спецификаций — системными числами.

С методологической точки зрения полезно отметить, что содержательно системные числа, описывающие представление информационного объекта, несут ту же смысловую нагрузку, что и квантовые числа, описывающие состояние квантового объекта в физике микрочастиц. Вообще говоря, при изучении информационных объектов с помощью квантовой информатики обнаруживается много сходств в описании их свойств с описанием свойств квантовых объектов. По мнению автора, это явление не случайно. Напротив, оно порождено общей методологией системного подхода и косвенно подтверждает концептуальную общность в изучении свойств объектов как носителей информации независимо от того, является информационным объектом микрочастица или какой-либо другой объект.

Применительно к категории конечных моделей теоретически можно построить шкалу, в которой системный тип объекта взаимно однозначно определен некоторым натуральным числом (номером объекта). Для этого можно использовать любое взаимно однозначное отображение $f:N^l\rightarrow N$, где N — множество натуральных чисел. Примером приемлемого отображения может служить канторовская функция нумерации l-к (кортежей из l элементов), определенная формулами. $\forall x_i\in N$, $i\in N$,

$$C\ (x_1,\ x_2)=\frac{(x_1+x_2)^2+3x_1+x_2}{2};\qquad C^3\ (x_1,\ x_2,\ x_3)=$$

$$=C\ (C\ (x_1,\ x_2),\ x_3),\ .\ .\ .\ .\ ,\ C^l\ (x_1,\ x_2,\ .\ .\ .\ .\ ,\ x_{l-1},\ x_l)=$$

$$=C^l\ (C^{l-1}\ (x_1,\ x_2,\ .\ .\ .\ .\ ,\ x_{l-1}),\ x_l).$$

По содержанию $C(x_1, x_2)$ — это номер пары $\langle x_1, x_2\rangle$ натуральных чисел упорядоченной последовательности таких пар. В данной последовательности пары идут в порядке возрастания суммы их членов, а из пар с одинаковой суммой членов первой является пара с меньшим первым членом. Таким образом, $C(0,\ 0)=0$, $C(0,\ 1)=1$, $C(1,\ 0)=2$ и т. д. Указанную шкалу можно именовать шкалой Кантора.

Следующий принцип уточняет содержание используемого в квантовой информатике понятия информации с учетом ее семантических свойств.

Принцип 6 (квантования). Всякая информация есть непустой класс отдельно взятых квантов (элементарных порций) информации. Каждый квант информации — это морфизм представления некоторого информационного объекта объектом подходящей категории в любой другой объект той же категории.

Легко видеть, что класс всех квантов информации, носителем которой является информационный объект при его фиксированном представлении объектом a подходящей категории, однозначно определен категорией этого представления и является классом всех морфизмов с началом в объекте a. Этот класс назовем потенциальной информацией.

С точки зрения содержания потенциальная информация описывает весь запас информации о всех возможных образах информационного объекта, которые могут существовать в субъекте при данном представлении. Каждый из этих образов однозначно определен своим квантом информации. Если, например, субъектом является человек, то это значит, что он «умеет мыслить» категорией представления информационного объекта, т. е. в его сознании имеется отраженный образ данной категории, включая образы тех объектов категории, в которые определены морфизмы с началом в объекте-представлении.

Для иллюстрации принципа 6 рассмотрим вновь информационный объект, заданный анкетируемой группой лиц и представленный выше моделью $M_A=\langle A, \theta_A\rangle$, $\theta_A=\{r_{\xi j}, \xi=\overline{1, n}, j=\overline{1, m_\xi}\}$ сигнатуры Σ. Пусть мы располагаем таблицей анкетных данных на множество лиц B, составленной по той же форме и с использованием тех же словарей, что и для группы лиц A. Эта таблица может содержаться, например, в памяти ЭВМ. Формально таблицу можно представить моделью $M_B=\langle B, \theta_B\rangle$ той же сигнатуры Σ, где $\theta_B=\{\rho_{\xi j}, \xi=\overline{1, n}, j=\overline{1, m_\xi}\}$ — множество отношений на B; $\rho_{\xi j}=\{x, g_\xi(x)=d_{\xi j}\}$ и $g_\xi(x):B\rightarrow D_\xi$ — отображение, определенное таблицей.

Рассмотрим три случая, связывая в каждом из них с точки зрения содержания существование в машине той или иной информации об объекте, представленном моделью M_A, с существованием некоторого частичного гомоморфизма модели M_A в модель M_B, при этом будем предполагать, что среди лиц множества $A\cup B$ «двойников», т. е. различных людей с одинаковыми анкетными данными, нет.

1. $A\cap B=\varnothing$. Информация в машине отсутствует. Частотных гомоморфизмов модели M_A в модель M_B тоже нет.

2. $A\cap B\neq 0$, $A\subseteq B$. Информация в машине есть. Существует частичный гомоморфизм модели M_A в модель M, в данном случае — всюду определенный тождественный изоморфизм $f:A\rightarrow B$ $(\forall x\in A, \forall\xi=\overline{1, n}, \forall j=\overline{1, m_\xi}, x\in r_{\xi j}\Leftrightarrow f(x)\in\rho_{\xi j})$.

3. $A\cap B=A\neq\varnothing$, $A\not\subseteq B$. В машине имеется частичная информация о группе лиц A, а именно о всех лицах из $\hat{A}\subseteq A$. Существует частичный гомоморфизм модели M_A в модель M_B с областью определения $\hat{M}_A=\langle\hat{A}, \hat{\theta}_A\rangle$, где $\hat{\theta}_A=\{\hat{r}_{\xi j}, \xi=\overline{1, n}, j=\overline{1, m_\xi}\}$ и $\hat{r}_{\xi j}$ — ограничения на A. В данном случае этим частичным гомоморфизмом является тождественный изоморфизм $\varphi:\hat{A}\rightarrow B$ подмодели $\hat{M}_A$ модели M_A $\{\forall x\in\hat{A}, \forall\xi=\overline{1, n}, \forall j=\overline{1, m_\xi}\ x\in\hat{r}_{\xi j}\Leftrightarrow\varphi(x)\in\rho_{\xi j}\}$ в модели M_B.

С методологической точки зрения полезно отметить, что любой квант информации является одновременно информацией и информационным объектом. Первое его качество по содержанию связано со свойствами кванта, характеризующими присущую ему функцию отражения объекта в субъекте (определяется согласно принципу 6 свойствами морфизма категории представления информационного объекта), второе качество связано со свойствами кванта, характеризующими функцию его как носители информации (определяется согласно принципам 1—3 свойствами объекта подходящей категории представления кванта информации). Например, квант информации, определенный гомоморфизмом $f:A\rightarrow B$ модели M_A в модель M_B, может быть представлен моделью $M_D=\langle D, R_f\rangle$ сигнатуры $\Sigma=\{P_f\}$, где $D=A\times B$, $R_f\subseteq A\times B$ и $\forall x\in A$, $\forall y\in B\langle x, y\rangle\in R_f\Leftrightarrow x\in A\,\&\,f(x)=y\in B$, $P_f(x, y)=1\Leftrightarrow\langle x, y\rangle\in R_f$.

Как известно, у квантовых объектов в физике микрочастиц проявляются аналогичные свойства, известные под названием корпускулярно-волнового дуализма [например, корпускулярно-волновая природа кванта света (фотона), методологические аспекты которой до сих пор обсуждаются на страницах журналов по философским проблемам естествознания]. Это дополнительно свидетельствует в пользу ранее высказанного соображения об общности развиваемых концепций.

Принцип 7 (полноты). Всякий квант информации является полным тогда и только тогда, когда он определен мономорфизмом.

По содержанию это означает, что если a — представление информационного объекта объектом подходящей категории и $\mu:a\rightarrow b$ — мономорфизмом этой категории, то объект b полностью (без потерь) отражает все свойства информационного объекта при его представлении объектом a. (Возможность существования дополнительно каких-либо свойств у объекта b при этом не исключается). В категории моделей всякий мономорфизм $f:A\rightarrow B$ модели $M_A=\langle A, \theta_A\rangle$ в модель $M_B=\langle B, \theta_B\rangle$ есть инъективный гомоморфизм. Поэтому для любого свойства p информационного объекта, представленного моделью M_A и определенного предикатом $p(x_1, \ldots, x_n)$, имеет место $\forall x_1, \ldots, x_n \in A$,

$$p(x_1, \ldots, x_n) = 1 \Rightarrow p(f(x_1), \ldots, f(x_n)) = 1.$$

В приведенной выше иллюстрации предыдущего принципа квант информации, описанный во втором случае, является полным, а в третьем — неполным.

Перед тем как сформулировать следующий принцип, напомним, что в теории категорий морфизм $\alpha: a \to b$ всякой категории называется обратимым справа, если существует морфизм $\beta: b \to a$ той же категории, такой, что $\alpha\beta = 1_a$.

Принцип 8 (точности). Образом, определяющим квант морфизма, является ретракт, изоморфный некоторому ретракту представления информационного объекта.

Точный квант информации — это такой квант, который точно (без искажений) отражает свойства информационного объекта, представленного объектом подходящей категории.

Заметим, что кванты информации могут быть неполными и неточными, полными, но неточными, неполными, но точными и, наконец, полными и точными. В приведенной выше иллюстрации принципа 6 точными являются оба кванта (случаи 2 и 3). В то же время полным и точным квантом информации является лишь квант, описанный в случае 2.

Рассмотрим условия полноты и точности кванта информации. Пусть квант информации определен морфизмом $\alpha: a \to b$ и существует образ $[p, \mu]$ объекта a при морфизме α, т. е. $\alpha = \pi\mu$ для некоторого нормального эпиморфизма π и мономорфизма μ. В этом случае квант информации является полным и точным тогда и только тогда, когда α — мономорфизм, обратимый справа.

Действительно, предположим, что квант информации является полным и точным. Тогда согласно принципу 7 α является мономорфизмом, а поскольку $\alpha = \pi\mu$, то π — тоже мономорфизм. С другой стороны, согласно принципу 8 существует подобъект $[q, \nu]$ объекта a, такой, что $\mu t \nu \pi = 1_p$ для некоторого морфизма t. Но в этом случае $\pi = \pi 1_p = \pi\mu t \nu \pi$, и поскольку π — мономорфизм, то имеет место $1_a \pi = \pi\mu t \nu \pi \Rightarrow \pi\mu t \nu = 1_a$ или, учитывая, что $\alpha = \pi\mu$, $\alpha(t\nu) = 1_a$, и, следовательно, мономорфизм α обратим справа.

Наоборот, пусть α — обратимый справа мономорфизм. Это значит, что существует морфизм β, такой, что $\alpha\beta = 1_a$. Согласно принципу 7 квант информации, определенный α, является полным. Покажем, что он точен. Рассмотрим ретракт $[a, 1_a]$ объекта a. Так как $\alpha = \pi\mu$, то $1_a \pi\mu\beta = 1_a$. Кроме того, $\pi = 1_a \pi = \pi\mu\beta\pi$. Учитывая, что π — эпиморфизм, получаем

$$\pi\mu\beta\pi = \pi 1_p \Rightarrow \mu\beta\pi = 1_p.$$

Следовательно, образ $[p, \mu]$ является ретрактом объекта b. Кроме того, $\mu\beta 1_a \pi = 1_p$. Таким образом, ретракт $[a, 1_a]$ изоморфен ретракту $[p, \mu]$ — образу объекта a при морфизме α, что и требовалось показать.

В порядке следствия из этого утверждения получаем, что квант информации, определенный морфизмом α, является полным и точным тогда и только тогда, когда нормальный эпиморфизм π, формирующий образ $[p, \mu]$ объекта a, при морфизме a является изоморфизмом. Это значит, что $[a, \alpha] = [p, \mu]$. В категории моделей это соответствует случаю, когда квант информации, определенный частичным гомоморфизмом f модели M_a представления информационного объекта в модель M_b, является изоморфизмом M_a в M_b (т. е. M_a является изоморфным вложением в M_b) [55].

Следующий принцип имеет целью сформулировать концептуально-методологическую основу для ответа на вопрос: сколько информации об объекте отражает в субъекте каждый отдельно взятый ее квант?

Ясно, что любая формализация понятия семантической информации малопригодна для практического использования, если она не позволяет производить количественную оценку информации.

Назовем системным типом произвольно взятого подобъекта $[c, \mu]$ объекта b системный тип объекта c. Так как всякий объект c' любого другого представителя (c', μ') подобъекта $[c, \mu]$ изоморфен (по определению подобъекта) объекту c, то он (объект c') согласно аксиоме о системном типе информационного объекта имеет одинаковый с объектом c системный тип. Поэтому данное определение системного типа подобъекта корректно относительно указанной аксиомы.

Принцип 9 (объемности). Каждый квант информации обладает единственным, присущим этому кванту объемом информации. Если образ объекта при морфизме, определяющем квант информации, существует, то объем такого кванта взаимно однозначно определен системным типом указанного образа.

Данный принцип сводит проблему количественной оценки объема произвольно взятого кванта информации, определенного морфизмом $\alpha: a \to b$, к количественной оценке системного типа объекта c образа $[c, \mu]$ объекта a при морфизме α. В силу того что в любой категории каждый морфизм имеет один и только один образ [53—68], объем каждого кванта информации определен единственным образом.

Обозначим объем кванта информации, определенного морфизмом α, через $V(\alpha)$ и положим по определению, что $V(\alpha) \leqslant V(\beta) \Leftrightarrow [c, \mu] \leqslant [d, V]$, где $[d, V]$ — образ морфизма $\beta: a \to b$. Тогда можно констатировать, что частичная упорядоченность по включению подобъектов объекта b индуцирует частичную упорядоченность квантов информации, определенных морфизмами с концом в b по их объемам.

Для иллюстрации принципа объемности воспользуемся иллюстрацией принципа 6. Чтобы избежать громоздких рассуждений, предположим, что словарь признаков, который используется для описания лиц, образующих анкетируемую группу A и множество B таблицы, находящейся в памяти ЭВМ, содержит единственный признак — фамилию лица, а словарь значений признаков — единственную фамилию — Иванов. Предположим, что оба множества A и B содержат, по крайней мере, по одному лицу с фамилией Иванов, а также другие лица.

В этом случае информационный объект может быть представлен моделью $M_A=\langle A, r\rangle$ сигнатуры $\Sigma=\{p\}$, где p — имя предиката $p(x)$: «фамилия лица X есть Иванов» и $\forall x\in A, p(x)=1\Leftrightarrow x\in r\neq\varnothing$, $\mathrm{Card}\,A>\mathrm{Card}\,r$. Содержащаяся в памяти ЭВМ таблица, в свою очередь, может быть представлена моделью $M_B=\langle B, \rho\rangle$ той же сигнатуры Σ, причем в этом случае $P(x)=1\Leftrightarrow x\in\rho\neq\varnothing$, $x\in B$, $\mathrm{Card}\,B>\mathrm{Card}\,\rho$.

Пусть $\widehat{M}_B=\langle\widehat{B}, b_1\rangle$ — подмодель модели M_B, где $\widehat{B}=\{b_1, b_2\}$, $p(b_1)=1$ и $b_1, b_2\in B$ (т. е. лицо b_1 есть Иванов, а лицо b_2 может носить как фамилию Иванов, так и другую фамилию). Построим отображение $f:A\rightarrow B$, положив по определению $\forall x\in r, f(x)=b_1$ и $\forall x\in A\setminus r, f(x)=b_2$. Тогда, как легко видеть, f — гомоморфизм модели M_A в модель M_B. Квант информации, определенный гомоморфизмом f, является информацией о том, что в анкетируемой группе лиц имеются лица по фамилии Иванов, а также лица с другими фамилиями.

Рассмотрим класс изоморфных моделей вида $M_D=\langle D, d_1\rangle$ сигнатуры Σ, где $D=\{d_1, d_2\}$, $p(d_1)=1$, $p(d_2)=0$. Очевидно, что для любой модели этого класса гомоморфизм f может быть описан композицией отображений $f=\pi\mu$, где $\pi:A\rightarrow D$, $\forall x\in r$, $\pi(x)=d_1$, $\forall x\in A\setminus r$, $\pi(x)=d_2$ и $\mu:D\rightarrow B$, $\mu(d_1)=b_1$, $\mu(d_2)=b_2$.

Нетрудно установить, что отображение π является эпиморфизмом, а отображение μ — мономорфизмом, причем, если $p(b_2)=0$, то мономорфизм является изоморфизмом модели M_D в модель M_B.

Для любой модели M_D указанного класса системный тип один и тот же и в шкале, определенной отображением $T:M\rightarrow S$, где M — множество всех конечных моделей сигнатуры Σ; S — множество всех пар $\langle n_1, n_2\rangle$ натуральных чисел и $\forall\langle n_1, n_2\rangle\in S$, $\forall M_D\in M$, $T(M_D)=\langle n_1, n_2\rangle\Leftrightarrow n_1=\mathrm{Card}\,D\,\&\,n_2=\mathrm{Card}\,d_1$, и может быть основан парой $T(M_D)=\langle 2, 1\rangle$. Следовательно, объем $V(f)$ кванта информации, определенного гомоморфизмом f, в той же шкале равен $V(f)=\langle 2, 1\rangle$. В шкале Кантора ему соответствует число $C(2, 1)=8$.

Рассмотренный в примере квант информации не является ни полным, ни точным как по смыслу, так и согласно принципам 7 и 8. По смыслу полным и точным квантом информации является информация о том, сколько лиц в группе A, а также сколько и какие из них носят фамилию Иванов. Как указано выше, квант полной и точной информации определяется изоморфизмом модели M_A в модель M_B. Так как системный тип образа M_A при изоморфизме совпадает с системным типом модели M_A, то объем полного и точного кванта информации определяется системным типом модели M_A. Легко видеть, что в указанной шкале $T(M_A)=\langle n_1, n_2\rangle$, где $n_1=\mathrm{Card}\,A$, $n_2=\mathrm{Card}\,r$. Например, если $n_1=17$, $n_2=5$, то $T(M_A)=\langle 17{,}5\rangle$. В шкале Кантора этому соответствует число $C(17{,}5)=270$.

В заключение необходимо сказать, что описанный подход к оценке объемов квантов информации не противоречит общепринятой оценке количества информации в битах. Чтобы выразить оценку количества содержащейся в кванте информации в битах, достаточно положить для произвольно взятого кванта информации α

$$V(\alpha)=\begin{cases}0, & C(\alpha)=0;\\ \log C(\alpha), & C(\alpha)>0,\end{cases}$$

где $C(\alpha)$ — объем кванта информации в шкале Кантора.

4.11.2. Оценка объемов квантовой информации на базе теории категорий

Общие соображения

Мы сформулировали девять главных принципов, составляющих концептуально-методологическую основу теории квантовой информатики. Развиваемый подход к формализованному описанию информации о свойствах предметных областей знаний (информационных объектов) опирается на использование идей математической теории категорий. Согласно предложенному подходу любая информация, носителем которой является информационный объект (или множество таких объектов), есть множество отдельно взятых квантов (элементарных порций) информации. Каждый квант информации определяется некоторым морфизмом категории представления информационного объекта.

Способ оценки объема произвольно взятого кванта информации постулируется принципом объемности. Этот принцип утверждает, что если K — категория представления информационного объекта A объектом $a\in ob\,K$, α — квант информации от A, определенный морфизмом $\alpha:a\rightarrow b$, $\alpha\in\mathrm{Mor}\,K$, и подобъект $[p, \mu]$ объекта b является образом a при морфизме α, то объем

$V(\alpha)$ кванта информации α однозначно определен системным типом объекта p.

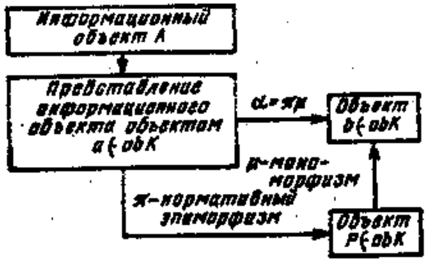


Рис. 4.138. Пояснение к понятию образа объекта *a* при морфизме α

Методологической основой выбранного подхода к оценке объема кванта информации могут служить следующие соображения.

Прежде всего, с содержательной точки зрения морфизм α реализует функцию отражения информационного объекта в субъекте в виде образа $[p, \mu]$ объекта a, представляющего информационный объект (рис. 4.138). Поэтому естественно оценить объем кванта информации, определенного морфизмом α, как объем порции информации, необходимой и достаточной для описания свойств указанного образа.

Далее можно предположить, что объем порции информации, необходимой и достаточной для описания свойств образа $[p, \mu]$ объекта a, совпадает с объемом порции информации, необходимой и достаточной для описания свойств объекта p. Предпосылкой для такого утверждения может служить следующий тезис.

Тезис объемности. В любой категории представления информационного объекта изоморфные объекты категории и только они являются носителями равной по объему информации, необходимой и достаточной для описания свойств этих объектов.

Данный тезис в принципе недоказуем, хотя и является интуитивно очевидным. Так, например, если представлением информационного объекта является некоторая таблица (матрица) данных, то любая другая таблица (матрица) данных, получаемая из первой произвольно взятой перестановкой строк и столбцов (матричный изоморфизм), в содержательном смысле несет одну и ту же информацию о свойствах информационного объекта, хотя является другим его представлением [53]. Другие примеры, подтверждающие справедливость тезиса, приводятся в тексте.

Если указанный тезис принять, то формально корректность выбранного подхода к оценке объема кванта информации обеспечивается следующим образом.

Во-первых, между морфизмом α и образом $[p, \mu]$ объекта a при этом морфизме существует однозначное соответствие. Действительно, в любой категории с нулевым объектом для любого морфизма категории с началом в произвольно взятом объекте этой категории существует не более одного образа данного объекта. При существовании образ единствен.

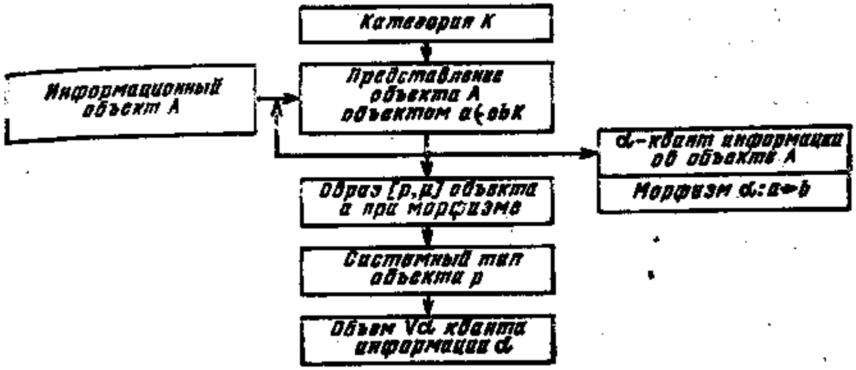


Рис. 4.139. Пояснение к понятию объема кванта информации

Во-вторых, между образом $[p, \mu]$ объекта a при морфизме α и системным типом объекта p имеется взаимно однозначное соответствие. Действительно, класс $[p]$ объектов категории, образованный из объектов всех возможных представителей вида $\langle p, \mu\rangle$ образа $[p, \mu]$,— это в точности класс объектов, изоморфных p. Поэтому согласно аксиоме о системном типе информационного объекта каждый объект из класса $[p]$ имеет один и тот же системный тип, совпадающий с системным типом объекта p. Следовательно, если предположить, что квант информации, необходимой и достаточной для описания свойств объекта p, определяется изоморфизмом с началом в объекте p, то между объемом кванта информации, определенного морфизмом, и системным типом объекта p устанавливается однозначное соответствие (рис. 4.139). Таким образом, достаточно убедиться в правомерности рассматриваемого подхода применительно к оценке объема кванта информации, определенного изоморфизмом.

Ниже описывается ряд содержательных примеров в пользу предложенного подхода. Дополнительной аргументацией служит сравнение этого подхода с известным подходом к оценке количества информации по формуле К. Шеннона.

Содержательное обоснование концепции оценки объемов информации

В целях большей простоты и наглядности рассмотрим примеры, в которых информационные объекты представлены объектами категории моделей фиксированной сигнатуры [54, 55]. Построение примеров, в которых информационные объекты представляются объектами других категорий, оставим читателю. При этом рекомендуется предварительное сведéние категории представления информационного объекта к изоморфной ей подкатегории категории моделей с помощью взаимно однозначного ковариантного функтора.

1. Пусть информационным объектом является множество A, для которого определено одно и только одно свойство — принадлежности элемента данному множеству. Тем самым определено некоторое представление о множестве A. Оно может быть описано моделью $M_A=\langle A, \theta_A\rangle$, $\theta_A=A$ сигнатуры $\Sigma=\{P\}$, где P — отмеченное свойство, которое определено предикатом $p(x)$, заданным условием

$$\forall x p(x)=1 \Leftrightarrow x \in A.$$

Предположим, что B — любое другое множество, которое может быть представлено моделью $M_B=\langle B, \theta_B\rangle$, $\theta_B=B$ той же сигнатуры $\Sigma=\{P\}$, причем $\forall y p(y)=1 \Leftrightarrow y \in B$, и что существует взаимно однозначное отображение $f: A \rightarrow B$.

Так как $\forall x\, x \in A \Leftrightarrow f(x) \in B$, то f — изоморфизм модели M_A на модель M_B, и, следовательно, модели M_A, M_B изоморфны.

Очевидно, что $A=\{x, f(x) \in B\}$. Поэтому квант информации, определенный изоморфизмом f, является порцией информации, необходимой и достаточной для описания элементного состава множества A. В то же время с точки зрения содержания описание элементного состава множества A является полным и точным описанием этого множества по условию и соответствует его представлению моделью M_A.

Указанное свойство кванта не зависит от выбора конкретного представителя из множества квантов, определенных всеми возможными изоморфизмами с началом в модели M_A. Иными словами, если другой квант определен изоморфизмом $f': A \rightarrow C$ модели M_A на модель $M_c=\langle c, \theta_c\rangle$, $\theta_c=C$ той же сигнатуры $\Sigma=\{P\}$, причем $\forall z\ p(z)=1 \Leftrightarrow z \in c$, то $A=\{x, f'(x) \in c\}$. Поэтому естественно допустить, что каждый квант, определенный изоморфизмом с началом в модели M_A, характеризуется одним и тем же объемом информации.

Как известно, любые два множества A и B имеют одну и ту же мощность тогда и только тогда, когда существует взаимно однозначное отображение $f: A \rightarrow B$. Мощность множества принято оценивать кардинальным числом. Поскольку в нашем случае отображение f является изоморфизмом модели M_A на модель M_B, то кардинальное число множества A описывает одновременно реляционный тип модели M_A [55], т. е. системный тип информационного объекта (множества A) при его представлении моделью M_A. Следовательно, между объемом кванта информации f и кардинальным числом множества A существует взаимно однозначное соответствие.

Таким образом, кардинальное число (мощность) множества A (или множества B в силу равномощности этих множеств) описывает объем указанного кванта информации.

2. Пусть информационным объектом является множество, для которого определено одно и только одно свойство — свойство линейной упорядоченности его элементов. (Здесь и далее будем полагать, что свойство, описанное в п. 1 и определяющее элементный состав множества по умолчанию, имеет место всегда.) Таким образом, о множестве A определено некоторое представление. Его можно описать моделью $M_A=\langle A, \theta_A\rangle$, $\theta_A=r \subseteq A^2$ сигнатуры $\Sigma=\{g\}$, где r — отношение линейного порядка на A [55], определенное предикатом $g(x, y)$:

$$\forall x, \quad y \in A \quad g(x, y)=1 \Leftrightarrow \langle x, y\rangle \in r.$$

Предположим, что B — любое другое линейно упорядоченное множество, представимое моделью $M_B=\langle B, \theta_B\rangle$, $\theta_B=\rho \subseteq B^2$ той же сигнатуры $\Sigma=\{g\}$, $\forall x, y \in B\ g(x, y)=1 \Leftrightarrow \langle x, y\rangle \in \rho$ и что при этом существует взаимно однозначное отображение $f: A \rightarrow B$, удовлетворяющее условию

$$\forall x, \ y \in A \quad \langle x, y\rangle \in r \Leftrightarrow \langle f(x), \ f(y)\rangle \in \rho.$$

Тогда f — изоморфизм модели M_A на модель M_B, и, значит, эти модели изоморфны.

Очевидно, что $A=\{x, f(x) \in B\}$ и $r=\{\langle x, y\rangle, \langle f(x), f(y)\rangle \in \rho\}$. Отсюда следует, что квант информации, определенный изоморфизмом f, является порцией информации, необходимой и достаточной для описания множества A вместе с отношением r линейного порядка на A. С точки зрения содержания это описание является полным и точным описанием указанного множества по условию и соответствует его представлению моделью M_A.

Как и в предыдущем примере, данное свойство кванта не зависит от выбора конкретного представителя из множества всех изоморфизмов с началом в модели M_A. Иными словами, если f' — другой квант информации, определенный изоморфизмом $f': A \rightarrow C$ модели M_A на модель $M_c=\langle c, \theta_c\rangle$, $\theta_c=l \subseteq c^2$ той же сигнатуры $\Sigma=\{g\}$ и $\forall u, v \in c$, $g(u, v)=1 \Leftrightarrow \langle u, v\rangle \in l$, то $A=\{x, f'(x) \in c\}$ и $r=\{\langle x, y\rangle, \langle f'(x), f'(y)\rangle \in l\}$. Поэтому естественно допустить, что каждый квант, определенный изоморфизмом с началом в модели M_A, характеризуется одним и тем же объемом информации.

Как известно [55], в теории множеств каждому линейно упорядоченному множеству A ставится в соответствие один и только один объект t_A, называемый порядковым типом множества A, причем по определению для любых двух множеств A и B равенство $t_A=t_B$ имеет место тогда и только тогда, когда модели M_A и M_B изоморфны. Таким образом, порядковый тип множества A описывает реляционный тип модели M_A, т. е. системный тип данного множества при его представлении моделью M_A.

Непосредственно видно, что между объемом кванта информации, определенного изоморфизмом f, и порядковым типом множества A существует взаимно однозначное соответствие. Следовательно, порядковый тип множества A (или множества B в силу совпадения порядковых типов этих двух множеств) описывает объем указанного кванта информации.

Порядковый тип вполне упорядоченного множества называется порядковым (трансфинитным) числом или ординалом. Проделав рассуждения, аналогичные приведенным выше, заключаем, что ординал всякого вполне упорядоченного множества описывает объем кванта информации, определенного изоморфизмом, с началом в модели представления этого множества.

3. Рассмотрим теперь более общий случай. Пусть дан информационный объект, для которого определено его представление моделью $M_A = \langle A, \theta_A \rangle$ сигнатуры $\Sigma = \{p_i^{(l_i)}\} i \in I$, где $\theta_A = \{r_i^{(l_i)}\} i \in I$ — множество l_i-арных отношений $r_i^{(l_i)}$ на A; I — индексирующее множество и $\forall i \in I p_i^{(l_i)}$ — имя l_i-местного предиката $p_i^{(l_i)}(x_1, \ldots, x_{l_i})$, для которого выполняется условие

$$\forall x_1, \ \ldots, x_{l_i} \in A \, p_i^{(l_i)}(x_1, \ \ldots, x_{l_i}) =$$
$$= 1 \Leftrightarrow (x_1, \ \ldots, x_{l_i}) \in r_i^{(l_i)}.$$

Известно [54], что для описания свойств моделей вида M_A используется язык формул логики предикатов сигнатуры Σ, при этом вводится понятие множества элементарных формул — $\{p_i^{(l_i)}(x_1, \ldots, x_{l_i})\} i \in I$, описывающих основные свойства модели, т. е. отношения из θ_A. Остальные формулы определяются индуктивно из элементарных с помощью специальных правил с использованием логических операций.

Обозначим через $\Phi^{(l)}(x_1, \ldots, x_l)$ произвольно взятую l-местную формулу сигнатуры Σ со свободными переменными $x_1, \ldots, x_l$, принимающими значения из множества A. При подстановке в формулу различных значений из A формула, так же как и всякий предикат $p_i^{(l_i)}(x_1, \ldots, x_{l_i})$, может принимать значение 1 (истинно) или 0 (ложно). Если при этом формула не равна тождественно нулю, то она называется выполнимой в модели M_A. Всякая выполнимая формула описывает производное (выводимое из основных) свойство модели M_A. В частности, если $\Phi^{(l)}(x_1, \ldots, x_l)$ выполнима, то она описывает свойство кортежей $\langle x_1, \ldots, x_l \rangle$, $x_i \in A$, $i = \overline{1, l}$ некоторых элементов из A находиться в l-арном отношении $r_\Phi^{(l)}$, определяемом выражением

$$r_\Phi^{(l)} = \{\langle x_1, \ \ldots, x_l \rangle, \ \Phi^{(l)}(x_1, \ \ldots, x_l) = 1\}.$$

Пусть $M_B = \langle B, \theta_B \rangle$, $\theta_B = \{\rho_i^{(l_i)}\} i \in I$ — другая модель той же сигнатуры Σ, изоморфная модели M_A. Это значит, что существует изоморфизм $f: A \to B$ модели M_A на модель M_B, т. е. взаимно однозначное отображение A и B, удовлетворяющее условию

$$\forall i \in I, \qquad \forall x_1, \ \ldots, x_{l_i} \in A \langle x_1, \ \ldots, x_{l_i} \rangle \in r_i^{(l_i)} \Leftrightarrow$$
$$\Leftrightarrow \langle f(x_1), \ \ldots, f(x_l) \rangle \in \rho_i^{(l_i)}.$$

Доказано [54], что в этом случае для всякой формулы $\Phi^{(l)}(x_1, \ldots, x_l)$ сигнатуры Σ имеет место

$$\forall x_1, \ \ldots, x_l \in A \Phi^{(l)}(x_1, \ \ldots, x_l) =$$
$$= 1 \Leftrightarrow \Phi^{(l)}(f(x_1), \ \ldots, f(x_l)) = 1.$$

Пусть

$$\rho_\Phi^{(l)} = \{\langle f(x_1), \ \ldots, f(x_l) \rangle, \qquad \Phi^{(l)}(f(x_1), \ \ldots, f(x_l)) = 1\}.$$

Тогда $A = \{x, f(x) \in B\}$ и

$$r_\Phi^{(l)} = \{\langle x_1, \ \ldots, x_l \rangle, \ \langle f(x_1), \ \ldots, f(x_l) \rangle \in \rho_\Phi^{(l)}\}.$$

Отсюда следует, что квант информации, определенный изоморфизмом f, является порцией информации, необходимой и достаточной для описания любого, в том числе выводного, свойства модели M_A, т. е. любого отношения, определимого на A в языке (посредством) формул логики предикатов сигнатуры модели M_A. С точки зрения содержания данное описание является полным и точным описанием свойств информационного объекта при его представлении моделью M_A.

Легко проверить, что, как и ранее (п.п. 1, 2), указанное свойство кванта не зависит от выбора конкретного представителя из множества квантов, определенных всеми возможными изоморфизмами с началом в модели M_A. Поэтому естественно допустить, что каждый квант из указанного множества характеризуется одним и тем же объемом информации.

В теории множеств [44] для всякой модели определен единственный, присущий этой модели реляционный тип. Две модели имеют по определению один и тот же реляционный тип тогда и только тогда, когда они изоморфны. Таким образом, между объемом кванта информации, определенного изоморфизмом f с началом в модели M_A, и реляционным типом модели M_A, т. е. системным типом информационного объекта при его представлении моделью M_A, имеется взаимно однозначное соответствие. Следовательно, реляционный тип модели M_A (или модели M_B в силу равенства реляционных типов этих моделей) описывает объем указанного кванта информации.

Заметим, что с помощью понятия реляционного типа модели можно описывать объемы квантов полной и точной информации не только о модели в целом, но и об отдельных, в том числе выводных, свойствах этой модели. Действительно, пусть некоторое свойство модели M_A определено выполнимой на M_A формулой $\Phi^{(l)}(x_1, \ldots, x_l)$ и описывается отношением

$$r_\Phi^{(l)} = \{\langle x_1, \; . \; . \; . \; . \; , x_l\rangle, \; \Phi^{(l)}(x_1, \; . \; . \; . \; . \; , x_l) = 1\}.$$

Предположим, что если $\operatorname{Card}\Sigma = 1$, то $\Phi^{(l)}(x_1, \ldots, x_l)$ не является элементарным предикатом сигнатуры Σ.

Образуем систему с отношением $M_\Phi = \langle A, r_\Phi^{(l)}\rangle$. Эта система не является моделью сигнатуры Σ. Вместе с тем очевидно, что для нее можно определить сигнатуру $\Sigma_\Phi = \{\Phi^{(l)}\}$, полагая, что $\Phi^{(l)}(x_1, \ldots, x_l)$ является новым элементарным l-местным предикатом (в частности, одним из элементарных предикатов сигнатуры Σ), а $\Phi^{(l)}$ — именем отношения $r_\Phi^{(l)}$. Тогда система M_Φ становится моделью сигнатуры Σ_Φ. Модель M_Φ является представлением некоторого нового информационного объекта, а именно — указанного свойства модели M_A. Подобные объекты можно именовать виртуальными. Они не существуют сами по себе, а порождены представлениями о первоначальном информационном объекте, в нашем случае — представлением первоначально заданного информационного объекта, представленного моделью M_A. Однако несмотря на специфику информационного объекта, представленного моделью M_Φ, для самой модели M_Φ можно, повторив проведенные выше рассуждения, убедиться в следующем. Каждый квант полной и точной информации из множества всех квантов, определенных изоморфизмами с общим началом в модели M_Φ, имеет один и тот же объем. Этот объем может быть описан реляционным типом модели M_Φ.

Так как реляционые типы моделей M_A и M_Φ не совпадают, то объемы квантов информации $V(f)$ и $V(f')$, определенных изоморфизмами f и f' с началом в M_A и M_Φ соответственно, также различны. С точки зрения содержания квант информации, определенный изоморфизмом f', описывает лишь часть информации, определенной изоморфизмом f. Легко проверить, что изоморфизм f модели M_A на модель M_B влечет изоморфизм f' модели M_Φ на модель $M'_\Phi = \langle B, \rho_\Phi^{(l)}\rangle$, где $\rho_\Phi^{(l)} = \{\langle f'(x_1), \; \ldots, \; f'(x_l)\rangle$, $\Phi(f'(x_1), \ldots, f'(x_l)) = 1\}$. Обратное в общем случае неверно.

Сравнение подходов к оценке объемов информации в квантовой информатике и по формуле К. Шеннона

Пусть информационным объектом является сообщение $x = \lim\limits_{t\to\infty} x(t)$, образованное последовательностью $x(t) = x_1, x_2, \ldots, x_t$ элементарных сигналов $x_i \in A = \{a_1, \ldots, a_k\}$, $i = \overline{1, t}$, передаваемых эргодическим источником сообщений. В математической теории связи [29] количество информации, содержащейся в сообщении x, оценивается формулой К. Шеннона

$$I_x = -\sum_{\nu=1}^{k} P_\nu \log P_\nu,$$

где P_ν — вероятность появления в сообщении сигнала a_ν. В силу эргодичности источника сообщений

$$\forall \nu = \overline{1, k}, \qquad P_\nu = \lim_{t\to\infty} f_\nu,$$

где $f_\nu = t_\nu / t$ — частота, а t_ν — число встречаемости сигнала a_ν в последовательности $x(t)$. Для любого фиксированного t последовательность $x(t)$ можно также рассматривать как сообщение, содержащее следующее количество информации:

$$I_{x(t)} = -\sum_{\nu=1}^{k} f_\nu \log f_\nu.$$

Предположим, что для каждого сообщения $x(t)$ определен один и тот же и только один набор $\Sigma = \{g_\nu, \; \nu = \overline{1, k}\}$ свойств, заданный предикатами $g_\nu(i)$:

$$\forall i = \overline{1, t}, \qquad g_\nu(i) = 1 \Leftrightarrow x_i = a_\nu.$$

Тогда каждое сообщение $x(t)$ может быть представлено моделью $M_x(t) = \langle J_t, \; \theta_{x(t)}\rangle$ сигнатуры Σ, где $J_t = \{1, \ldots, t\}$, $\theta_{x(t)} = \{r_\nu, \; \nu = \overline{1, k}\}$ и $\forall i = \overline{1, t}$ $i \in r_\nu \Leftrightarrow g_\nu(i) = 1$. Указанную модель можно именовать моделью сообщения $x(t)$. Очевидно, что при любом t для модели $M_{x(t)}$ выполняются свойства

$$\forall \nu, \; \mu = \overline{1, k}, \qquad \nu \neq \mu \Leftrightarrow r_\nu \cap r_\mu = \varnothing \quad \text{и} \quad \bigcup_{\nu=1}^{k} r_\nu = J_t.$$

Согласно концепциям квантовой информатики объем $v(\varphi_t)$ кванта информации, определенного произвольно взятым изоморфизмом φ_t с началом в модели $M_{x(t)}$, описывается реляционным типом этой модели [т. е. системным типом информационного объекта, в данном случае — сообщения $x(t)$ при его представлении моделью $M_{x(t)}$]. Как указано в предыдущем параграфе, для описания реляционного типа модели $M_{x(t)}$ может быть использована подходящая шкала измерения. Такой шкалой может служить, например, шкала, определенная отображением $T: M \to S$, где M — множество всех моделей вида $M_{x(t)}$; S — множество всех кортежей $\langle n, n_1, \ldots, n_k\rangle$ натуральных чисел длины $k+1$ и T удовлетворяет условию

$$\forall M_{x(t)} \in M \qquad T(M_{x(t)}) = \langle t, \; t_1, \; . \; . \; . \; . \; , t_k\rangle \Leftrightarrow$$
$$\Leftrightarrow \forall \nu = \overline{1, k} \qquad t_\nu = \operatorname{Card} r_\nu \,\&\, t = \operatorname{Card} J_t.$$

Действительно, пусть $y = \lim\limits_{t\to\infty} y(t)$ — другое сообщение, образованное последовательностью $y(t) = y_1, y_2 \dots y_t$ элементарных сигналов $y_j \in A$, $j = \overline{1, t}$. Очевидно, что для каждого t существует представление $y(t)$ моделью $M_{y(t)} = \langle J_t, \theta_{y(t)} \rangle$ сигнатуры $\Sigma = \{g_\nu, \nu = \overline{1, k}\}$, где $J_t = \{1, \dots, t\}$, $\theta_{y(t)} = \{\rho_\nu, \nu = \overline{1, k}\}$, $\forall \nu = 1, k$ $\forall j = \overline{1, t}$ $g_\nu(j) = 1 \Leftrightarrow g_\nu = y_j$ и $j \in \rho_\nu \Leftrightarrow g_\nu(j) = 1$. Предположим, что модели $M_{x(t)}$ и $M_{y(t)}$ изоморфны. Это значит, что существует изоморфизм модели $M_{x(t)}$ на модель $M_{y(t)}$, т. е. взаимно однозначное отображение $\varphi_t : J_t \to J_t$, удовлетворяющее условию

$$\forall \nu = \overline{1, k}, \qquad \forall i \in J_t, \qquad i \in r_\nu \Leftrightarrow \varphi_t(i) \in \rho_\nu.$$

Легко видеть, что

$$\forall M_{x(t)}, \qquad M_{y(t)} \in M\, T(M_{x(t)}) = T(M_{y(t)}) \Leftrightarrow M_{x(t)} \sim M_{y(t)},$$

где символ $\sim$ используется для обозначения изоморфизма моделей.

Таким образом, объем $V(\varphi_t)$ кванта информации, определенного изоморфизмом φ_t, описывается в указанной шкале кортежем

$$V(\varphi_t) = T(M_{x(t)}) = \langle t, t_1, \ \dots, t_k \rangle.$$

Примером другого описания объема кванта информации φ_t, взаимно однозначного приведенному, может служить кортеж

$$V(\varphi_t) = \langle t, f_1, \ \dots, f_k \rangle,$$

где $f_\nu = t_\nu / t$ — частота сигнала a_ν в сообщении $x(t)$, $\nu = \overline{1, k}$.

Оценка $\nu(\varphi_t)$ кортежем $\langle t, f_1, \dots, f_k \rangle$ правомерна для любого t. Поэтому при $t \to \infty$ объем кванта информации, определенного изоморфизмом φ с началом в модели $M_x = \lim\limits_{t\to\infty} M_{x(t)}$, можно оценить следующим кортежем:

$$V(\varphi) = \langle \omega, P_1, \ \dots, P_k \rangle,$$

где P_ν — вероятность сигнала a_ν в сообщении x, $\nu = \overline{1, k}$ и ω — мощность множества натуральных чисел.

Поскольку для бесконечных моделей рассматриваемого класса мощность множества одна и та же, оценку объема указанного кванта можно представить в виде

$$V(\varphi) = \langle P_1, \ \dots, P_k \rangle.$$

Обозначим через $V(\psi)$ объем кванта информации, определенного изоморфизмом ψ с началом в модели $M_y = \lim\limits_{t\to\infty} M_y(t)$. Тогда, как нетрудно видеть, справедлива импликация

$$V(\varphi) = V(\psi) \Rightarrow I_x = I_y,$$

где I_x, I_y — оценка количества информации в сообщениях x и y соответственно по формуле К. Шеннона. Иными словами, равенство объемов квантовой информации, определенных изоморфизмами с началом в моделях M_x, M_y представления сообщений x, y, влечет равенство количества информации, содержащейся в сообщениях x и y по К. Шеннону. Обратная импликация неверна.

Таким образом, количество информации, определяемое по формуле К. Шеннона, является интегральной оценкой объема кванта полной и точной информации, описывающей сообщение x на уровне его представлению моделью M_x.

Сказанное не означает, что оценка количества информации по формуле К. Шеннона получает ограниченное применение. В литературе часто утверждается о неправомерности применения этой формулы для оценки количества семантической информации [58]. По мнению автора, напротив, предлагаемый в квантовой информатике подход к оценке объемов информации позволяет утверждать о возможности использования формулы для оценки количества семантической информации и при этом достаточно точно установить границы ее применимости и для этого случая. Действительно, пусть произвольный информационный объект имеет представление моделью M_A. Предположим, что кортеж $T(M_A) = \langle n, n_1, \dots, n_k \rangle$, $n_1, \dots, n_k \leqslant n$ описывает системный тип модели M_A. Тогда согласно принципу 9 (объемности) данный кортеж может служить описанием объема кванта полной и точной информации, носителем которой является информационный объект на уровне его представления моделью M_A. Следовательно, имеется возможность оценки указанной информации по формуле

$$I = -\sum_{\nu=1}^{k} \frac{n_\nu}{n} \log \frac{n_\nu}{n}.$$

Отсюда напрашивается вывод о том, что с точки зрения концепций квантовой информатики использование формулы К. Шеннона $I = -\sum\limits_{\nu=1}^{k} P_i \log P_i$ для оценки количества семантической информации, носителем которой служит информационный объект, является правомерным, если кортеж $\langle P_1, \dots, P_k \rangle$ описывает системный тип информационного объекта. Не следует только забывать, что такая оценка остается интегральной (между кортежем $\langle P_1, \dots, P_k \rangle$ и I существует лишь однозначность, но не взаимно однозначное соответствие).

Целесообразно отметить следующее. Полученные результаты можно было бы сформулировать, не обращаясь к представлению сообщения x моделью M_x. Действительно, обозначим через K_A класс всех возможных сообщений вида $x(t) = x_1, \dots, x_t$, $t = 1, 2 \dots$, образованных последовательностями элементарных сигналов $x_i \in A=\{a_1, \ldots, a_k\}$, $i=\overline{1, t}$. Для всякой пары сообщений $x(t)$, $y(t)$, $y(t)=y_1, \ldots, y_t$, $t=1, 2$, $y_j \in A$, $j=\overline{1, t}$, определим множество взаимно однозначных отображений вида $\varphi_t: X_t \to Y_t$, $X_T=\{x_i, i=\overline{1, t}\}$, $Y_T=\{y_j, j=\overline{1, t}\}$, удовлетворяющих условию $\forall i, j=\overline{1, t}$ $\forall \nu=\overline{1, k}$, $x_i=a_\nu \Leftrightarrow \varphi_t(x_i)=y_j=a_\nu$.

Обозначим через F_A класс всех возможных указанных отображений сообщений $x(t) \in K_A$ друг на друга. Легко проверить, что пара $K=\langle K_A, F_A \rangle$ образует категорию, в которой $K_A = obK$, $F_A = \mathrm{Mor}\, K$ (роль единичных морфизмов категории выполняет тождественные отображения каждого сообщения в себя). Очевидно, что не для всякой пары сообщений отображение φ_t существует. Однако если оно существует, то является изоморфизмом категории. Квант полной и точной информации, носителем которой является сообщение $x(t)$, определяется изоморфизмом φ_t с началом в $x(t)$, а его объем — системным типом $x(t)$. Нетрудно убедиться, что для описания системного типа $T(x(t))$ сообщение $x(t)$, а значит, и объема $v(\varphi_t)$ указанного кванта информации может быть использована шкала $T: K_A \to \mathrm{S}$, аналогичная использованной выше, с той разницей, что вместо множества M всех моделей вида $M_{x(t)}$ рассматривается множество K_A всех сообщений вида $x(t)$. Поэтому и в данном случае

$$v(\varphi_t)=T(x(t))=\langle t, t_1, \ldots, t_k \rangle.$$

Такое совпадение не случайно. Его можно строго обосновать. Для этого достаточно убедиться в том, что категория K является категорией, изоморфной подкатегории L категории моделей, в которой объектами являются модели $M_{x(t)}$, а морфизмами — всевозможные изоморфизмы этих моделей друг на друга. Как известно (см. гл. 4), произвольные категории K и L называются изоморфными, если существует взаимно однозначный ковариантный функтор $F: K \to L$, т. е. взаимно однозначное отображение K на L, удовлетворяющее следующим условиям:

1) $\forall \alpha \in Ob\, K$, $F(a) \in Ob\, L$;
2) $\forall \alpha \in \mathrm{Mor}_{(a, b)} K$, $F(\alpha) \in \mathrm{Mor}(F(a), F(b))^L$;
3) $\forall 1_a \in \mathrm{Mor}\, K$, $F(1_a)=1_{F(a)}$;
4) если $\alpha \in \mathrm{Mor}_{(a, b)} K$, $\beta \in \mathrm{Mor}_{(b, c)} K$, то $F(\alpha\beta)=F(\alpha)F(\beta)$.

Надо сказать, что и в общем случае оценка объема кванта полной и точной информации, носителем которой является информационный объект, не зависит от его представления в изоморфных категориях.

Часть 3

ЛОГИКИ

Глава 5

КЛАССИЧЕСКАЯ МАТЕМАТИЧЕСКАЯ ЛОГИКА

5.1. АЛГЕБРА И ИСЧИСЛЕНИЕ ВЫСКАЗЫВАНИЙ

В настоящем параграфе рассматриваются связи между высказываниями, которые воспринимаются через выражающие их предложения соответствующего предметного языка.

Для исследования логико-алгебраической модели требуется построить целую иерархию языков: U-язык, A-язык, O-язык [30]. С их помощью различаются изучаемая логика и логика, с помощью которой это делается. Выделим предметный язык (язык-объект, O-язык), который является объектом исследования, и язык исследования (U-язык), в рамках которого исследуется предметный язык. Высказывания считаются смыслом предложений [30]. Будем изучать исключительно то, как одни высказывания строятся из других, простых, «элементарных», играющих роль строительных блоков.

Алгебра высказываний. В алгебре высказываний в соответствии с теоретико-множественным подходом в качестве элементов множества выступают простые высказывания. Содержанием этой алгебры являются операции над этими высказываниями.

Под простым высказыванием понимается предложение, утверждающее что-то о чем-то. Рассматривая простое высказывание в определенных условиях времени и места, можно дать определенные утверждения о его истинности или ложности.

Пример 5.1. «Москва — столица СССР» в настоящий момент — истинное высказывание, а «2=1» — ложное.

Из простых высказываний можно образовывать новые составные высказывания. Для этих целей используются грамматические связки: «не», «и», «или», «если..., то», «тогда и только тогда». Построение из данных высказываний нового составного высказывания называется логической операцией над высказываниями.

В алгебре высказываний все высказывания рассматриваются только с точки зрения их логического значения, а от их содержания отвлекаются. Считается, что каждое высказывание либо истинно, либо ложно и что ни одно высказывание не может быть одновременно истинным и ложным. Будем обозначать высказывания латинскими буквами A, B, C... Основными логическими операциями над высказываниями являются отрицание, конъюнкция, дизъюнкция, импликация и эквивалентность.

Под отрицанием высказывания A понимается новое высказывание, обозначаемое $\neg A$, которое считается истинным, если A ложно, и ложно, если A истинно.

Пример 5.2. Для истинного высказывания «Восемь делится на четыре» отрицанием является ложное высказывание «Неверно, что восемь делится на четыре», или «Восемь не делится на четыре».

Конъюнкцией двух высказываний A и B считается новое высказывание, обозначаемое символом $A \wedge B$ (читается так: «A и B»), которое истинно, если оба высказывания A и B истинны, и ложно, если хотя бы одно из них ложно. Высказывания A и B называются членами конъюнкции.

Пример 5.3. Для высказываний «Семь меньше десяти» и «Восемь делится на четыре» конъюнкцией является истинное высказывание «Семь меньше десяти и восемь делится на четыре».

В алгебре высказываний союз «и» употребляется в том же значении, что и в повседневной речи. Но в обычной речи не принято соединять союзом «и» два высказывания, далекие друг от друга по содержанию. В алгебре высказываний рассматривается конъюнкция двух любых высказываний.

Дизъюнкцией двух высказываний A и B называется новое высказывание, которое обозначается символом $A \vee B$ (читается так: «A или B») и считается истинным, и ложным, если оба они ложны. A и B называются членами дизъюнкции.

Пример 5.4. Для двух высказываний «$\sqrt{2} > \sqrt{3}$», «Трава зеленая» дизъюнкция представляет собой истинное высказывание «$\sqrt{2} > \sqrt{3}$ или трава зеленая».

В повседневной речи «или» употребляется в различном смысле: во-первых, в неисключающем, когда оно выражает, что из двух высказываний, по крайней мере, одно истинно, а возможно, что и оба истинны, и, во-вторых, в исключающем, в смысле «либо..., либо...», когда оно выражает, что из двух высказываний истинно только одно, а другое ложно.

Пример 5.5. В предложении «В жаркую погоду пьют воду или едят мороженое» союз «или» имеет неисключающий смысл, а в предложении «Сегодня мы пойдем на экскурсию или на пляж» «или» употребляется в исключающем смысле.

В алгебре высказываний «или» употребляется в неисключающем смысле. Кроме того, в повседневной речи употребляют только дизъюнкции, члены которых как-то связаны между собой по смыслу, тогда как в логике рассматриваются дизъюнкции любых двух высказываний.

Импликацией двух высказываний A и B называется новое высказывание, которое обозначается символом $A \rightarrow B$ (читается так: «Если A, то B») и считается ложным, если A истинно и B ложно, и истинным при всех других логических значениях высказываний A и B. Высказывание A называется условием или посылкой, высказывание B — заключением или следствием импликации.

Импликация $A \rightarrow B$ читается также следующим образом: «A влечет B», или «Из A следует B», или «A имплицирует B».

Пример 5.6. Высказывания «Если $2 \times 2 = 5$, то трава синяя» и «если $1 > 1$, то восемь делится на четыре» истинны, так как у первого из них посылка ложна, а у второго истинно следствие. Импликация «Если $2 = 2$, то $4 > 8$» ложна, поскольку ее условие истинно, а заключение ложно.

Под эквивалентностью высказываний A и B понимается новое высказывание, которое обозначается символом $A \leftrightarrow B$ (читается так: «A тогда и только тогда, когда B», или короче: «A эквивалентно B») и считается истинным, когда оба высказывания A и B либо истинны, либо ложны, и ложным в остальных случаях. Высказывания A и B называются членами эквивалентности.

Эквивалентность $A \leftrightarrow B$ читается также следующим образом: «Для того чтобы A, необходимо и достаточно, чтобы B», или «A, если и только если B», или «Если A, то B, и обратно», или «A равносильно B».

Пример 5.7. Высказывание «Треугольник ABC с вершиной A и основанием BC равнобедренный тогда и только тогда, когда $\angle B = \angle C$» есть пример эквивалентности.

Все рассмотренные операции можно наглядно представить с помощью так называемых *таблиц истинности*. В их строках указываются два возможных логических значения высказываний A и B: 1 (истина) и 0 (ложь), а также значения высказывания, полученного в результате применения соответствующей операции.

Отрицанию можно сопоставить таблицу истинности

A	$\neg A$
1	0
0	1

понимаемую так: «Когда A истинно, $\neg A$ ложно; когда A ложно, $\neg A$ истинно».

Конъюнкция высказываний A и B имеет следующую таблицу истинности:

A	B	$A\wedge B$
1	1	1
0	1	0
1	0	0
0	0	0

Высказывание «$A\wedge B$» истинно тогда и только тогда, когда истинны оба высказывания A и B.

Для дизъюнкции (связки «или») получается следующая таблица:

A	B	$A\vee B$
1	1	1
0	1	1
1	0	1
0	0	0

Таким образом, $A\vee B$ ложно только тогда, когда A и B ложны.

Для импликации таблица истинности имеет вид

A	B	$A\to B$
1	1	1
0	1	1
1	0	0
0	0	1

Такое уточнение смысла высказывания «Если A, то B» не противоречит обычной практике.

С помощью рассмотренных логических операций из заданной совокупности элементарных высказываний можно строить различные составные высказывания, при этом порядок выполнения операций указывается скобками, как в арифметике.

Пример 5.8. Из трех высказываний A, B, C можно построить высказывания

$$\neg(A\wedge B)\vee C,\quad A\to[B\leftrightarrow(A\vee C)].$$

Логическое значение составного высказывания зависит только от логических значений образующих его элементарных высказываний. Оно может быть найдено на основании определений логических операций над высказываниями.

5.2. ПРОПОЗИЦИОНАЛЬНЫЕ ТАВТОЛОГИИ

Для всех дедуктивных рассуждений используются некоторые предложения, которые считаются истинными исключительно в силу своего синтаксиса. Такие предложения, называемые *тавтологиями*, играют очень важную роль, так как помогают точно описать логические средства, применяемые в формализованных математических теориях.

Займемся изучением таких тавтологий, которые считаются истинными только в силу своего связочного синтаксиса, т. е. тавтологиями, построенными из предположений a, b посредством связок $\vee$, $\wedge$, $\to$, $\neg$ так, что они должны считаться истинными в силу особенностей их строения и совершенно независимо от того, истинны или ложны исходные предложения. Например, интуиция подсказывает, что следующие предложения истинны в силу лишь своего связочного синтаксиса, т. е. являются пропозициональными тавтологиями:

1) $\neg(a\wedge\neg a)$ — закон отрицания противоречия;

2) $((a\vee b)\to(\neg a\to b))$ — выражение дизъюнкции через отрицание и импликацию;

3) $(((a\to b)\wedge a)\to b)$.

Поясним точнее, что понимается под пропозициональной тавтологией с интуитивной точки зрения. Логическая интуиция побуждает связывать с предложениями их логические значения: истину или ложь. Вопрос о том, какие критерии могут быть выдвинуты для истинности или ложности предложений a, $b\ldots$, не рассматривается. Не будем также обсуждать, всегда ли в конкретных случаях возможно сопоставление предложений a, $b\ldots$ с какими-нибудь логическими значениями. Будем интересоваться только тем, что может быть сказано о ложности или истинности предложений α, $\beta\ldots$, построенных из предложений a, $b\ldots$ с помощью $\vee$, $\wedge$, $\to$, $\neg$, если известны логические значения предложений a, $b\ldots$

Пусть α и β — предложения, логические значения которых не известны. В соответствии с интуитивным смыслом пропозициональных связок дизъюнкция $\alpha\vee\beta$ истинна в том и только в том случае, когда истинно хотя бы одно из предложений α, β.

Конъюнкция $\alpha\wedge\beta$ истинна в том и только в том случае, когда оба предложения α, β истинны. Отрицание $\neg\alpha$ истинно тогда и только тогда, когда α ложно. Импликация $\alpha\to\beta$ считается ложной в том и только в том случае, когда α истинно, а β ложно. Во всех остальных случаях импликация истинна. Этот способ понимания импликации, впрочем, отличается от ее понимания в обиходном языке, поскольку в последнем случае импликация $\alpha \to \beta$ обычно понимается как возможность вывода β из α. Поэтому на первый взгляд обычное понимание импликации ничего не подсказывает в отношении логических значений следующих предложений:

1. Если $1+1=2$, то Париж есть столица Франции.
2. Если $1+1 \neq 2$, то Париж есть столица Франции.
3. Если $1+1 \neq 2$, то Рим есть столица Франции[1].

Заметим, что импликация в таком понимании имеет следующее важное свойство, называемое *правилом отделения* (modus ponens): если $(\alpha \to \beta)$ *истинно* и α истинно, то β истинно. Иначе это правило называется первой формой гипотетического *силлогизма*. Под *силлогизмом* понимается дедуктивное умозаключение, в котором одно суждение является необходимым следствием двух других. Это содержание предложения 3. Можно легко убедиться в том, что 3 является тавтологией в соответствии с установленным ранее значением связок. В самом деле, 3 может быть ложным только в том случае, если $((a \to b) \wedge a)$ истинно и b ложно. Но первое предложение истинно только в том случае, когда a истинно и $(a \to b)$ истинно. Поэтому предложение 3 всегда истинно независимо от логических значений a и b.

5.3. ИСЧИСЛЕНИЕ ВЫСКАЗЫВАНИЙ

В данном параграфе будет дано изложение так называемого исчисления высказываний (пропозициональных исчислений), хотя, как уже говорилось в § 5.1, этот термин используется в силу традиций и не совсем правомочен, так как в данном случае связанные переменные не рассматриваются. Изложение производится в соответствии с лингвистической концепцией.

Исчисление высказываний строится как формальная аксиоматическая система на основе пропозициональных тавтологий, используемых для дедуктивных рассуждений. Языки пропозициональных исчислений отличаются от языков элементарных математических теорий. Опишем сначала их алфавиты. Под алфавитом пропозиционального исчисления (или под алфавитом нулевого порядка) будет пониматься любая упорядоченная система:

$$A_0 = \{V_0, \quad L_1, \quad L_2, \ U\},$$

---

[1] Смысл этих предложений неясен, поскольку мы привыкли к тому, что между посылкой и заключением имеется определенная (причинная) связь. Например, высказывание «Если этот кусок железа положен в воду в момент времени t, то железо растворится» рассматривается как ложное даже в том случае, если кусок железа не положен в воду в момент времени t, т. е. даже если посылка ложна.

где V_0, L_1, L_2, U — непересекающиеся множества; множество V_0 бесконечно; объединение L_1, L_2 не пусто.

Элементы множества V_0 называются пропозициональными переменными и обозначаются a, b, c, если нужно — с индексами.

Элементы множества L_1 называются унарными пропозициональными связками, элементы множества L_2 бинарными пропозициональными связками.

Множество L_1 содержит ровно один элемент, обозначаемый символом $\urcorner$ и называемый знаком отрицания; множество L_2 содержит три элемента, обозначаемых символами $\wedge$, $\vee$, $\to$ и называемых знаками конъюнкции, дизъюнкции и импликации соответственно.

Элементы множества U называются вспомогательными знаками. Множество U содержит ровно два элемента, обозначаемых (,) и называемых скобками.

Объединение множеств V_0, L_1, L_2, U есть множество знаков алфавита A_0.

Множество F_0 формул, образованных из знаков алфавита A_0,— это наименьшее из таких множеств конечных последовательностей знаков в алфавите A_0, которые удовлетворяют требованиям:

а) все пропозициональные переменные, рассматриваемые как одноэлементные последовательности, принадлежат F_0;

б) если α принадлежит множеству F_0 и $\circ$ — унарная пропозициональная связка в A_0, то $\circ\,\alpha$ принадлежит множеству F_0;

в) если α и β принадлежат множеству F_0 и $\circ$ — бинарная пропозициональная связка в алфавите A_0, то $(\alpha \circ \beta)$ принадлежит множеству F_0. Пара

$$\mathscr{L}_0 = \{A_0 F_0\}, \qquad (5.1)$$

где A_0 — алфавит нулевого порядка и F_0 — множество всех формул, образованных из знаков алфавита A_0, называется формализованным языком нулевого порядка (или формализованным языком пропозиционального исчисления).

Пропозициональным исчислением будем называть любую дедуктивную систему $\mathscr{L}_0 = \{\mathscr{L}_0, C\}$, где $\mathscr{L}_0$ — формализованный язык нулевого порядка, а C — операция присоединения следствий в язык $\mathscr{L}_0$.

Остановимся на некоторых основных свойствах языков $\mathscr{L}_0$ нулевого порядка. Если α — формула в языке $\mathscr{L}_0$ и все пропозициональные переменные $a, b, c, \ldots, b\alpha$ интерпретированы как предложения, то и α может быть интерпретирована как предложение, причем знаки $\vee$, $\wedge$, $\to$, $\urcorner$ интерпретируются как «или», «и», «если..., то...», «не».

Формула α в языке $\mathscr{L}_0$ называется пропозициональной тавтологией (или просто тавтологией), если α — истинное предложение независимо от того, интерпретируются a, b, $c\ldots$ как истинные или как ложные предложения.

Под интерпретацией языка $\mathcal{L}_0$ нулевого порядка будем понимать любое отображение 1, каждая пропозициональная переменная которого сопоставляется с предложением (истинным или ложным). Поэтому каждая интерпретация 1 для языка $\mathcal{L}_0$ переводит любую формулу $\alpha(b\mathcal{L}_0)$ в предложение α_1 (истинное или ложное).

Формула α является пропозициональной тавтологией тогда и только тогда, когда α_1 является истинным предложением при любой интерпретации 1 для $\mathcal{L}_0$.

5.4. БУЛЕВА АЛГЕБРА

В алгебре высказываний для проверки истинности тех или иных утверждений использовался метод таблиц истинности. Этот метод громоздок и обладает определенными недостатками. В булевой алгебре рассматривается совокупность операций над множеством, состоящим из двух элементов: 0 и 1, причем ложности и истинности высказываний ставятся в соответствие 0 и 1. Благодаря этому по сравнению с использованием высказываний многие выкладки упрощаются, хотя полностью избавиться от громоздких таблиц истинности здесь не удается.

Однако следует заметить, что обобщение исчисления высказываний приводит к построению исчисления предикатов (см. § 5.8). Далее дано изложение булевой алгебры на основе структурной концепции.

В соответствии с этой концепцией вводится понятие структуры. Под *структурой* понимается такая система, в которой выполняется свойство наличия верхней границы (sup) и нижней границы (inf) для любой пары элементов. В качестве верхней и нижней границ выступают сами элементы 0 и 1. В соответствии с общим разделением систем на реляционные и ассерторические [30] данную структуру можно отнести к классу реляционных систем.

В булевой алгебре в отличие от алгебры высказываний используется другая система обозначений операций. Так, символ $\vee$ соответствует $\cup$, а символ $\wedge$ — $\cap$.

В соответствии с теорией структур для введения упорядоченности во множествах (сравнения 0 и 1) вводится в рассмотрение бинарное отношение $\leqslant$.

С позиций теории множеств *алгебра Буля* есть непустое множество A_B элементов a, b, $c\ldots$, для которых определены:

1) двухместная операция (называемая *булевым сложением*), сопоставляющая каждым двум элементам a и b из множества A_B какой-либо элемент из того же множества, который обозначается символом $a\vee b$ (и называется *булевой суммой* элементов a и b);

2) двухместная операция (называемая *булевым умножением*), сопоставляющая каждым двум элементам a и b из множества A_B какой-либо элемент из того же множества, который обозначается символом $a\wedge b$ (и называется *булевым произведением* элементов a и b);

3) одноместная операция, сопоставляющая каждому элементу a из множества A_B какой-либо элемент из того же множества, который обозначается символом $\overline{a}$ (и называется *булевым дополнением* к элементу a), причем выполняются следующие аксиомы:

1) $a\vee b=b\vee a$; $a\wedge b=b\wedge a$ (коммутативность);

2) $(a\vee b)\vee c=a\vee(b\vee c)$; $(a\wedge b)\wedge c=a\wedge(b\wedge c)$ (ассоциативность);

3) $a\vee(b\wedge c)=(a\vee b)\wedge(a\vee c)$; $a\wedge(b\vee c)=(a\wedge b)\vee(a\wedge c)$ (дистрибутивность);

$$\left.\begin{array}{lll} 4) & (a\vee b)\wedge b=b; & (a\wedge b)\vee b=b; \\ 5) & (a\vee a)\wedge b=b; & (a\wedge\overline{a})\vee b=b. \end{array}\right\}\text{(поглощение);}$$

Из этих аксиом вытекают следующие теоремы:

1) $a\vee a=a$, $a\wedge a=a$ (идемпотентность);

2) $a\vee\overline{a}=b\vee\overline{b}$, $a\wedge\overline{a}=b\wedge\overline{b}$ для произвольных a и b из множеств A_B.

Из теоремы следует, что элементы $a\vee\overline{a}$ и $a\wedge\overline{a}$ не зависят от выбора $a\in A_B$. Элемент $a\vee\overline{a}$ называется *булевой единицей* и обозначается символом e, элемент $a\wedge\overline{a}$ называется *булевым нулем* и обозначается символом 0;

3) $a\vee 0=a$, $a\wedge 0=0$, $a\vee e=e$, $a\wedge e=a$, для любого $a\in A_B$;

4) $a=a$ для любого $a\in A_B$;

5) $\overline{a\vee b}=\overline{a}\wedge\overline{b}$, $\overline{a\wedge b}=\overline{a}\vee\overline{b}$;

6) $\overline{0}=e$, $\overline{e}=0$.

Один из наиболее важных примеров булевых алгебр — *алгебра логики*, которая изучает высказывания и логические операции над высказываниями.

Бинарное отношение $\leqslant$, определенное на множестве A, называется *отношением порядка* на множестве A, если оно рефлексивно, антисимметрично и тразитивно, т. е.

1) $x\leqslant x$;

2) если $x\leqslant y$ и $y\leqslant x$, то $x=y$;

3) если $x\leqslant y$ и $y\leqslant z$, то $x\leqslant z$ для любых x, y, $z\in A$.

Вместо $x\leqslant y$ можно писать $y\geqslant x$. Если $x\leqslant y$, то говорят, что y больше или равно x, или x меньше или равно y, или x включено в y.

Другими словами, упорядоченное множество — это пара $\{A, \leqslant\}$, где A — непустое множество и $\leqslant$ — отношение порядка на множестве A. Пусть S — непустое множество элементов упорядоченного множества A. Элемент a_0 называется *верхней* (*нижней*) *границей* множества S, если $a \leqslant a_0 (a \geqslant a_0)$ при всех $a \in S$. Если множество всех верхних (нижних) границ множества S содержит наименьший (наибольший) элемент, то он называется *точной верхней (точной нижней) границей* множества S и обозначается $\sup S$ ($\inf S$). Особенно важен случай, когда каждое двухэлементное подмножество A имеет точную верхнюю и точную нижнюю границы.

Отношение порядка $\leqslant$ на множестве A называется *отношением структурного порядка*, если для любых a и $b \in A$ элементы

$$\sup(a, b) \quad \text{и} \quad \inf(a, b)$$

существуют.

В этом случае упорядоченное множество A называется *решеткой* (*структурой*). Точная верхняя граница элементов a, $b \in A$ обозначается $a \bigcup b$ и называется объединением элементов a и b; точная нижняя граница элементов a, $b \in A$ обозначается $a \bigcap b$ и называется пересечением элементов a и b.

Следующие свойства объединения и пересечения следуют непосредственно из определений (тождеств):

1) $a \bigcup a = a$, $a \bigcap a = a$ (законы идемпотентности);

2) $a \bigcup b = b \bigcup a$, $a \bigcap b = b \bigcap a$ (законы коммутативности $\bigcup$ и $\bigcap$);

3) $a \bigcup (b \bigcup c) = (a \bigcup b) \bigcup c$, $a \bigcap (b \bigcap c) = (a \bigcap b) \bigcap c$ (законы ассоциативности $\bigcup$ и $\bigcap$);

4) $(a \bigcap b) \bigcup b = b$, $a \bigcap (a \bigcup b) = a$ (законы поглощения);

5) $a \bigcap (b \bigcup c) = (a \bigcap b) \bigcup (a \bigcap c)$, $a \bigcup (b \bigcap c) = (a \bigcup b) \bigcap (a \bigcup c)$ — (законы дистрибутивности).

Заметим, что если для любых a и $b \in A$ выполняется один из законов дистрибутивности, то выполняется и другой. Следовательно, чтобы установить дистрибутивность структуры A, достаточно показать, что одно из тождеств 5 выполняется в структуре A.

В самом деле, предположим, что выполняется первое из тождеств 5, тогда

$$(a \cup b) \cap (a \cup c) = ((a \cup b) \cap a) \cup ((a \cup b) \cap c) = a \cup ((a \cap c) \cup (b \cap c)) =$$
$$= (a \cup (a \cap c)) \cup (b \cap c) = a \cup (b \cap c)$$

в силу тождеств 1—3 и первого закона дистрибутивности.

Как сказано в гл. 2, теоретико-множественное дополнение подмножества A в пространстве X может быть определено или как наибольшее подмножество пространства X, не пересекающееся с подмножеством A, или как наименьшее подмножество пространства X, дающее в объединении с подмножеством A все пространство X. Каждое из этих замечаний уточняет определение дополнения элементов в структурах. Но, к сожалению, эти определения неэквивалентны. Поэтому следует определить два понятия дополнения элемента a в структуре A.

Предположим, что в структуре A имеется нулевой[1] элемент (0). Элемент c в структуре A называется $\bigcap$-дополнением элемента a в структуре A, если c — наибольший элемент со свойством $a \bigcap c = 0$ (т. е. если c — наибольший элемент в множестве таких X, что $a \bigcap X = 0$).

Предположим, что в структуре A имеется единичный элемент 1. Элемент $c \in A$ называется $\bigcup$-*дополнением* элемента a в структуре A, если c — наименьший элемент a в структуре A, если c — наименьший элемент со свойством $a \bigcup c = 1$ (т. е. если c — наименьший элемент в множестве таких x, что $a \bigcup x = 1$). Элемент $c \in A$ называется дополнением элемента a, если c является $\bigcap$- и $\bigcup$-дополнением для a.

Структура A называется *булевой алгеброй*, если она дистрибутивна и каждый элемент $a \in A$ имеет дополнение, обозначаемое $-a$. Второе условие предполагает, что в булевой алгебре имеются нулевой и единичный элементы.

Другими словами, булевой алгеброй является такая дистрибутивная решетка A, в которой каждому $a \in A$ сопоставлен элемент $-a$ со свойствами

6) $(a \bigcap -a) \bigcup b = b$, $(a \bigcup -a) \bigcap b = b$.

Кроме того, эти свойства означают, что

$$a \cap -a = 0 \quad \text{и} \quad a \cup -a = 1.$$

Поэтому булеву алгебру A_0 можно записать в виде

$$A_0 = \langle \{0, 1\}, \ \cup, \ \cap, \ -, \ 0, \ 1 \rangle.$$

5.5. БУЛЕВА АЛГЕБРА И ПРОПОЗИЦИОНАЛЬНЫЕ ТАВТОЛОГИИ

Правила, определяющие истинность или ложность дизъюнкции, конъюнкции, импликации и отрицания в соответствии с истинностью или ложностью их компонентов могут быть записаны

---

[1] Наибольший (наименьший) элемент структуры A, если он существует, будем называть *единичным* (*нулевым*) *элементом* и обозначать 1(0). По определению $x < 1$; $x > 0$; $x \bigcup 1 = 1$; $x \bigcap 1 = x$; $x \bigcup 1 = x$, $x \bigcap 0 = 0$, где $\bigcup$ и $\bigcap$ — двухместные функциональные символы, а 0 и 1 — константные символы.

в форме следующих схем, где 0 и 1 означают соответственно какое-нибудь ложное или истинное предложение исчисления высказываний:

$$0 \cap 0 = 0; \qquad 0 \cup 1 = 1 \cup 0 = 1 \cup 1 = 1;$$

$$1 \cap 1 = 1; \qquad 0 \cap 1 = 1 \cap 0 = 0 \cap 0 = 0;$$

$$1 \to 0 = 0; \qquad 1 \to 1 = 0 \to 0 = 0 \to 1 = 1;$$

$$-0 = 1; \qquad -1 = 0.$$

Это обстоятельство позволяет решить очень простым способом, истинно или ложно предложение α, составленное из предложений $a, b, \ldots$, логическое значение которых мы знаем, с помощью пропозициональных связок.

Для этого достаточно:

а) заменить каждое из элементарных предложений $a, b \ldots$, встречающихся в предложении α, на 1, если оно истинно, и на 0, если оно ложно;

б) интерпретировать 0 и 1 как нулевой и единичный элементы двухэлементной булевой алгебры;

в) понимать пропозициональные связи $\vee$, $\wedge$, $\to$ как $\cup$, $\cap$, $\to$ в булевой алгебре (соответственно).

Если в результате выполнения булевых операций в булевом выражении, полученном описанным способом, получаем 1, то предложение α истинно. Если же получаем 0, то α ложно.

Предположим, например, что в предложении

$$(\neg(a \wedge b) \to (\neg a \cup \neg b))$$

a — истинное предложение, а b — ложное. Тогда это предложение истинно, так как $-1(1 \cap 0) \to (-1 \cup -0) = 1 \to 1 = 1$.

Эта проверка истинности или ложности составных предложений осуществляется с помощью таблиц истинности.

Рассмотрим, например, построение устройства, моделирующего принятие некоторой резолюции «комитетом трех». Каждый член комитета при одобрении резолюции нажимает свою кнопку; если большинство членов согласны, то резолюция принимается, что фиксируется регистрирующим прибором. Устройство реализует высказывание

$$f(x_1, x_2, x_3) = \neg x_1 \; x_2 \; x_3 \vee x_1 \neg x_2 x_3 \vee x_1 x_2 \neg x_3 \vee x_1 x_2 x_3,$$

при этом таблица истинности (табл. 5.1) имеет следующий вид:

Таблица 5.1

x_1	x_2	x_3	$f(x_1 x_2 x_3)$
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1

Если предложение α, составленное из предложений a, b, c с помощью пропозициональных связок, всегда истинно независимо от истинности или ложности $a, b \ldots$, то α называется пропозициональной тавтологией, если соответствующее булево выражение в двухэлементной булевой алгебре $\mathscr{L}_0$ принимает значение 1 при всех подстановках 1 и 0 вместо $a, b \ldots$

Чтобы проверить, будет ли тавтологией предложение, образованное из элементарных предложений, нужно выполнить все возможные подстановки 1, 0 вместо $a_1, a_2, \ldots, a_n$, т. е. всего 2^n подстановок. Вообще говоря, излагаемый далее метод поиска подстановки, дающий значение 0, более прост, так как он позволяет избежать большого числа подстановок.

Поясним этот метод на двух примерах.

Пример 5.9. Рассмотрим предложение

$$((a_1 \to (a_2 \to a_3)) \to ((a_1 \to a_2) \to (a_1 \to a_3))). \qquad (5.2)$$

Предположим, что для некоторых a_1, a_2, a_3 в двухэлементной булевой алгебре A получает значение 0:

$$((a_1 \to (a_2 \to a_3)) \to ((a_1 \to a_2) \to (a_1 \to a_3))) = 0.$$

Это возможно в том и только в том случае, когда

$$(a_1 \to (a_2 \to a_3)) = 1 \text{ и } ((a_1 \to a_2) \to (a_1 \to a_3)) = 0. \qquad (5.3)$$

Последнее равенство возможно тогда и только тогда, когда

$$(a_1 \to a_2) = 1 \text{ и } (a_1 \to a_3) = 0, \qquad (5.4)$$

т. е.

$$a_1 = 1 \text{ и } a_3 = 0. \qquad (5.5)$$

Из (5.3) и (5.4) следует, что $a_2 = 1$. Отсюда

$$(a_1 \to (a_2 \to a_3)) = 1 \to (1 \to 0) = 1 \to 0 = 0,$$

что противоречит (5.3). Поэтому не существует подстановки 1, 0 вместо a_1, a_2, a_3, дающей 0, и предложение (5.2) является тавтологией.

Пропозициональные тавтологии образуют исходный пункт всякого дедуктивного рассуждения.

Пример 5.10. Докажем утверждение вида $(a \to b)$. Легко установить, что выражение

$$((\neg b \to \neg a)) \to (a \to b) \tag{5.6}$$

является пропозициональной тавтологией. Отсюда в силу правила отделения достаточно доказать, что $(\neg b \to \neg a)$. Этот метод часто применяется в математической практике.

5.6. БУЛЕВА АЛГЕБРА И ИСЧИСЛЕНИЕ ВЫСКАЗЫВАНИЙ

При интерпретации языка нулевого порядка $\mathscr{L}_0$ все пропозициональные переменные $a, b, c \ldots$, фигурирующие в формуле предложения α, понимаются как переменные, истинно (1) и ложно (0) — как множество элементов двухэлементной булевой алгебры A_0; знаки $\vee$, $\wedge$, $\to$, $\neg$ интерпретируются как знаки булевых операций в A_0. При этом условии α понимается как отображение $A_0 \times A_0 \times \ldots \times A_0$ в A_0. Это отображение обозначается α_{A_0} и называется булевым многочленом, определяемым формулой предложения α в алгебре A_0.

Если, например, α — формула

$$\neg a_1 \wedge a_2 \wedge a_3 \vee a_1 \wedge a_2 \wedge a_3 \vee a_1 \wedge a_2 \wedge a_3 \vee a_1 \wedge$$
$$\wedge a_2 \wedge a_3 \vee a_1 \wedge a_2 \wedge a_3 \vee a_1 \wedge a_2 \wedge a_3,$$

то α_{A_0} — это отображение, сопоставляющее каждую «тройку» (a, b, c) с элементом α_{A_0}

$$(a, b, c) = a_1 \cap a_2 \cap a_3 \cup a_1 \cap -a_2 \cap a_3 \cup a_1 \cap a_2 \cap -$$
$$-a_3 \cup a_1 \cap a_2 \cap a_3 \cup a_1 \cap a_2 \cap a_3 \cup a_1 \cap a_2 \cap a_3 =$$
$$= (a_1 \cup -a_1) \cap a_2 \cap a_3 \cup a_3 \cup (a_2 \cup -a_2) \cap a_1 \cap a_3 \cup$$
$$\cup (a_3 \cup -a_3) \cap a_1 \cap a_2 = a_2 \cap a_3 \cup a_1 \cap a_3 \cup a_1 \cap a_2 = a_3 \cap (a_1 \cup a_2) \cup a_1 \cap a_2.$$

Используя указанную интерпретацию формул как булевых многочленов, сформулируем более точно определение пропозициональной тавтологии: формула в языке $\mathscr{L}_0$ является пропозициональной тавтологией, если отображение

$$\alpha_{A_0} : A_0 \times A_0 \times \ . \ . \ . \times A_0 \to A_0$$

тождественно равно единичному элементу алгебры A_0.

Воспользовавшись булевыми многочленами, легко доказать следующие утверждения:

1. Если формула α — пропозициональная тавтология в языке $\mathscr{L}_0$, $\beta_1, \ldots, \beta_m$ — формулы в языке $\mathscr{L}_0$ и $a_1, \ldots, a_m$ — пропозициональные переменные, то формулы $\alpha(a_1/\beta_1, \ldots, a_m/\beta_m)$ — также формулы в языке $\mathscr{L}_0$.

2. Если α — пропозициональная тавтология в языке $\mathscr{L}_0$, $\beta_1, \ldots, \beta_m$ — формулы в языке $\mathscr{L}_0$ и $a_1, \ldots, a_m$ — пропозициональные переменные, то формулы $\alpha(a_1/\beta_1, \ldots, a_m/\beta_m)$ — также пропозициональная тавтология.

3. Если предложение α и $(\alpha \to \beta)$ — пропозициональные тавтологии, то и β — также тавтология.

Для иллюстрации докажем утверждение 3. Легко заметить, что

$$\alpha_{A_0} \to \beta_{A_0} = (\alpha \to \beta)_{A_0} = 1 \quad \text{и} \quad \alpha_{A_0} = 1,$$

поэтому $\beta_{A_0} = 1$ тождественно.

5.7. ПРИКЛАДНЫЕ ВОПРОСЫ ПРОПОЗИЦИОНАЛЬНЫХ ИСЧИСЛЕНИЙ

Рассмотрим проблему построения вопросно-ответной системы с использованием ЭВМ.

Во-первых, покажем, как можно применить математическую логику при проектировании отдельных компонентов информационной системы. Во-вторых, попытаемся с помощью аппарата математической логики выполнить некоторый анализ естественного языка.

Компоненты информационной системы. На рис. 5.1 представлена схема информационной системы. У нее два входа: данные некоторого вида и запросы-требования информации со стороны пользователя.

Проблема заключается в сопоставлении входного запроса с данными. Основой для такого сравнения может послужить представление как данных, так и запросов в виде строк символов. С этой целью обе точки входа системы располагают процедурами представления, которые завершаются статьями описания файла и вопросами в символическом виде. Пока оставим открытым вопрос о процессах представления (можно считать, что такая процедура выполняется либо вручную, либо автоматически). Предположим, что пользователь формулирует запросы на естественном языке, не уточняя форму представления данных. Далее выполним, скорее, логический, чем грамматический анализ запросов.

Рис. 5.1. Схема информационной системы

Анализ вопроса на естественном языке. Пусть в информационной системе хранятся библиографические данные по математической логике, а вопрос имеет вид:

Какие книги написаны Ю. Л. Ершовым? (5.7)

Рассмотрим связь между вопросом и ответом (связь между соответствующими выражениями). Одним из ответов явится наименование «Теория нумераций», поскольку

Ю. Л. Ершов написал «Теорию нумераций» (5.8)

«Теория нумераций» есть книга. (5.9)

Представим символически оба элементарных предложения. Предложение (5.7) представляет собой отношение, имеющее место между двумя индивидуальностями: человек и написанная работа. Обозначим это отношение W, имея в виду, что у этого отношения два аргумента. Пусть a и b — символические обозначения для «Ю. Л. Ершова» и «Теории нумераций» соответственно. Тогда предложение приобретает вид

$$aWb. \qquad (5.10)$$

Отношение W (вместе с позициями для аргументов) называют двухместным предикатом, соответствующим данному отношению; a и b — выражения аргументов (в данном случае это индивидные константы, обозначающие индивидуальности). Аналогично выполняется символическая запись предложения (5.9):

$$Bb, \qquad (5.11)$$

где символу b соответствует «книга», а в качестве одноместного предиката выступает некоторое свойство, а именно быть книгой.

Отметим, что понятие одноместного предиката тесно связано с понятием класса; так, предложение (5.11) можно понимать как b принадлежит классу B.

Метод пропозиционального хранения данных. Этот метод предполагает хранение символических выражений, таких же, как и рассмотренные ранее. Например, при хранении в памяти ЭВМ предложения (5.8) слова русского языка переводятся (транслируются) в машинные слова (аналогичные W, a, b). Так, в файле на диске предложение (5.10) хранится в виде четырех слов

1	*a*	*W*	*b*

,

где четвертое слово играет роль идентификаторов. Аналогично предложения (5.9) или (5.11) имеют машинное представление

Таблица 5.2

Символ	Части предложения и отношения
a	Ю. Л. Ершов
b	Теория нумераций
c	Об одной иерархии множеств
d	С. Клини
e	Р. Весли
f	Основания интуиционистской математики
B	(свойство быть) книга(свойство быть) статья
	написал (отношение «быть автором» в широком смысле)

2	—	*B*	*b*

,

причем здесь указывается, что второе место пусто (помечено символом —).

При таком подходе база данных состоит из двух частей: словаря индивидных констант и предикатов (их называют дескриптивными именами) и набора элементарных предложений, представляющих собой конечные строки дескриптивных имен.

Пример словаря приведен в табл. 5.2. Будем считать его запасом (хранилищем) полного пространства, в рамках которого выполняется рассуждение.

В табл. 5.3 представлен соответствующий пример файла. Он описывает возможное состояние пространства рассуждений.

Таблица 5.3

Символ	Отношение (предикат)	Символ
a	*W*	*b*
—	*B*	*b*
a	*W*	*c*
—	*P*	*c*
d	*W*	*f*
e	*W*	*f*
—	*B*	*f*

Вернемся теперь к вопросу (5.7) и запишем его в символическом виде. Выделим логические «части речи»: выражения аргументов, предикаты и логические символы. Результатом анализа вопроса следует считать идентификацию логических единиц, соответствующих выделенным трем классам.

Значениями аргументов являются «какие» и «Ю. Л. Ершов». Часть предложения «какие» имеет статус переменной (дескриптивной переменной вопроса). Другой частью предложения (5.7) является составное выражение-отношение, которое обозначим R_3, так как в нем можно выделить три уровня сложности. Таким образом, символическая запись вопроса приобретает вид

$$xR_3a. \tag{5.12}$$

Кандидатуры для ответа на вопрос (5.7) ограничиваются классом книг. Поэтому R_3 является отношением с ограниченной областью определения (отношением с ограниченным доменом, которое символически изображается предикатом, определяющим ограничение(в данном случае B), и обозначается следующим образом:

$$BR_2 = R_3. \tag{5.13}$$

Заметим, что формальное определение отношения с ограниченным доменом записывается так:

$$x_F R_y = F_x \wedge_x R_y. \tag{5.14}$$

Теперь отношение R_2 приближено к отношению W из примера файла с точностью до порядка следования аргументов. Это приводит к необходимости обращения отношения, которое в русском языке выражают страдательным залогом. А поэтому отношение «написаны» имеет обращение «написал». В качестве другого примера укажем, что отношение «быть ребенком» является обращением «быть родителем». Определим теперь строго операцию обращения отношения как

$$x\breve{R}y = yRx, \tag{5.15}$$

отсюда получаем, что

$$R_2 = \breve{R}_1, \tag{5.16}$$

и, наконец,

$$R_1 = \text{написал}. \tag{5.17}$$

Если отождествить R_1 с W файл-таблицы, то из (5.12), (5.13), (5.16) и (5.17) символизация предложения (5.17) выглядит как

$$x_B[\breve{W}]a. \tag{5.18}$$

Заменим символы выражениями русского языка:

$$\text{Какие } [(\text{написал})^{\smile}] \text{ Ю. Л. Ершов книги} \tag{5.19}$$

Общие правила символизации предложений можно реализовать в виде программы для ЭВМ. Программа анализирует входное предложение-запрос и применяет последовательность правил перезаписи. Такой анализ завершается, когда весь запрос представлен в символическом виде.

5.8. АЛГЕБРА И ИСЧИСЛЕНИЕ ПРЕДИКАТОВ

Общие сведения. Как уже отмечалось, для решения многих кибернетических задач требуется их логический анализ. Для автоматизации логического рассуждения необходим некоторый формальный язык, на котором можно формулировать посылки и делать верные логические выводы. В этом языке должна быть предусмотрена возможность описания интересующей нас задачи.

Исчисление предикатов первого порядка — это такая логическая система, с помощью которой можно выразить большую часть знаний, относящихся к математике, а также к естественному разговорному языку. Эта система содержит правила логического вывода, позволяющие делать верные логические построения новых утверждений, исходя из некоторого заданного множества утверждений. Благодаря своей общности и логической силе исчисление предикатов может претендовать на использование его для машинного построения умозаключений.

Язык, применяемый в исчислении предикатов, определяется его синтаксисом. Чтобы определить синтаксис, надо задать алфавит символов и правила соединения этих символов друг с другом в виде выражений, допустимых на этом языке. Один из важных классов выражений в исчислении предикатов — это класс правильно построенных формул (ППФ).

Обычно пользуются каким-либо языком для того, чтобы делать утверждения, касающиеся определенной предметной области (ПО). Отношения между языком и описываемой им областью определяются семантикой этого языка.

Говорят, что ППФ принимают значение T или F в зависимости от того, являются утверждения в этой области истинными или ложными. Приемы обращения в ППФ позволяют строить умозаключения, относящиеся к некоторой области и, следовательно, представляют интерес при создании процессов принятия решения, требующих такого умозаключения.

Ниже зададим синтаксис одного из вариантов языка исчисления предикатов. Затем покажем, как на этом языке можно делать утверждения, касающиеся описываемой им области.

Основные понятия. В исследуемых ранее формальных системах структура языка была связана с подразумеваемым смыслом языка, причем в первую очередь исследовались понятия, встречающиеся в математических аксиоматических системах. В математике, кибернетике и других науках наряду с высказываниями встречаются выражения, грамматически имеющие форму высказывания, но содержащие предметные переменные некоторых множеств. Такие выражения можно получить из любых высказываний, заменив в них обозначения предметов предметными переменными множеств, к которым принадлежат эти предметы. Если в их выражениях все предметные переменные снова заменить какими-либо элементами данных множеств, то опять получим высказывания. Например, предложение «2 — простое число» есть высказывание. Заменив в этом высказывании «2» предметным переменным n множества натуральных чисел, получим выражение «n — простое число». Грамматически оно имеет ту же форму, что и первое высказывание, но не является таковым. При замене предметного переменного n любым натуральным числом 1, 2, 3... построенное выражение будет снова обращаться в высказывание. Все подобные выражения называются предикатами или функциями-высказываниями. Таким образом, применяя понятия исчисления высказываний, можно аналогично выразить довольно сложные конкретные факты, однако при этом часто возникают трудности. Приведем пример такого сложного рассуждения.

Пример 5.11. Всякий друг Михаила есть друг Евгения. Петр не есть друг Евгения. Следовательно, Петр не есть друг Михаила. Для формального описания этого рассуждения введем специальные обозначения. Если запись $P(x)$ означает, что она обладает свойством P, то запись $\forall xP(x)$ обозначает утверждение «все x обладают свойством P». Далее, пусть символы m, e, p обозначают соответственно Михаил, Евгений, Петр, а запись $F(x, y)$ означает, что x есть друг y. Тогда приведенное рассуждение записывается как

$$\forall (F(x, m) \rightarrow F(x, e)),$$

или в других обозначениях:

$$\frac{\neg F(p, e)}{\neg F(p, m)}.$$

Вхождение переменной x в формулу $\forall x(F(x, m) \rightarrow F(x, e))$ считается связанным.

Запись $\exists xP(x)$ означает, что существует по крайней мере один предмет x, обладающий свойством P.

Пример 5.12. В формуле

$$\forall x (x = 5) \vee \exists x (x < 2) \forall x = 7$$

первые два вхождения не связаны со следующими двумя вхождениями, а последнее вхождение не связано с первыми четырьмя. Во избежание таких формул переименовывают переменные

$$\forall y (y = 5) \vee \exists z (z < 2) \vee x = 7.$$

В этой формуле переменная x входит свободно, а переменные y и z связаны кванторами общности и существования соответственно.

Необязательно выбирать оба символа $\forall$ и $\exists$; можно определить символ $\forall$ в терминах $\exists$. Вначале заметим, что запись $\exists x \neg (x=0)$ означает, что некоторое натуральное число не равно 0. Следовательно, запись $\neg \exists \neg (x=0)$ означает, что *нет натурального числа, не равного* 0, *т. е. каждое натуральное число равно* 0. Это означает то же, что и формула $\forall x(x=0)$. Рассуждая таким же образом, можно показать, что запись $\forall xP$ всегда означает то же, что и $\neg \exists x \neg P$; следовательно, можно определить $\forall xP$ как $\neg \exists \neg P$. Аналогично можно было бы определить $\exists xP$ как $\neg \forall x \neg P$.

Введем точное определение предиката. Для этого примем некоторые обозначения. Рассмотрим предикаты, определенные на n множествах: $M_1, M_2, \ldots, M_n (n=1, 2, 3\ldots)$. Предметные переменные множества M_1 обычно обозначают $x_1, x_1', x_1''\ldots$, его элементы $x_{01}, x_{11}, x_{21}\ldots$, аналогично предметные множества M_2 обозначают $x_2, x_2', x_2''\ldots$, а его элементы $x_{02}, x_{12}, x_{22}\ldots$ и т. д.

Назовем n-местным предикатом, определенным на множествах $M_1, M_2, \ldots, M_n$, выражение, содержащее предметные переменные данных множеств и обращающееся в высказывание при замене последних любыми элементами множеств $M_1, M_2, \ldots, M_n$ соответственно. Множества $M_1, M_2, \ldots, M_n$ называются *базисными множествами предиката*, их элементы — *аргументами предиката*: 1-м, 2-м, ..., n-м. Назовем n-местный предикат ($n \geqslant 2$), все базисные множества которого совпадают, *однородным*.

Обозначим n-местные предикаты, определенные на множествах $M_1, M_2, \ldots, M_n$, содержащих предметные переменные $x_1, x_2, \ldots, x_n$, через $A(x_1, x_2, \ldots, x_n)$, $B(x_1, x_2, \ldots, x_n)$, $C(x_1, x_2, \ldots, x_n)$.

Высказывание, в которое обращается предикат $A(x_1, x_2, \ldots, x_n)$ при замене предметных переменных $x_1, x_2, \ldots, x_n$ аргументами $x_{01}, x_{02}, \ldots, x_{0n}$, соответственно обозначается

$$A(x_{01}, x_{02}, \ldots, x_{0n}).$$

Возможен и другой, эквивалентный способ определения предиката. Напомним некоторые определения. Пусть A и B — множества; отображением из множества A в множество B называется сопоставление каждого объекта из множества A с некоторым объектом из множества B. Назовем n-й в A последовательность, состоящую из n (необязательно различных) объектов из множества A. Обозначим n-ку, состоящую из объектов $a_1, a_2, \ldots, a_n$, через $(a_1, a_2, \ldots, a_n)$. Отображение из множества n-ок в A в множество B называется n-арной функцией из A в B. Подмножество множества n-к в A называется n-арным предикатом на множестве A. Если P обозначает такой предикат, то $P(a_1, \ldots, a_n)$ будет обозначать, что n-ка $(a_1, \ldots, a_n)$ принадлежит P. Заметим, что унарная (одинарная) функция из A в B есть отображение из множества A в множество B, а унарный предикат на множестве A есть подмножество множества A. Приведем пример унарного предиката.

Пример 5.13. Пусть x-переменная, принимающая какое-либо значение в множестве действительных чисел. Тогда выражение $x>2$ есть одноместный (унарный) предикат, определенный на множестве действительных чисел. Заменив x числом 3, получим истинное высказывание: $3>2$, а заменив x числом 1, получим ложное высказывание: $1>2$.

Рассмотрим n-местный предикат $A(x_1, x_2, \ldots, x_n)$, определенный на множествах $M_1, M_2, \ldots, M_n$, где $x_{01}, x_{02}, \ldots, x_{0n}$ — некоторые элементы множеств $M_1, M_2, \ldots, M_n$. Значением предиката $A(x_1, x_2, \ldots, x_n)$ для аргументов $x_{01}, x_{02}, \ldots, x_{0n}$ называется логическое значение высказывания $A(x_{01}, x_{02}, \ldots, x_{0n})$, в которое обращается данный предикат при замене предметных переменных $x_1, x_2, \ldots, x_n$ аргументами $x_{01}, x_{02}, \ldots, x_{0n}$ соответственно.

Если значение предиката $A(x_1, x_2, \ldots, x_n)$ для аргументов $x_{01}, x_{02}, \ldots, x_{0n}$ есть истина, то говорят, что аргументы $x_{01}, x_{02}, \ldots, x_{0n}$ удовлетворяют данному предикату, в противном случае говорят, что аргументы $x_{01}, x_{02}, \ldots, x_{0n}$ не удовлетворяют данному предикату.

Пример 5.14. Относительно предиката $x>2$, рассмотренного в примере 5.13, можно сказать, что его значение для чисел 3 и 1 есть соответственно истина и ложь, а следовательно, число 3 удовлетворяет, а число 1 не удовлетворяет этому предикату.

Два n-местных предиката, определенных на одних и тех же множествах $M_1, M_2, \ldots, M_n$, называются равносильными, если их значения для любых аргументов совпадают, иначе говоря, если они удовлетворяются одними и теми же аргументами.

Пример 5.15. Пусть x — переменная, принимающая какое-либо значение в множестве действительных чисел. Тогда одноместные предикаты $x>2$ и $x-2>0$ равносильны.

Пусть $A(x_1, x_2, \ldots, x_n)$ и $B(x_1, x_2, \ldots, x_n)$ — два n-местных предиката, определенных на одних и тех же множествах $M_1, M_2, \ldots, M_n$. Тогда предикат $B(x_1, x_2, \ldots, x_n)$ называется *следствием* предиката $A(x_1, x_2, \ldots, x_n)$, если $B(x_1, x_2, \ldots, x_n)$ удовлетворяется любыми аргументами, удовлетворяющими $A(x_1, x_2, \ldots, x_n)$.

Пример 5.16. Одноместный предикат «число x делится на число 3», определенный на множестве целых чисел, есть следствие одноместного предиката «число x делится на число 6», определенного на том же множестве.

Назовем n-местный предикат, определенный на множествах $M_1, M_2, \ldots, M_n$, тождественно истинным, если значение его для любых аргументов есть истина; тождественно ложным, если значение его для любых аргументов есть ложь; выполнимым, если существует, по крайней мере, одна n-система его аргументов, для которой значение предиката есть истина.

Пример 5.17. Пусть x и y — переменные, принимающие какие-либо значения в множестве действительных чисел. Тогда однородный двухместный предикат $x+y=y+x$ будет тождественно истинным, одноместный предикат $x+1=x$ будет тождественно ложным, однородный двухместный предикат $x^2+y^2=5$ будет выполнимым, но не тождественно истинным.

Множеством истинности n-местного предиката, определенного на множествах $M_1, M_2, \ldots, M_n$, называется совокупность всех n-систем его аргументов, для которых значение предиката есть истина.

Множество истинности предиката $A(x_1, x_2, \ldots, x_n)$ обозначается символом $\lambda\, x_1, x_2, \ldots, x_n \cdot A(x_1, x_2, \ldots, x_n)$ [читается так: «Множество всех n-систем $(x_1, x_2, \ldots, x_n)$, таких, что $A(x_1, x_2, \ldots, x_n)$»].

Знак $\lambda x_1, x_2, \ldots, x_n$ называется символом абстракции по переменным $x_1, x_2, \ldots, x_n$.

Пример 5.18. Для рассмотренного ранее одноместного предиката $x>2$, определенного на множестве действительных чисел, множество истинности есть бесконечный интервал

$$(2, \infty), \text{ т. е. } \lambda x.\,(x>2)=(2, \infty).$$

Из только что приведенного определения видно, что множество истинности любого одноместного предиката, определенного на множестве M, есть подмножество множества M, множество истинности n-местного предиката, определенного на множествах $M_1, M_2, \ldots, M_n$ при $n \geqslant 2$, есть n-арное отношение между элементами множеств $M_1, M_2, \ldots, M_n$. Такое рассуждение объясняет связь двух определений предиката, приведенных ранее.

Напомним, что пропозициональная функция определялась как функция, принимающая значения в множестве, состоящем из двух элементов: истина и ложь. Пропозициональные функции часто определяются с помощью предикатов.

Пусть дан n-местный предикат $A(x_1, x_2, \ldots, x_n)$, определенный на множествах $M_1, M_2, \ldots, M_n$.

Пропозициональной функцией, соответствующей предикату $A(x_1, x_2, \ldots, x_n)$, называется функция, которая определена на множествах $M_1, M_2, \ldots, M_n$ и значение которой для аргументов $x_{01}, x_{02}, \ldots, x_{0n}$ есть значение данного предиката для этих же аргументов.

Пример 5.19. Для одноместного предиката $x>2$, определенного на множестве действительных чисел, соответствующей пропозициональной функцией будет функция, которая определена на множестве действительных чисел и значение которой для действительного числа есть логическое значение высказывания $x_0>2$. Так, для числа 1 ее значение ложь, а для числа 3 — истина.

Алгебра предикатов. Синтаксис и семантика. Простейшими логическими операциями над предикатами, так же как и для высказываний, являются отрицание, конъюнкция, дизъюнкция, импликация и эквивалентность, которые обозначаются с помощью тех же символов, что и для высказываний. При этом от содержания предикатов отвлекаются; они рассматриваются только с точки зрения их значений. Другими словами, равносильные предикаты не различаются. Заметим, что поскольку перечисленные операции не приводят к связыванию переменных, то уместно говорить об *алгебре предикатов* (см. § 5.8).

Отрицанием n-местного предиката $A(x_1, x_2, \ldots, x_n)$, определенного на множествах $M_1, M_2, \ldots, M_n$, называется новый n-местный предикат, определенный на тех же множествах, обозначаемый $\neg A(x_1, x_2, \ldots, x_n)$ [читается: *неверно, что* $A(x_1, x_2, \ldots, x_n)$] и имеющий значение «истина» для тех и только тех его аргументов $x_{01}, x_{02}, \ldots, x_{0n}$, для которых значение данного предиката $A(x_1, x_2, \ldots, x_n)$ есть «ложь».

Пример 5.20. Для одноместного предиката $x>2$, определенного на множестве действительных чисел, отрицанием будет одноместный предикат «x не больше двух», также определенный на множестве действительных чисел.

Конъюнкцией n-местного предиката $A(x_1, x_2, \ldots, x_n)$, определенного на множествах $M_1, M_2, \ldots, M_n$, и m-местного предиката $B(y_1, y_2, \ldots, y_m)$, определенного на множествах $N_1, N_2, \ldots, N_m$, называется новый $(n+m)$-местный предикат, определенный на множествах $M_1, M_2, \ldots, M_n; N_1, N_2, \ldots, N_m$ и обозначаемый $A(x_1, x_2, \ldots, x_n) \wedge B(y_1, y_2, \ldots, y_m)$. Он имеет значение «истина» для тех и только тех его аргументов $x_{01}, x_{02}, \ldots$ $\ldots, x_{0n}$ и $y_{01}, \ldots, y_{02}, \ldots, y_{0m}$, для которых значения соответственно предикатов $A(x_1, x_2, \ldots, x_n)$ и $B(y_1, y_2, \ldots, y_m)$ для аргументов $x_{01}, x_{02}, \ldots, x_{0n}$ и $y_{01}, y_{02}, \ldots, y_{0m}$ суть «истина».

Пример 5.21. Для одноместного предиката $x>2$, определенного на множестве действительных чисел, и двухместного предиката «точка M лежит на прямой l», определенного на множестве всех точек и множестве всех прямых плоскости, конъюнкцией будет трехместный предикат $x>2$ и «точка M лежит на прямой l», определенный на множествах действительных чисел, точек и прямых плоскости.

Дизъюнкцией n-местного предиката $A(x_1, x_2, \ldots, x_n)$, определенного на множествах $M_1, M_2, \ldots, M_n$, и m-местного предиката $B(y_1, y_2, \ldots, y_m)$, определенного на множествах $N_1, N_2, \ldots, N_m$, называется новый $(n+m)$-местный предикат, определенный на множествах $M_1, M_2, \ldots, M_n; N_1, N_2, \ldots, N_m$, обозначаемый $A(x_1, x_2, \ldots, x_n) \vee B(y_1, y_2, \ldots, y_m)$ (читается: $A(x_1, x_2, \ldots, x_n)$ или $B(y_1, y_2, \ldots, y_m)$) и имеющий значение «истина» для тех и только тех его аргументов $x_{01}, x_{02}, \ldots, x_{0n}$; $y_{01}, y_{02}, \ldots, y_{0m}$, для которых значение по крайней мере одного из предикатов $A(x_1, x_2, \ldots, x_n)$ или $B(y_1, y_2, \ldots; y_m)$ для аргументов $x_{01}, x_{02}, \ldots, x_{0n}$; $y_{01}, y_{02}, \ldots, y_{0m}$ соответственно есть «истина».

Операции конъюнкции и дизъюнкции можно применять также к предикатам, у которых имеются общие переменные. В таком случае число переменных конъюнкции и дизъюнкции будет равняться числу различных переменных этих предикатов. В частности, конъюнкция и дизъюнкция двух n-местных предикатов, зависящих от одних и тех же переменных, будет n-местным предикатом, зависящим от этих же переменных.

Можно доказать, что

$$\begin{gathered}
\lambda x_1 x_2 \ldots x_n. \\
[A(x_1, x_2, \ldots, x_n) \wedge B(x_1, x_2, \ldots, x_n)] = \lambda x_1 x_2 \ldots x_n. \\
A(x_1, x_2, \ldots, x_n) \cap \lambda x_1 x_2 \ldots x_n. \\
B(x_1, x_2, \ldots, x_n); \\
\lambda x_1 x_2 \ldots x_n. \\
[A(x_1, x_2, \ldots, x_n) \vee B(x_1, x_2, \ldots, x_n)] = \lambda x_1 x_2 \ldots x_n. \\
A(x_1, x_2, \ldots, x_n) \cup \lambda x_1 x_2 \ldots x_n. \\
B(x_1, x_2, \ldots, x_n).
\end{gathered}$$

Действительно, например, множество истинности конъюнкции $A(x_1, x_2, \ldots, x_n) \wedge B(x_1, x_2, \ldots, x_n)$ совпадает с множеством всех n-систем $(x_{01}, x_{02}, \ldots, x_{0n})$ ее аргументов, удовлетворяющих обоим данным предикатам, которое в то же время является пересечением множеств истинности данных предикатов $A(x_1, x_2, \ldots, x_n)$ и $B(x_1, x_2, \ldots, x_n)$.

Импликацией n-местного предиката $A(x_1, x_2, \ldots, x_n)$, определенного на множествах $M_1, M_2, \ldots, M_n$, и m-местного предиката $B(y_1, y_2, \ldots, y_m)$, определенного на множествах $N_1, N_2, \ldots, N_m$, называется предикат, определенный на множествах $M_1, M_2, \ldots, M_n; N_1, N_2, \ldots, N_m$, обозначаемый $A(x_1, x_2, \ldots$ $\ldots, x_n) \rightarrow B(y_1, y_2, \ldots, y_m)$ [читается: *если* $A(x_1, x_2, \ldots, x_n)$, то $B(y_1, y_2, \ldots, y_m)$] и имеющий значение «ложь» для тех и только тех его аргументов $x_{01}, x_{02}, \ldots, x_{0n}$; $y_{01}, y_{02}, \ldots, y_{0m}$, для которых значение предиката $A(x_1, x_2, \ldots, x_n)$ для x_{01},

$x_{02}, \ldots, x_{0n}$ есть «истина», а значение предиката $B(y_1, y_2, \ldots, y_m)$ для $y_{01}, y_{02}, \ldots, y_{0m}$ есть «ложь».

Эквивалентностью n-местного предиката $A(x_1, x_2, \ldots, x_n)$, определенного на множествах $M_1, M_2, \ldots, M_n$, и m-местного предиката $B(y_1, y_2, \ldots, y_m)$, определенного на множествах $N_1, N_2, \ldots, N_m$, называется предикат, определенный на множествах $M_1, M_2, \ldots, M_n; N_1, N_2, \ldots, N_m$, обозначаемый $A(x_1, x_2, \ldots, x_n) \leftrightarrow B(y_1, y_2, \ldots, y_m)$ [читается: $A(x_1, x_2, \ldots, x_n)$ *тогда и только тогда, когда* $B(y_1, y_2, \ldots, y_m)$] и имеющий значение «истина» для тех и только тех его аргументов $x_{01}, x_{02}, \ldots, x_{0n}; y_{01}, y_{02}, \ldots, y_{0m}$, для которых значения предикатов $A(x_1, x_2, \ldots, x_n)$ и $B(y_1, y_2, \ldots, y_m)$ для аргументов $x_{01}, x_{02}, \ldots, x_{0n}$ и $y_{01}, y_{02}, \ldots, y_{0m}$ соответственно совпадают.

Операции импликации и эквивалентности можно применять также к предикатам, у которых имеются общие переменные. В этом случае число переменных импликации и эквивалентности будет равно числу различных переменных этих предикатов. В частности, импликация и эквивалентность двух n-местных предикатов, зависящих от одних и тех же переменных, будет n-местным предикатом, зависящим от тех же переменных.

Пример 5.22. Для одноместных предикатов, определенных на множестве целых чисел, «x делится на шесть» и «x делится на три» импликацией будет тождественно истинный предикат, определенный на том же множестве: «если x делится на шесть, то x делится на три».

Определим синтаксис алгебры (и исчисления) предикатов, построив соответствующий язык (так называемый язык первого порядка), и семантику, установив интерпретацию алгебры (и исчисления) предикатов.

Язык первого порядка имеет следующие символы: предметные (индивидуальные) переменные x_i, n-арные функциональные символы и n-арные предикатные символы для каждого n, символы $\neg, \vee, \exists$.

Подобно тому как пропозициональные формы были использованы для выявления логической структуры, зависящей только от пропозициональных связок, можно и умозаключения, содержащие кванторы, представлять в абстрактной форме. Множество всех знаков в формализованном языке произвольной формализованной математической теории является объединением непересекающихся множеств знаков следующих категорий:

V — свободных индивидных переменных;

Ξ — связанных индивидных переменных;

Φ_m — m-местных функций (функторов);

P_m — m-местных предикатов;

L_1 — унарных пропозициональных связок и L_2 — бинарных пропозициональных связок;

Q — из двух знаков: кванторов общности и существования;

U — вспомогательных знаков, скобок и т. п.

Объединение всех множеств

$$V,\ \Xi,\ \Phi,\ P,\ L_1,\ L_2,\ Q,\ U \tag{5.20}$$

называется множеством знаков алфавита A.

Его элементы называются знаками алфавита A.

Под примитивным термом будем понимать любую конечную последовательность знаков из алфавита A вида

$$\Phi(x_1, \ \ldots, x_m), \tag{5.21}$$

где Φ обозначает m-местный функтор ($m=1, 2\ldots$), а $x_1, \ldots, x_m$ — свободные переменные.

Множеством термов, образованных из знаков алфавита A, называется наименьшее множество T конечных последовательностей знаков из алфавита A, такое, что:

все свободные индивидные переменные и все 0-местные функторы (т. е. индивидные константы) принадлежат множеству T;

если τ — примитивный терм со свободными индивидными переменными $x_1, \ldots, x_m$ и $\tau_1, \ldots, \tau_m$ принадлежит множеству T, то результат подстановки $\tau_1, \ldots, \tau_m$ соответственно вместо $x_1, \ldots, x_m$ в τ также принадлежит множеству T.

Терм обычно обозначается буквой τ (иногда с индексами). Если в алфавите A нет функторов, то множество всех термов T совпадает с множеством свободных индивидных переменных из алфавита A.

Если при $m>0$ не существует m-местных функторов, то множество всех термов T является объединением множества свободных индивидных переменных и множества всех индивидных констант в алфавите A.

Под примитивной формулой будем понимать любую конечную последовательность знаков из алфавита A вида

$$\rho(x_1, \ \ldots, x_m),$$

где ρ обозначает m-местный предикат ($m=1, 2\ldots$), а $x_1, \ldots, x_m$ — свободные индивидные переменные.

Наименьшим множеством конечных последовательностей знаков из алфавита A называются следующие формулы:

Φ_1, если α — примитивная формула со свободными индивидными переменными, $x_1, \ldots, x_m$, а $\tau_1, \ldots, \tau_m$ — термы, то результат подстановки термов $\tau_1, \ldots, \tau_m$ вместо $x_1, \ldots, x_m$ соответственно в L принадлежит множеству F (в частности, все примитивные формулы принадлежат множеству F);

Φ_2, если α — принадлежит множеству F и $\circ$ — знак какой-нибудь унарной пропозициональной связи, то $\circ\alpha$ принадлежит F;

Φ_3, если α и β принадлежат множеству F и $\circ$ — знак какой-нибудь бинарной пропозициональной связки, то $\alpha\circ\beta$ также принадлежит F;

Φ_4, если $\alpha(x)$ (где x обозначает некоторую свободную индивидную переменную) принадлежат множеству F, то при любой связанной индивидной переменной ξ, которая не встречается в $\alpha(x)$, выражения $(\exists\xi)\alpha(\xi)$ и $(\forall\xi)\alpha(\xi)$ принадлежит F.

Формулы типа Φ_1 называются элементарными. Буквы α, β, γ используются для обозначения формул. Наконец, система $\mathcal{L}=A, I, F$ называется формализованным языком первого порядка, основанным на алфавите A.

Отметим следующие свойства термов и формул:

если τ — терм, $x_1, \ldots, x_n$ — свободные индивидные переменные и $\tau_1, \ldots, \tau_n$ — термы, то $\tau(x_1/\tau_1, \ldots, x_n/\tau_n)$ есть терм;

если α — формула, $x_1, \ldots, x_n$ — свободные индивидные переменные и $\tau_1, \ldots, \tau_n$ — термы, то $\alpha(x_1/\tau_1, \ldots, x_n/\tau_n)$ есть формула.

Определим интерпретацию исчисления предикатов. Пусть $\mathcal{L}=\{A, I, F\}$ — формализованный язык первого порядка; I — непустое множество. Сопоставим каждый m-местный функтор φ с отображением: $\varphi_I : I^m \to I$.

В частности, каждую индивидную константу φ сопоставим с фиксированным элементом $\varphi_I \in I$. Каждый m-местный предикат ρ сопоставим с m-местным отношением ρ_I в I. Каждое такое сопоставление I называют интерпретацией (псевдомоделью, каркасом) языка $\mathcal{L}$ на множестве I, так как оно однозначно определяет:

интерпретацию каждого терма τ в языке $\mathcal{L}$ как конкретного отображения $\tau_I : I\times I\times\ldots\times I \to I$ или, в случае индивидных констант, как конкретного элемента из множества I;

интерпретацию каждой формулы α в языке $\mathcal{L}$ как конкретной пропозициональной функции α_I, определенной на множестве I. Для получения отображения τ_I достаточно интерпретировать:

каждую свободную переменную в τ как переменную, пробегающую множество I;

каждый встречающийся в τ функтор φ как отображение φ_I.

Для построения α_I достаточно интерпретировать:

все свободные и связанные переменные, встречающиеся в α как переменные, пробегающие множество I;

каждый встречающийся в α функтор φ как отображение φ_I;

каждый встречающийся в α предикат ρ как отношение ρ_I;

все пропозициональные связки и кванторы как соответствующие связки и кванторы, обычно используемые в математике.

Пример 5.23. Рассмотрим язык

$$\mathcal{L} = \{\nu, \sigma, \mu, \rho, \pi\}.$$

Пусть α — формула

$$((\exists\xi)\,\rho\,(x\sigma\,(\xi\xi))) \to ((\exists\eta)\,\rho\,(x\mu\,(\sigma\,(\nu\nu)\,\eta)) \wedge \pi(\nu\chi),$$

где ν — свободная индивидная переменная, а ξ, η — связанные индивидные переменные.

Пусть I — множество всех положительных целых чисел и пусть $\mathcal{L}_I$ — следующая интерпретация языка $\mathcal{L}$ в I. Тогда α_I — это пропозициональная функция

$$(\exists\xi)\,x = (\xi + \xi) \to ((\exists\eta)\,(x = (1+1)\,\eta) \wedge (1 < x)).$$

Исчисление предикатов. Свободные и связанные переменные. В исчислении предикатов по сравнению с алгеброй предикатов появляются новые, дополнительные логические операции *квантификации*, которые делают его значительно богаче по содержанию, при этом, как и в случае простейших операций, предикаты рассматриваются только с точки зрения их значений, т. е. равносильные предикаты не различаются. Основными операциями квантификации являются *квантор общности* и *квантор существования*, которые называются *двойственными* друг для друга. Рассмотрим их сначала для одноместных предикатов.

Пусть $A(x)$ — одноместный предикат, определенный на множестве M. Универсальным высказыванием, соответствующим предикату $A(x)$, называется высказывание *«Каждый элемент множества M удовлетворяет предикату Ax»*, которое обозначается символом $(\forall(x)A(x)$ и считается истинным, если данный предикат тождественно истинный, и ложным в противном случае.

Символ $(\forall x)$ называется квантором общности по переменной x, его читают так: *для всех x* и *для каждого x*. Универсальное высказывание $(\forall x)A(x)$ читают кратко: *для всех x* $A(x)$ или *для каждого x* $A(x)$. Говорят, что высказывание $(\forall x)A(x)$ есть результат применения квантора общности к предикату $A(x)$.

Пример 5.24. Рассмотрим ППФ

$$P\,(a,\ b)$$

и следующую ее интерпретацию: D — конечное множество целых чисел 1, 2, ..., 99, 100; a — число 30; b — число 1; P — отношение «больше или равно».

При такой интерпретации ППФ утверждает, что $30>1$.

Это утверждение истинное, и наша ППФ при данной интерпретации имеет значение T.

Припишем значения 1, 2, ..., 100 константным буквам f_1^0, f_2^0, ..., f_{100}^0 соответственно. С каждой из этих букв можно сконструировать ППФ вида

$$P(f_i,\ f_i^0).$$

при этом каждая из ППФ при данной интерпретации имеет значение T.

Легко показать, что

$$(\forall x)\,A(x) \leftrightarrow [A(a_1) \wedge A(a_2) \wedge\ .\ .\ .\ \wedge A(Q_m)],$$

где $A(x)$ — одноместный предикат, определенный на множестве $M=\{a_1,\ a_2,\ \ldots,\ a_m\}$ (доказательство этого утверждения оставляем читателю в качестве упражнения). Таким образом, для предикатов, определенных на конечном множестве, операция квантора общности может быть выражена через конъюнкцию. Для предикатов, определенных на бесконечном множестве, это сделать невозможно. В этом случае операция квантора общности является существенно новой.

Пусть $A(x)$ — одноместный предикат, определенный на множестве M. *Экзистенциональным* высказыванием, соответствующим предикату $A(x)$, называется высказывание *«Существует элемент множества M, удовлетворяющий предикату $A(x)$»*, которое обозначается символом $(\exists x)A(x)$ и считается истинным, если предикат $A(x)$ выполнимый, и ложным в противном случае.

Символ $(\exists x)$ называется квантором существования по переменной x, его читают так: *существует x такой, что...* или *для некоторого x...* Экзистенциональное высказывание $(\exists x)(Ax)$ читают кратко: *существует x такой*, что $A(x)$ или *для некоторого x*, $A(x)$. Говорят, что высказывание $(\exists x)A(x)$ есть результат применения квантора существования к предикату $A(x)$.

Пример 5.25. Для предикатов $x>2$ и $x=x+1$, определенных на множестве действительных чисел, соответствующие им экзистенциональные высказывания будут: $(\exists\ x)(x>2)$ — существует действительное число, большее двух (истинное), и $(\exists\ x)(x=x+1)$ — существует действительное число, на единицу большее самого себя (ложное).

Легко показать, как и в случае квантора общности, что

$$(\exists x)\,A(x) \leftrightarrow [A(a_1) \vee (a_2) \vee\ .\ .\ .\ \vee A(a_m)].$$

Таким образом, для предикатов, определенных на конечном множестве, операция квантора существования может быть выражена через дизъюнкцию: Для предикатов, определенных на бесконечном множестве, это сделать невозможно, операция квантора существования тогда является существенно новой.

Заметим, что выражения $(\forall(x)A(x)$ и $(\exists x)$ (Ax) есть высказывания, а не предикаты, хотя в них присутствует предметная переменная множества M. Присутствие x здесь связано с принятым способом обозначений. Действительно, универсальное и экзистенциональное высказывания, соответствующие заданному предикату, выражаются без переменной. Поэтому переменная x, входящая в выражения $(\forall x)A(x)$ и $(\exists x)A(x)$, называется *связанной*, тогда как переменная x, входящая в предикат $A(x)$, называется *свободной*. Иногда требуется высказать утверждение, касающееся всех «пар» элементов из некоторой области или всех «троек» и т. п. Тогда используются несколько переменных и для каждой из них символ $\forall$. Так, ППФ, соответствующую утверждению «Для всех пар целых чисел, расположенных между 1 и 100, первое больше второго», можно записать в виде

$$(\forall x)\,(\forall y)\,P\,(x,\ y).$$

Очевидно, что эта ППФ (двойная конъюнкция) при указанной ранее интерпретации имеет значение F.

Вообще, пусть $A(x_1,\ x_2,\ \ldots,\ x_n)$ — n-местный предикат, определенный на множествах $M_1,\ M_2,\ \ldots,\ M_n$ $(n \geqslant 2)$. *Предикатом, полученным из $A(x_1,\ x_2,\ \ldots,\ x_n)$ применением квантора общности (существования) по переменной x*, называется $(n-1)$-местный предикат, определенный на множествах $M_2,\ \ldots,\ M_n$ значением универсального (экзистенционального) высказывания, соответствующего одноместному предикату $A(x_{01},\ x_{02},\ \ldots x_{0n})$, получающемуся из данного n-местного предиката заменой переменных $x_2,\ \ldots,\ x_n$ аргументами $x_{02},\ \ldots,\ x_{0n}$ соответственно. Такой предикат обозначается $(\forall x_1)A(x_1,\ x_2,\ \ldots,\ x_n)$ или $(\exists x_1)A(x_1,\ x_2,\ \ldots,\ x_n)$ соответственно. К этим $(n-1)$-местным предикатам снова можно применить один из кванторов по любой свободной переменной, получая $(n-2)$-местные предикаты, и т. д. После n-кратного применения кванторов к n-местному предикату все его свободные переменные будут связаны и получится высказывание.

До сих пор предикаты противопоставлялись высказываниям: предикаты — это выражения, содержащие предметные переменные некоторых множеств, а высказывания — предложения, не содержащие никаких переменных. Однако высказывание можно рассматривать как предикат особого вида, с числом переменных, равным нулю, а его значение считать совпадающим с логическим значением данного высказывания.

На основании определений операций квантификации синтаксис исчисления предикатов представляет собой язык первого порядка, а алгебра предикатов получается из исчисления путем устранения операций квантификации. В данном случае множество знаков алфавита A составят

$$V,\ \Xi,\ \Phi,\ P,\ L_1;\ L_2,\ U.$$

Правила вывода исчисления предикатов. Любое рассуждение в математическом доказательстве состоит из простых шагов, каждый из которых заключается в опознавании того, что одни предложения являются непосредственными логическими следствиями других. Под *правилом вывода* понимают операцию, которая конечную последовательность формул $\alpha_1, \ldots, \alpha_n$ $(n \geqslant 1)$ формализованного языка сопоставляет с некоторой формулой β этого языка, причем в соответствии с интуитивными представлениями β является логическим следствием формул $\alpha_1, \ldots, \alpha_n$. Это правило можно записать как

$$\frac{\alpha_1, \ldots, \alpha_n}{\beta} \tag{5.22}$$

и читается оно так: β *в силу* $\alpha_1, \ldots, \alpha_n$. Формулы $\alpha_1, \ldots, \alpha_n$ называют *посылками*, а β — *заключением* правила.

Более точно: формула β в формированном языке $\mathscr{L}$ является *логическим следствием* множества формул S языка $\mathscr{L}$, если при каждой интерпретации I языка $\mathscr{L}$ в любом множестве $I \neq \Phi$. Из того, что все пропозициональные функции α_i, где $\alpha \in S$, истинны в множестве I, следует, что β_1 также истинна в I.

Пусть A_0 — множество формул в языке $\mathscr{L}$ и $R_1, \ldots, R_k$ правила вывода. Говорят, что формула α *выводима* из множества A_0 с помощью правил вывода $R_1, \ldots, R_k$, если существует последовательность формул

$$\alpha_1, \ldots, \alpha_n$$

такая, что:

α совпадает с α_n;

для каждого $i \leqslant n$ α_i либо принадлежит множеству A_0, либо является непосредственным следствием из некоторых формул $\alpha_1, \ldots, \alpha_{i-1}$ в силу одного из правил вывода $R_1, \ldots, R_k$.

Последовательность $\alpha_1, \ldots, \alpha_n$ называется тогда *формальным доказательством* формулы α из множества A_0 с помощью правил вывода $R_1, \ldots, R_k$.

Пример 5.26. Возьмем в качестве аксиом формулы

$$(P \vee P) \to P; \tag{5.23}$$

$$P \to (P \vee P_1); \tag{5.24}$$

$$(P \vee P_1) \to (P_1 \vee P); \tag{5.25}$$

$$(P \to P_1) \to [(P_2 \vee P) \to (P_2 \vee P_1)], \tag{5.26}$$

причем $F_1 \to F_2$ рассматривается как сокращение формулы $\neg F_1 \vee F_2$, а в качестве правил вывода используются правило modus poneus и правило подстановки. Тогда последовательность, состоящая из следующих восьми формул, может служить примером доказательства:

$$P \to (P \vee P_1); \tag{5.27}$$

$$P \to (P \vee P); \tag{5.28}$$

$$(P \vee P) \to P; \tag{5.29}$$

$$(P \to P_1) \to [(P_2 \vee P) \to (P_2 \vee P_1)]; \tag{5.30}$$

$$[(P \vee P) \to P] \to \{\neg P \vee (P \vee P) \to (\neg P \vee P)\}; \tag{5.31}$$

$$[\neg P \vee (P \vee P)] \to (\neg P \vee P); \tag{5.32}$$

$$[P \to (P \vee P)] \to (P \to P); \tag{5.33}$$

$$P \to P. \tag{5.34}$$

Формула (5.27) есть аксиома; формула (5.28) получается из (5.27), заменой P_1 на P; формула (5.29) есть аксиома (5.23); формула (5.30) — аксиома (5.26); формула (5.31) получается из (5.30) заменой P на $P \vee P$, P_1 на P и P_2 на $\neg P$; формула (5.32) есть непосредственное следствие формул (5.21) и (5.23) по правилу modus poneus; формула (5.33) есть формула (5.32), записанная сокращенно; наконец, формула (5.34) есть непосредственное следствие формул (5.28) и (5.33) по правилу modus poneus.

Предложения исчисления предикатов. Заметим, что термин *предложение* не совпадает с тем определением, которое дано в [30]. Точнее было бы говорить *дизъюнкт*, однако здесь будем пользоваться термином *предложение*, имея в виду следующее. Для того чтобы показать, что некоторое множество ППФ неудовлетворимо, надо доказать, что нет такой интерпретации, при которой каждая из ППФ в этом множестве имеет значение T. Хотя эта задача и кажется трудоемкой, существуют довольно эффективные процедуры ее решения. Для выполнения этих процедур требуется представить ППФ данного множества в специальном удобном виде — в виде предложений.

Любую ППФ исчисления предикатов можно представить в виде предложения, применив к ней последовательность простых операций. Задача состоит в том, чтобы показать, как придать произвольной ППФ форму *предложения*. Проиллюстрируем этот процесс на ППФ:

$$(\forall x)\{P(x) \to \{(\forall (y)\{P(y) \to P(f(x, y))\} \wedge \neg (\forall y)\{Q(x, y) \to \\ \to P(y)\}\}\} \tag{5.35}$$

Процесс преобразования ППФ в форму предложения состоит из следующих этапов.

1. Исключение знаков импликации. В форме предложения в исчислении предикатов явно используются лишь связки $\vee$

и $\neg$. Знак импликации можно исключить подстановкой в исходном утверждении вместо $A \rightarrow B$ $\neg A \vee B$. Такая подстановка дает возможность записать

$$(\forall x)\{\neg P(x) \vee \{(\forall y)\{\neg P(y) \vee P(f(x, y))\} \times$$
$$\times \vee \neg(\forall y)\{\neg Q(x, y) \vee P(y)\}\}\}. \qquad (5.36)$$

2. Уменьшение области действия знаков отрицания. Необходимо, чтобы знак отрицания $\neg$ применялся не более чем к одной предикатной букве. С помощью повторного применения указанных далее подстановок можно свести область действия каждого знака $\neg$ до отдельной предикатной буквы:

заменить $\neg(A \wedge B)$ на $\neg A \vee \neg B$;
заменить $\neg(A \vee B)$ на $\neg A \wedge \neg B$;
заменить $\neg\neg A$ на A;
заменить $\neg(\forall x) A$ на $(\exists x)\{\neg A\}$;
заменить $\neg(\exists x) A$ на $(\forall x)\{\neg A\}$.

Тогда ППФ (5.35) примет сначала вид

$$(\forall x)\{\neg P(x) \vee \{(\forall y) \neg P(y) \vee P(f(x, y))\} \times$$
$$\times \wedge (\exists y)\{\neg\{\neg Q(x, y) \vee P(y)\}\}\}, \qquad (5.37)$$

а затем

$$(\forall x)\{\neg P(x) \vee \{(\forall y)\{\neg P(y) \vee P(f(x, y))\} \wedge$$
$$\wedge (\exists y)\{Q(x, y) \wedge \neg P(y)\}\}\}. \qquad (5.38)$$

3. Стандартизация переменных. В области действия любого квантора переменная, связываемая им, является «немой» переменной. Поэтому везде в области действия квантора ее можно заменить другой переменной, а это не приведет к изменению значения истинности ППФ. Стандартизация переменных в ППФ означает переименование «немых» переменных, с тем чтобы каждый квантор имел свою собственную «немую» переменную. Так, вместо $(\forall x)\{P(x) \rightarrow (\exists x) Q(x)\}$ следует написать $\forall x \rightarrow \{P(x) \rightarrow (\exists y) Q(y)\}$.

Стандартизация в этом примере дает запись вида

$$(\forall x)\{\neg P(x) \vee \{(\forall y)\{\neg P(y) \vee P(f(x, y))\} \wedge$$
$$\wedge (\exists w)\{Q(x, w) \wedge \neg P(w)\}\}\}. \qquad (5.39)$$

4. Исключение кванторов существования. Рассмотрим ППФ

$$(\forall y \exists x) P(x, y),$$

которую можно интерпретировать, например, так: *для всех y существует такой x (возможно, зависящий от y), что x больше* y. Заметим, что, поскольку квантор существования $(\exists x)$ находится внутри области действия квантора всеобщности $(\forall y)$, допускается, что значение x может зависеть от значения y. Пусть эта зависимость определяется явно с помощью некоторой функции $g(y)$, отображающей каждое значение y в x, который «существует». Такая функция называется *функцией Сколема*. Если вместо x, который «существует», взять функцию Сколема, то можно исключить квантор существования:

$$(\forall y) P(g(y), y).$$

Общее правило исключения из ППФ квантора существования состоит в замене в ней всюду переменной, относящейся к квантору существования, функцией Сколема, аргументами которой служат переменные, относящиеся к тем кванторам всеобщности, области действия которых охватывают область действия исключаемого квантора существования. Функциональные буквы для функций Сколема должны быть «новыми» в том смысле, что они не должны совпадать с теми буквами, которые уже имеются в ППФ.

Так, исключая $(\exists z)$ из ППФ

$$\{(\forall w) Q(w)\} \rightarrow (\forall x)\{(\forall y)\{(\exists z)\{P(x, y, z) \rightarrow$$
$$\rightarrow (\forall u) R(x, y, u, z)\}\}\},$$

получаем

$$\{(\forall w) Q(w)\} \rightarrow (\forall x)\{(\forall y)\{P(x, y, g(x, y)) \rightarrow$$
$$\rightarrow (\forall u) R(x, y, u, g(x, y))\}\}.$$

Если исключаемый квантор существования не принадлежит области действия ни одного из кванторов всеобщности, то функция Сколема не содержит аргументов, т. е. является просто константой. Так, ППФ $(\exists x) P(x)$ превращается в $P(a)$, где a — константа, про которую известно, что она «существует».

Чтобы исключить все переменные, относящиеся к кванторам существования, надо применить описанную ранее процедуру по очереди к каждой переменной. В нашем примере исключение кванторов существования в формуле (5.39) (здесь лишь один такой квантор) дает ППФ вида

$$(\forall x)\{\neg P(x) \vee \{(\forall y)\{\neg P(y) \vee Pf(x, y)\} \wedge$$
$$\wedge \{Q(x, g(x)) \wedge \neg P(g(x))\}\}\}, \qquad (5.40)$$

где $g(x)$ — функция Сколема.

5. Приведение к предваренной нормальной форме. На этом этапе уже не осталось кванторов существования, а каждый квантор всеобщности имеет свою переменную. Теперь можно перенести все кванторы всеобщности в начало ППФ и считать

областью действия каждого квантора всю часть ППФ, расположенную за ним. Про полученную таким образом ППФ говорят, что она имеет *предваренную нормальную форму*. Правильно встроенная формула в предваренной нормальной форме состоит из цепочки кванторов, называемой *префиксом*, и расположенной за ней формулы, не содержащей кванторов, и называемой *матрицей*. Предваренная нормальная форма для ППФ (5.40) имеет вид

$$(\forall x \forall y)\{\neg P(x) \lor \{(\neg P(y) \lor P(f(x, y))\} \land$$
$$\land \{Qx, g(x) \land \neg P(g(x))\}\}\}. \qquad (5.4l)$$

6. Приведение матрицы к конъюнктивной нормальной форме. Любую матрицу можно представить в виде конъюнкции конечного множества дизъюнкций предикатов и (или) их отрицаний.

Говорят, что такая матрица имеет конъюнктивную нормальную форму. Дадим примеры матриц в конъюнктивной нормальной форме:

$$\{P(x) \lor Q(x, y)\} \land \{P(w) \lor \neg R(y)\} \land Q(x, y);$$
$$P(x) \lor Q(x, y);$$
$$P(x) \land Q(x, y);$$
$$R(y).$$

Любую матрицу можно привести в конъюнктивную нормальную форму, применив несколько раз следующее правило:

Заменить $A \lor \{B \land C\}$ на $\{A \lor B\} \land \{A \lor C\}$.

После приведения матрицы ППФ (5.33) в конъюнктивную нормальную форму она примет вид

$$(\forall x \forall y)\{\{\neg P(x) \lor \neg P(y) \lor Pf(x, y)\} \land$$
$$\land \{\neg P(x) \lor Q(x, g(x))\} \land \{\neg P(x) \lor \neg P(g(x))\}\}. \qquad (5.42)$$

7. Исключение кванторов всеобщности. Так как все переменные ППФ должны быть связанными, то все оставшиеся на этом этапе переменные относятся к кванторам всеобщности. Так как порядок расположения кванторов всеобщности несуществен, то эти кванторы можно явным образом не указывать, условившись, что все переменные в матрице относятся к кванторам всеобщности. Таким образом, остается лишь матрица в конъюнктивной нормальной форме.

8. Исключение связок $\land$. Теперь можно исключить знак $\land$, заменив запись $A \land B$ двумя ППФ A и B. Результатом многократной замены будет конечное множество ППФ, каждая из которых представляет собой дизъюнкцию атомных формул и

(или) их отрицаний. Атомную формулу или ее отрицание называют *литералом*, а ППФ, состоящую лишь из дизъюнкций литералов,— предложением. Итак, каждая ППФ в рассмотренном множестве будет предложением.

Таким образом, ППФ (5.35) представляется теперь в виде следующих предложений:

$$\left.\begin{array}{l} \neg P(x) \lor \neg P(y) \lor P(f(x, y)); \\ \neg P(x) \lor Q(x, g(x)); \\ \neg P(x) \lor \neg P(g(x)). \end{array}\right\} \qquad (5.43)$$

Заметим, что в литералах предложения могут содержаться переменные, которые всегда следует считать относящимися к кванторам всеобщности. Если вместо переменных литерала подставляются выражения, не содержащие переменных, то получается так называемый константный частный случай этого литерала. Так, $Q(a, f(g(b)))$ — константный частный случай ППФ $Q(x, y)$.

Процесс, демонстрирующий то, что некоторое множество S ППФ неудовлетворимо (невыполнимо), начинается с превращения каждой ППФ из множества S в предложения. В результате возникает некоторое множество S предложений. Можно показать, что если множество S неудовлетворимо, то неудовлетворимо и множество S', и, наоборот, из неудовлетворимости множества S' вытекает неудовлетворимость множества S.

Пример использования исчисления предикатов. Продолжим обсуждение проблемы получения ответа на запрос к ЭВМ (см. § 5.7). В преобразованиях от (5.7) к (5.19) указан общий семантический анализ вопроса, заданного на естественном языке. Было установлено, что «написал» соответствует отношению «книга» — свойству, «Ю. Л. Ершов» — индивидуальности, «какие» — неизвестной величине. Однако получение ответа на запрос требует более глубокого анализа предложения, при этом следует учитывать два аспекта значения: расширение и содержание. Содержание одноместного предиката — это свойство, обозначаемое предикатом, а расширение представляет собой класс предметов, обладающих указанным свойством. Аналогично содержание двухместного предиката трактуется как отношение, обозначаемое предикатом, а расширение — это класс упорядоченных пар индивидуальностей, которые состоят в заданном отношении.

Тогда пример файла из табл. 5.3 задает применительно к рассматриваемому пространству рассуждения расширение некоторого «простого» предиката. По этой причине расширение W задано строками 1, 3, 5 и 6, а именно $\{(a, b), (a, c)(d, f)(e, f)\}$, расширение B — строками 2 и 7: $\{b, f\}$, расширение P — строкой 4: {*c*}. Теперь весь файл можно назвать расширенным файлом.

Для ответа на входной запрос требуется преобразование значения фраз на русском языке в списке в памяти ЭВМ. Другими словами, ответ следует искать в расширенном списке $B[W]$ (список пар) среди тех имеющихся входов на месте первого аргумента, которые соответствуют символу a, стоящему на втором месте.

Такое «вычисление» можно произвести над матрицами. В этом случае каждому оператору из символьного запроса ставится в соответствие некоторая операция над матрицей. Например, матрица, соответствующая обращению $\breve{W}$, получается из матрицы, соответствующей W, путем простой перестановки столбцов. Матрица для ограничения домена, предположим R на B, получается вычеркиванием тех строк в матрице R, для которых первый элемент не содержится в матрице B.

Далее, при разборе запроса «написал» можно отождествить с «написаны», а следовательно, и с W. Это содержательная сторона проблемы. Реляционное выражение (отношение) следует хранить для указания синонимии (равнозначности) двух выражений. Точно так же следует хранить логические определения типа (5.14) для ограничения домена и (5.15) для обращения. В этом случае файл называют содержательным. Сам словарь может рассматриваться как часть содержательного файла.

Синонимия представляет собой простой пример общего класса представлений значений, называемых постулатами значений. Рассмотрим следующий запрос:

С. КЛИНИ соавтор Р. ВЕСЛИ? (5.44)

В этом предложении использован предикат «соавтор» (быть соавтором). Проблема состоит в сопоставлении «соавтор» с «написал». Это выполняется посредством получения относительного произведения W со своим обращением. Относительное произведение двух отношений R и S, символически R/S, строго определяется выражением

$$x\,(R/S)_y = (\exists z)\,(xRz \wedge zSy), \tag{5.45}$$

где $\exists z$ — квантор существования. Поэтому справедливо, что

$$\text{соавтор} = W/\breve{W}. \tag{5.46}$$

Вопросы в символическом виде и множества значений. Легко заметить, что проблема ответа на запрос с помощью ЭВМ требует для своего решения обработки формулы исчисления предикатов. Формулы можно классифицировать по наличию свободных переменных. Например, можно привести выражение

$$B_x \wedge aW_x, \tag{5.47}$$

в которое входит свободная переменная; с другой стороны, в выражении

$$(\exists x)\,(dWx \wedge eW_x) \tag{5.48}$$

переменная x связана. Выражение (5.47) называют открытой сентенциальной формулой, а (5.48) — замкнутой сентенциальной формулой, или просто сентенцией.

Для сентенции множество значений образует истинное выражение, например, числовое: 1 — для истины и 0 — для лжи. Можно предположить, что информационная система состоит из базы данных и сентенции, которую следует обработать. Проблема заключается в вычислении множества значений.

Рассмотрим запрос

Кто не написал «Теорию нумераций»? (5.49)

В символическом виде это выглядит так:

$$\neg\,(xWb), \tag{5.50}$$

где $\neg$ — знак отрицания.

Должна ли в этом случае ЭВМ напечатать все фамилии из словаря, кроме Ю. Л. Ершова?

В более сложных случаях могут вообще появиться логические комбинации, ведущие к бессмыслице.

Так, дизъюнктивная формула

$$(xWb) \vee (yWb), \tag{5.51}$$

где $\vee$ — знак дизъюнкции, вызывает следующее рассуждение. Любая подстановка конкретного значения, делающая истинным один из компонентов формулы (5.5), делает истиной и всю формулу. Следовательно, (5.51) приводит к множеству значений, составленному из упорядоченных пар (соответствуют паре свободных переменных x), причем каждая такая пара имеет вид (a, α) либо (α, a) с любым допустимым дескриптивным именем вместо α. В качестве другого примера рассмотрим запрос

Все публикации Ю. Л. Ершова — книги? (5.52)

Символизация приводит к формуле

$$(\forall x)\,(aW_x \supset B_x). \tag{5.53}$$

Формула (5.45) обладает тем свойством, что любая подстановка вместо x, которая делает aW_x ложным, делает истинной и всю импликацию.

Можно избавиться от квантора общности, тогда (5.53) примет вид

$$\neg\,(\exists x)\,(aW_x - B_x),$$

где минус — знак теоретико-множественного значения.

Глава 6

НЕКЛАССИЧЕСКИЕ ЛОГИКИ И ТЕОРИЯ ДОКАЗАТЕЛЬСТВ

6.1. ОБЩИЕ СВЕДЕНИЯ

В гл. 5 рассмотрены классические математические логики в виде классического исчисления высказываний и классического исчисления предикатов. Можно рассматривать классическую логику как обобщение математической модели философской логики Аристотеля. Как уже указывалось, большинство задач кибернетики и особенно искусственного интеллекта шире модели «предпосылка—следствие», которая используется в этой логике и утверждает наличие, по крайней мере, третьей составляющей, которая существенно вмешивается в процесс получения следствия из посылки. Могут получаться противоречивые логики, если одно из таких следствий будет исключать другое (рис. 6.1).

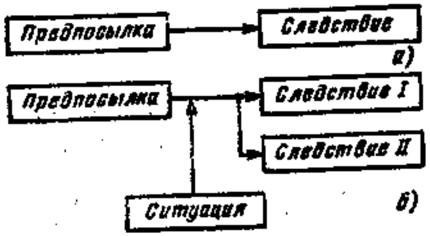


Рис. 6.1. Модели классической (*а*) и неклассической (*б*) логик

Кроме необходимости расширения классической логики имеется тенденция ее углубления за счет создания логики выше первого порядка (по сравнению с исчислением предикатов, которое является формализацией логики первого порядка). Третье направление в теории неклассических логик связано с изменением исходных положений, на которых строится логика, например логика нечетких множеств. Само разделение на классическую и неклассическую логики достаточно условно. Поэтому существует способ изложения (введения в проблемную область) неклассической логики с помощью исчисления секвенций (см. § 6.3). Это позволяет с более широких позиций рассмотреть классическую логику и сделать первые шаги в области логики выше первого порядка. Очень часто реакция кибернетической системы, особенно такой, как система искусственного интеллекта, получается путем доказательства непротиворечивости набора исходных положений. Поэтому инженеру знаний необходимо знание определенных результатов и методов теории доказательств. При доказательстве используются методы классической и неклассической математической логики. Поэтому эти два раздела математики в значительной степени пересекаются. В частности, для удобства доказательства математической логике придается форма исчисления. Напомним, что исчисление отличается от алгебры тем, что в нем имеются кванторы, которые связывают переменные, в алгебре все переменные свободны. В заключение приведем некоторую классификацию математических логик (рис. 6.2), которая не претендует на полноту.

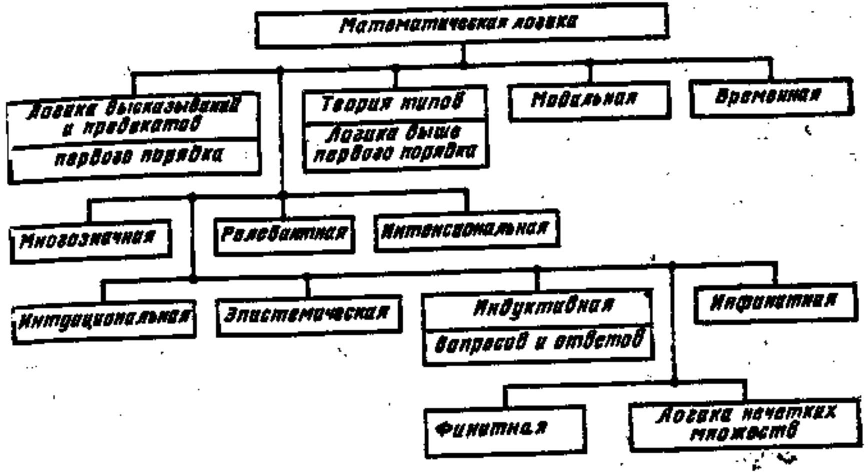


Рис. 6.2. Классификация математических логик

Считается, что различных логик существует более семидесяти [54—77]. На рис. 6.2 приведены только те, которые в том или ином виде используются в кибернетике и системах искусственного интеллекта. Некоторые из этих логик уже рассмотрены в предыдущих главах или будут рассмотрены в данной главе.

Специалисты высказывают серьезные соображения о том, что теория доказательств может быть положена в основания математики.

6.2. ТЕОРИЯ ДОКАЗАТЕЛЬСТВ НА БАЗЕ МЕТОДА РЕЗОЛЬВЕНТЫ

6.2.1. Общие сведения

Процедуры поиска доказательства по Эрбрану или методу резольвенты на самом деле являются процедурами опровержения, т. е. вместо доказательства общезначимости формулы доказывается, что отрицание формулы противоречиво. Таким образом, ищутся процедуры опровержения в общем виде, которые применяют к «стандартной форме» формулы (сколемовская стандартная форма), иногда называемой «предложением» [30, 61], при этом используются следующие идеи.

1. Формула логики первого порядка может быть сведена к *предваренной нормальной форме*, в которой матрица не содержит никаких кванторов, а префикс есть последовательность кванторов.

2. Поскольку матрица не содержит кванторов, она не может быть сведена к *конъюнктивной нормальной форме*.

3. Сохраняя противоречивость формулы, в ней можно устранить кванторы существования путем использования функций Сколема.

Считается, что формула *общезначима* (непротиворечива) тогда и только тогда, когда она истинна при всех возможных интерпретациях. Формула *необщезначима* (противоречива) тогда и только тогда, когда она не является общезначимой (непротиворечивой), т. е. истинной при всех интерпретациях.

Стандартная форма формулы F или *сколемова стандартная форма* получается из F исключением кванторов существования путем введения функций Сколема.

Дизъюнкт есть дизъюнкция литералов (или литер). Литерал (литер) есть атом или его отрицание. Если это удобно, будем *множество литералов* использовать как синоним *дизъюнкта*.

Пример 6.1. Дизъюнкции

$$\sim P(x,\ f(x)) \vee R(x,\ f(x),\ g(x)) \quad \text{и} \quad Q(x,\ g(x)) \vee R(x,\ f(x),\ g(x))$$

есть два дизъюнкта.

Теорема 6.1. Можно показать, что если S — множество дизъюнктов, которые представляют стандартную форму формулы F, то F противоречива в том и только в том случае, когда противоречиво S.

Справедливость этого положения основывается на том, что при переходе от формулы F к множеству дизъюнктов S, представляющих стандартную форму F, использовались корректные с точки зрения логики преобразования, не нарушающие смысл логических выражений. Далее, если S — стандартная форма формулы F, а F противоречива, то по теореме 6.1 $F=S$. Но если F непротиворечива, то в общем случае F неэквивалентна S.

Пример 6.2. Пусть $F \rightleftarrows (\exists x)P(x)$ и $S=P(a)$. (Знак $\rightleftarrows$ означает «равно по определению».) Очевидно, что S является стандартной формой (предложением) формулы F. Однако пусть I есть следующая интерпретация:

Область интерпретации: $D=\{1,\ 2\}$. Значения для a:

a
1

Значения для P:

$P(1)$.	$P(2)$
Л	И

Отсюда ясно, что F истинна в I, но S ложна в I. Поэтому $F \neq S$.

Одна и та же формула может иметь более чем одну стандартную форму (предложений).

Пример 6.3. Выразим следующую теорему в стандартной форме «Если $x \cdot x=e$ для всех x в группе G, где $\cdot$ есть бинарный оператор и e есть единица группы G, то G коммутативна».

Сначала эту теорему формализуем вместе с некоторыми основными аксиомами теории групп и затем представим отрицание этой теоремы множеством дизъюнктов. Нам известно, (см. гл. 3), что понятие группы G удовлетворяет следующим аксиомам:

$A_1: x,\ y \in G$ влечет $x \cdot y \in G$ (свойство замкнутости);

$A_2: x,\ y,\ z \in G$ влечет $x \cdot (y \cdot z) = (x \cdot y) z$ (свойство ассоциативности);

$A_3: x \cdot e = e \cdot x = x$ для всех $x \in G$ (свойство существования единичного элемента);

A_4: для каждого $x \in G$ существует элемент $x^{-1} \in G$, такой, что $x \cdot x^{-1} = x^{-1} \cdot x = e$ (свойство существования обратного элемента).

Пусть $P(x,\ y,\ z)$ обозначает $x \cdot y = z$ и $i(x)$ обозначает x^{-1}. Тогда вышестоящие аксиомы запишутся в виде

$$A_1': (\forall x)(\forall y)(\exists z)P(x,\ y,\ z);$$

$$A_2': (\forall x)(\forall y)(\forall z)(\forall u)(\forall v)\forall(w)(P(x,\ y,\ u) \wedge$$
$$\wedge P(y,\ z,\ v) \wedge P(u,\ z,\ w) \rightarrow P(x,\ v,\ w)) \wedge$$
$$\wedge(\forall z)(\forall x)(\forall y)(\forall u)(\forall v)(\forall w)(P(x,\ y,\ u) \wedge$$
$$\wedge P(y,\ z,\ v) \wedge P(x,\ v,\ w) \rightarrow P(u,\ z,\ w);$$

$$A_3': (\forall x)P(x,\ e,\ x) \wedge (\forall x)P(e,\ x,\ x);$$

$$A_4': (\forall x)P(x,\ i(x),\ e) \wedge (\forall x)P(i(x),\ x,\ e).$$

Заключение теоремы имеет вид:

B: Если $x \cdot x=e$ для всех $x \in G$, то G коммутативна, т. е. $u \cdot v=v \cdot u$ для всех $u,\ v \in G$. B может быть представлена формулой

$$B': (\forall x)P(x,\ x,\ e) \rightarrow ((\forall u)(\forall v)(\forall w)(P(u,\ v,\ w) \rightarrow P(v,\ u,\ w))).$$

Теперь вся теорема представляется формулой $F=A_1' \wedge \ldots \wedge A_4' \rightarrow B'$. Таким образом, $\sim F=A_1' \wedge A_2' \wedge A_3' \wedge A_4' \wedge B'$. Для получения множества дизъюнктов S для $\sim F$ сначала получим множество дизъюнктов S_i для каждой аксиомы A_i', $i=1,\ 2,\ 3,\ 4$, следующим образом:

$$S_1': \{P(x,\ y,\ f(x,\ y))\};$$

$$S_2': \{\sim P(x,\ y,\ u) \vee \sim P(y,\ z,\ v) \vee \sim P(u,\ z,\ w) \vee P(x,\ v,\ w),$$

$$\sim P(x, y, u) \vee \sim P(y, z, v) \vee \sim P(x, v, w) \vee P(u, z, w);$$
$$S_3': \{P(x, e, x), \quad P(e, x, x)\};$$
$$S_4': \{P(x, i(x), e), \quad P(i(x), x, e)\}.$$

Так как

$$\sim B' = \sim ((\forall x) P(x, x, e) \to ((\forall u)(\forall v)(\forall w) P(u, v, w) \to$$
$$\to P(v, u, w))) = (\sim (\forall x) P(x, x, e) \forall ((\forall u)(\forall v)(\forall w)(\sim$$
$$\sim P(u, v, w) \vee P(v, u, w)))) = (\forall x) P(x, x, e) \wedge \sim$$
$$\sim ((\forall u)(\forall v)(\forall w)(\sim P(u, v, w) \vee P(v, u, w))) =$$
$$= (\forall x) P(x, x, e) \wedge (\exists u)(\exists v)(\exists w)(P(u, v, w) \wedge \sim P(v, u, w)),$$

то множество дизъюнктов для B' приводится ниже:

$$T: \{P(x, x, e), \quad P(a, b, c), \quad \sim P(b, a, c)\};$$

Таким образом, множество $S = S_1 \cup S_2 \cup S_3 \cup S_4 \cup T$ является множеством, состоящим из следующих дизъюнктов:

1) $P(x, y, f(x, y))$;
2) $\sim P(x, y, u) \vee \sim P(y, z, v) \vee \sim P(u, z, w) \vee P(x, v, w)$;
3) $\sim P(x, y, u) \vee \sim P(y, z, v) \vee \sim P(x, v, w) \vee P(u, z, w)$;
4) $P(x, e, x)$;
5) $P(e, x, x)$;
6) $P(x, i(x), e)$;
7) $P(i(x), x, e)$;
8) $P(x, x, e)$;
9) $P(a, b, c)$;
10) $\sim P(b, a, c)$.

На этом заканчивается рассмотрение процедуры получения множества дизъюнктов S для формулы $\sim F$. Формула F общезначима тогда и только тогда, когда S противоречива: для доказательства используется процедура опровержения. Будем считать, что на входе процедуры опровержения поступает множество дизъюнктов S. Далее будет использоваться термин «выполнимо» («невыполнимо») вместо «противоречиво» («непротиворечиво») для множества дизъюнктов S.

6.2.2. Универсум Эрбрана

Рассмотрим классический метод Эрбрана для доказательства непротиворечивости множества дизъюнктов.

Эрбрановский универсум множества дизъюнктов. По определению множество дизъюнктов невыполнимо тогда и только тогда, когда оно ложно при всех интерпретациях на всех областях. Так как неудобно и невозможно рассматривать все интерпретации на всех областях, было бы удобно, если бы можно было фиксировать одну такую специальную область H, в которой S невыполнимо тогда и только тогда, когда S ложно при всех интерпретациях на этой области. Такая область существует, называется *эрбрановским универсумом* множества S и определяется следующим образом.

Определение. Пусть H_0 — множество констант, встречающихся в S. Если никакая константа не встречается в S, то H_0 состоит из одной константы, скажем, $H_0 = \{a\}$. Для $i = 0, 1, 2\ldots$ пусть H_{i+1} есть объединение H_i и множества всех термов вида $f^n(t_1, \ldots, t_n)$ (при всех n) для всех функций f^n, встречающихся в S, где t_j $(j = 1, 2, \ldots, n)$ принадлежит H_j.

Тогда каждое H_i называется *множеством констант* i-го уровня для S, а H_∞ называется *эрбрановским универсумом* S.

Пример 6.4. Пусть $S = \{P(a), \sim P(x) \vee P(f(x))\}$.

Тогда

$$H_0 = \{a\};$$
$$H_1 = \{a, f(a)\};$$
$$H_2 = \{a, f(a), f(f(a))\};$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$H_\infty = \{a, f(a), f(f(a)), f(f(f(a))) \cdots\}.$$

Пример 6.5. Пусть $S = \{P(x) \vee Q(x), R(z), T(y) \vee \sim W(y)\}$. Так как не существует никаких констант в S, то положим $H_0 = \{a\}$. Так как не существует никаких функциональных символов в S, то, следовательно, $H = H_0 = H_1 = \ldots = \{a\}$.

Пример 6.6. Пусть $S = \{P(f(x)), a, g(y), b\}$.

Тогда

$$H_0 = \{a, b\};$$
$$H_1 = \{a, b, f(a), f(b), g(a), g(b)\};$$
$$H_2 = \{a, b, f(a), f(b), f(f(a)), f(f(b)), f(g(b)), g(f(a)), g(f(b)), g(g(a)),$$
$$g(g(b))\}.$$

В дальнейшем под *выражением* будет пониматься терм, множество термов, множество атомов, литерал, дизъюнкт, множество дизъюнктов или, короче, множество предложений. Если в выражении не встречаются никакие переменные, оно будет называться *основным выражением*. В связи с этим будут использоваться термины «основной терм», «основной атом», «основной литерал» и «основной дизъюнкт». *Подвыражение* выражения E есть выражение, которое встречается в E.

Определение. Пусть S — множество дизъюнктов. Тогда множество основных атомов вида $P^n(t_1, \ldots, t_n)$ для всех n-местных предикатов P^n, встречающихся в S, где $t_1, \ldots, t_n$ — элементы эрбрановского универсума S, называется *множеством атомов* множества S или *эрбрановским базисом*.

Определение. *Основной пример* дизъюнкта C множества дизъюнктов S есть дизъюнкт, полученный заменой переменных в C на члены эрбрановского универсума S.

Пример 6.7. Пусть $S=\{P(x), Q(f(y))\vee R(y)\}$, $C=P(x)$ — дизъюнкт в S и $H=\{a, f(a), f(f(a)), \ldots\}$ — эрбрановский универсум S. Тогда $P(a)$ и $P(f(f(a)))$ суть основные примеры C.

Рассмотрим интерпретации над эрбрановским универсумом. Пусть S — множество дизъюнктов. Как указывалось выше, интерпретация над эрбрановским универсумом множества S есть значение констант, функциональных символов и предикатных символов, встречающихся в S. Далее дадим определение специальной интерпретации над эрбрановским универсумом, которая называется H-интерпретацией множества S.

Определение. Пусть S — множество дизъюнктов, H — эрбрановский универсум S и I — интерпретация S над H. Говорим, что I есть *H-интерпретация множества S*, если она удовлетворяет следующим условиям.

1) I отображает все константы из S в самих себя;

2) пусть f есть n-местный функциональный символ и $h_1, \ldots, h_n$ — элементы H. В I через f обозначается функция, которая отображает $(h_1, \ldots h_n)$ (элемент из H^n) в $f(h_1, .., h_n)$ (элемент из H), при этом не возникает никаких ограничений при придании значения любому n-местному предикатному символу в S. Пусть $A=\{A_1, A_2, \ldots, A_n, \ldots\}$ — эрбрановский базис множества S. H-интерпретацию удобно представить в виде

$$I=\{m_1, \quad . \ . \ . \ , \ m_n, \ . \ . \ .\},$$

где m_j есть или A_j, или $\sim A_j$ для $j=1, 2 \ldots$ Смысл этого множества состоит в том, что если m_j есть A_j, то атому A_j присвоено значение «истинно», в противном случае — значение «ложно».

Пример 6.8. Рассмотрим множество

$$S=\{P(x)\vee Q(x), \ \acute{R}(f(y))\}.$$

Эрбрановский универсум H для S есть $H=\{a, f(a), f(f(a)), \ldots\}$. В S входят три предикатных символа: P, Q и R. Следовательно, эрбрановский базис S

$$A=\{P(a), Q(a), R(a), P(f(a)), Q(f(a)), R(f(a)) \cdot \cdot \cdot\}.$$

Некоторые H-интерпретации множества S суть

$$I_1=\{P(a), Q(a), R(a), P(f(a)), Q(f(a)), R(f(a)) \cdot \cdot \cdot\};$$
$$I_2^*=\{\sim P(a), \sim Q(a), \sim R(a), \sim P(f(a)), \sim Q(f(a)), \sim R(f(a)) \cdot \cdot \cdot\};$$
$$I_3^*=\{P(a), Q(a), \sim R(a), P(f(a)), Q(f(a)), \sim R(f(a)) \cdot \cdot \cdot\}.$$

Интерпретацию множества дизъюнктов S не обязательно задавать над эрбрановским универсумом — интерпретация может не быть H-интерпретацией (интерпретацией Herbrand, Эрбрана). Пусть, например, $S=\{P(x), Q(y, f(y, a))\}$. Возможна следующая интерпретация над областью $D=\{1, 2\}$ (табл. 6.1).

Таблица 6.1

$D=\{1, 2\}$;

a	$f(1,1)$	$f(1,2)$	$f(2,1)$	$f(2,2)$	$P(1)$	$P(2)$	$Q(1,1)$	$Q(1,2)$	$Q(2,1)$	$Q(2,2)$
2	1	2	2	1	*И*	*Л*	*Л*	*И*	*Л*	*И*

Для такой интерпретации можно определить H-интерпретацию I^*, соответствующую I. Проиллюстрируем это на том же примере. Сначала находится эрбрановский базис S:

$$A=\{P(a), Q(a, a), P(f(a, a)), Q(a, f(a, a)),$$
$$Q(f(a, a), a), \quad Q(f(a, a), f(a, a)) \ . \ . \ .\}.$$

Затем оценивается каждый член A с использованием данных табл. 6.1:

$$P(a)=P(2)=\text{Л};$$
$$Q(a, a)=Q(2,2)=\text{И};$$
$$P(a, f(a, a))=P(f(2,2))=P(1)=\text{И};$$
$$Q(a, f(a, a))=Q(2, f(2, 2))=Q(2, 1)=\text{Л};$$
$$Q(f(a, a), a)=Q(f(2, 2), 2)=Q(1, 2)=\text{И};$$
$$Q(f(a, a), f(a, a))=Q(f(2, 2), f(2, 2))=Q(1, 1)=\text{Л}.$$

Следовательно, H-интерпретация I^*, которая соответствует I, есть

$$I^*=\{\sim P(a), Q(a, a), P(f(a, a)), \sim Q(a, f(a, a)),$$
$$Q(f(a, a), a), \sim Q(f(a, a), f(a, a)) \ . \ . \ .\}.$$

Если в S отсутствуют константы, то элемент a, который используется, чтобы начать эрбрановский универсум, может быть отображен в произвольный элемент области D. В этом случае, если D имеет более одного элемента, то существует более одной H-интерпретации, соответствующей I [61].

Определение. Пусть I — интерпретация на области D. *H-интерпретацией I^*, соответствующей I*, является интерпретация, которая удовлетворяет следующим условиям. Пусть $h_1, \ldots, h_n$ — элементы эрбрановского универсума H. Пусть каждый h_i отображается в некоторый d_i в D. Если $P(d_1, \ldots, d_n)$ получает в интерпретации I значение *И* (*Л*), то $P(h_1, \ldots, h_n)$ также получает значение *И* (соответственно *Л*) в I^*.

Лемма. Если интерпретация I на некоторой области D удовлетворяет множеству дизъюнктов S, то любая из H-интерпретаций I^*, соответствующих I, также удовлетворяет S.

Теорема 6.2. Множество дизъюнктов S невыполнимо тогда и только тогда, когда S ложно при всех H-интерпретациях в S [61].

Таким образом, мы показали, что необходимо рассматривать только интерпретации над эрбрановским универсумом (или более строго, H-интерпретации) для проверки того, выполнимо множество дизъюнктов или нет. Поэтому далее под интерпретацией будет пониматься H-интерпретация. Нетрудно убедиться в справедливости следующих утверждений.

1. Основной пример C' дизъюнкта C выполняется в интерпретации тогда и только тогда, когда существует основной литерал L' в C', такой, что L' есть также в I, т. е. $C' \cap I = \varnothing$ (где $\varnothing$ — пустое множество).

2. Дизъюнкт C выполняется в интерпретации I тогда и только тогда, когда каждый основной пример C выполняется в интерпретации I.

3. Дизъюнкт опровергается интерпретацией I тогда и только тогда, когда существует по крайней мере один такой основной пример C' для C, что C' не выполняется в I.

4. Множество дизъюнктов S невыполнимо тогда и только тогда, когда для каждой интерпретации I существует, по крайней мере, один такой основной пример C' некоторого дизъюнкта C в S, что C' не выполняется в интерпретации I.

Пример 6.9. 1. Рассмотрим дизъюнкт $C = \sim P(x) \vee Q(f(x))$. Пусть I_1, I_2 и I_3 определяются следующим образом:

$$I_1 = \{\sim P(a), \sim Q(a), \sim P(f(a)), \sim Q(f(a)), \sim P(f(f(a))), \sim Q(f(f(a))) \cdots\};$$

$$I_2 = \{(P(a), Q(a), P(f(a)), Q(f(a)), P(f(f(a))), Q(f(f(a))) \cdots\},$$

$$I_3 = \{P(a), \sim Q(a), P(f(a)), \sim Q(f(a)), P(f(f(a))), \sim Q(f(f(a))) \cdots\}.$$

Нетрудно убедиться, что C выполняется в интерпретациях I_1 и I_2, но опровергается в интерпретации I_3.

2. Рассмотрим $S = \{P(x), \sim P(x)\}$. Существуют только две H-интерпретации:

$$I_1 = \{P(a)\} \text{ и } I_2 = \{\sim P(a)\}.$$

Множество S опровергается двумя интерпретациями и, следовательно, невыполнимо.

6.2.3. Семантические деревья

Как изложено далее, нахождение доказательства для множества дизъюнктов эквивалентно построению семантического дерева.

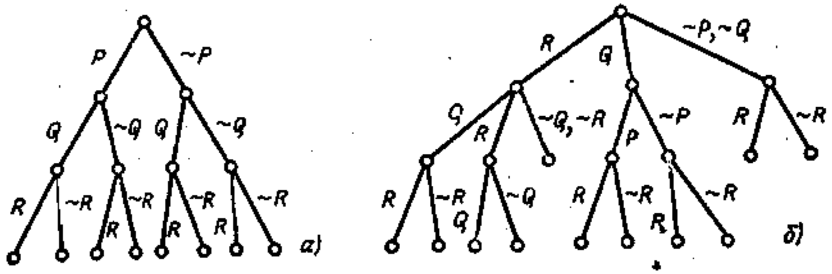


Рис. 6.3. Пояснение к примеру 6.10

Определение. Если A — атом, то говорят, что A и $\sim A$ *контрарны* друг другу, и множество $\{A, \sim A\}$ называется *контрарной парой*. Отметим, что дизъюнкт является тавтологией, если он содержит контрарную пару.

Определение. Пусть S — множество дизъюнктов и A — его эрбрановский базис. *Семантическое дерево* для S — это растущее вниз дерево T, в котором каждому ребру приписано конечное множество атомов или отрицаний атомов из A таким образом, что:

а) из каждого угла N выходит конечное число ребер $L_1, \ldots, L_n$. Пусть Q_i — конъюнкция всех литералов, приписанных к L_i, $i=1, \ldots, n$. Тогда $Q_1 \vee Q_2 \vee \ldots \vee Q_n$ — общезначимая пропозициональная формула;

б) пусть для каждого узла N $I(N)$ есть объединение всех множеств, приписанных ребрам ветви к N. Тогда $I(N)$ не содержит контрарных пар.

Определение. Пусть $A = \{A_1, A_2, \ldots, A_n, \ldots\}$ — эрбрановский базис множества S. Семантическое дерево для S будет *полным* тогда и только тогда, когда для каждого i $(i=1, 2, \ldots)$ и каждого конечного узла N семантического дерева (т. е. для узла, из которого не выходит никаких ребер) $I(N)$ содержит либо A_i, либо $\sim A_i$.

Пример 6.10. Пусть $A = \{P, Q, R\}$ — эрбрановский базис множества S. Тогда каждое из двух деревьев на рис. 6.3 является полным семантическим деревом для S.

Пример 6.11. Рассмотрим $S = \{P(x), P(a)\}$. Эрбрановский базис множества S есть $\{P(a)\}$. Полное семантическое дерево для S показано на рис. 6.4.

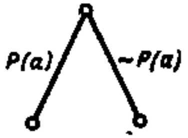


Рис. 6.4. Пояснение к примеру 6.11

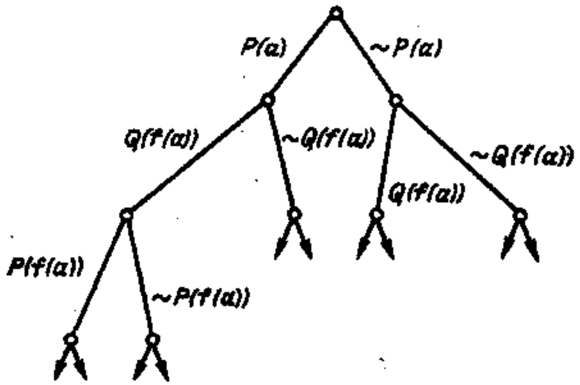


Рис. 6.5. Пояснение к примеру 6.12

Пример 6.12. Рассмотрим $S=\{P(x),\ Q(f(x))\}$. Эрбрановский базис множества S есть

$$\{P(a),\ Q(a),\ P(f(a)),\ Q(f(a)),\ P(f(f(a))),\ Q(f(f(a)))\ \cdot\ \cdot\ \cdot\}.$$

На рис. 6.5 представлено семантическое дерево для данного S. Для каждого узла N в семантическом дереве для S $I(N)$ является подмножеством некоторой интерпретации для S. Поэтому $I(N)$ будет называться *частичной интерпретацией* для S.

Для бесконечного эрбрановского базиса множеств S всякое полное семантическое дерево для S будет тоже бесконечно. Нетрудно убедиться, что полное семантическое дерево для S соответствует исчерпывающему перебору всех возможных интерпретаций для S. Если S невыполнимо, то оно не может быть истинным в каждой из этих интерпретаций. Таким образом, можно остановить рост дерева из узла N, если $I(N)$ опровергает S. Отсюда появляются следующие определения.

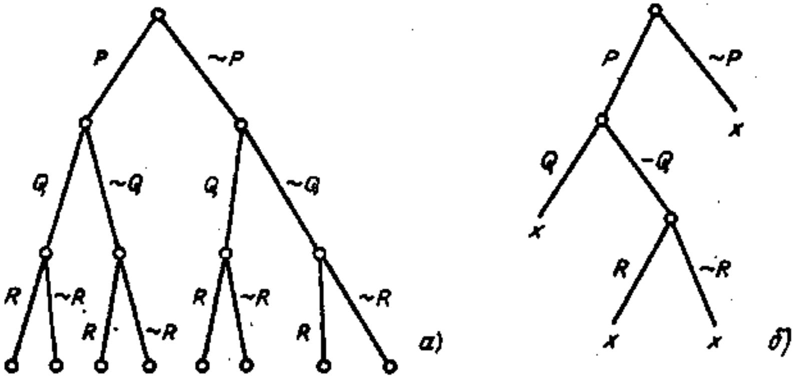
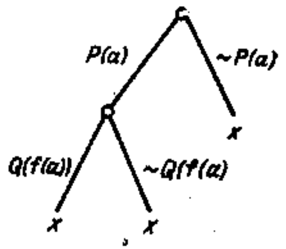


Рис. 6.6. Пояснение к примеру 6.13

Рис. 6.7. Пояснение к примеру 6.14



Определение. Узел N называется *опровергающим*, если $I(N)$ опровергает некоторый основной пример дизъюнкта в S, но для любого предшествующего N узла N' $I(N')$ не опровергает никакого основного примера дизъюнкта в S.

Определение. Считается, что семантическое дерево T будет *закрытым* тогда и только тогда, когда каждая ветвь T оканчивается опровергающим узлом.

Определение. Узел N закрытого семантического дерева называется *выводящим* узлом, если все непосредственно следующие за N узлы являются опровергающими.

Пример 6.13. Пусть $S=\{P,\ Q\vee R,\ \sim P\vee\sim Q,\ P\vee R\}$. Эрбрановский базис множества S есть $A=\{P,\ Q,\ R\}$. На рис. 6.6, *а* представлено полное семантическое дерево для S, а на рис. 6.6, *б* — закрытое семантическое дерево.

Пример 6.14. Рассмотрим $S=\{P(x),\ \sim P(x)\vee Q(f(x)),\ \sim Q(f(a))\}$. Эрбрановский базис множества S есть $A=\{P(a),\ Q(a),\ P(f(a)),\ Q(f(a))\ldots\}$.

На рис. 6.7 представлено семантическое дерево для S.

6.2.4 Теорема Эрбрана

Эта теорема является основой современных машинных алгоритмов доказательства теорем. Она тесно связана с упоминавшейся ранее теоремой 6.1, так как для проверки невыполнимости множества дизъюнктов необходимо рассмотреть только интерпретации над эрбрановским универсумом S. Если S ложно при всех интерпретациях над эрбрановским универсумом S, то можно заключить, что S невыполнимо.

Существует большое число таких интерпретаций, которые можно организовать в некоторую систему с помощью семантических деревьев. Есть два варианта теоремы Эрбрана.

Теорема 6.3 (теорема Эрбрана. I вариант). Множество дизъюнктов S невыполнимо тогда и только тогда, когда любому полному семантическому дереву S соответствует конечное замкнутое семантическое дерево. Это означает, что любая ветвь полного дерева ведет в опровергающий узел.

Доказательство. Допустим, что S невыполнимо. Пусть T — полное семантическое дерево для S. Для каждой ветви B

дерева T пусть I_B — множество всех литералов, приписанных всем ребрам ветви B. Тогда I_B есть интерпретация для S. Так как S невыполнимо, то I_B должна опровергать основной пример C' дизъюнкта C в S. Но так как C' конечно, то на B должен существовать опровергающий узел N_B (лежащий на конечном расстоянии от корня). Так как каждая ветвь T имеет опровергающий узел, то существует замкнутое семантическое дерево T' для S. Кроме того, так как из каждого узла в T' выходит только конечное число ребер, то T' должно быть конечным (т. е. число узлов в T конечно), иначе согласно лемме Кенига [61] мы могли бы найти бесконечную ветвь, не содержащую опровергающих узлов. Этим заканчивается первая часть доказательства (достаточность).

Наоборот, если для каждого полного семантического дерева T для S существует конечное закрытое семантическое дерево, то каждая ветвь T содержит опровергающий узел, т. е. каждая интерпретация опровергает S. Следовательно, S невыполнимо. Это завершает второе доказательство второй половины теоремы.

Теорема 6.4 (теорема Эрбрана, II вариант). *Множество дизъюнктов S невыполнимо тогда и только тогда, когда существует конечное невыполнимое множество S' основных примеров дизъюнктов S.*

Доказательство. Предположим, что S невыполнимо. Пусть T — полное семантическое дерево для S. Тогда по теореме Эрбрана (вариант I) существует конечное закрытое семантическое дерево T', соответствующее T. Пусть S' — множество всех основных примеров дизъюнктов, которые опровергаются во всех опровергающих узлах T'; S' конечно, так как в T' конечное число опровергающих узлов. Так как S' ложно в каждой интерпретации S', то S' невыполнимо. Тем самым доказана достаточность, условно обозначаемая ($\Rightarrow$).

Для доказательства необходимости ($\Leftarrow$) предположим, что существует конечное невыполнимое множество S' основных примеров дизъюнктов в S. Так как каждая интерпретация I для S содержит интерпретацию I' множества S' и I' опровергает S', то I должна опровергать S'. Однако S' опровергается в каждой интерпретации I множества S', значит, S' опровергается в каждой интерпретации I множества S. Поэтому S опровергается в каждой интерпретации множества S', значит, S невыполнимо.

Пример 6.15. Пусть $S=\{P(x),\ \sim P(f(a))\}$. Это множество S невыполнимо. Следовательно, по теореме Эрбрана существует конечное невыполнимое множество S' основных примеров дизъюнктов множества S. Мы установили, что одно из этих множеств есть

$$S'=\{P(f(a)),\ \sim P(f(a))\}.$$

Пример 6.16. Пусть $S=\{\sim P(x)\vee Q(f(x),\ x),\ P(g(b)),\ \sim Q(y,\ z)\}$. Это множество S невыполнимо. Одно из невыполнимых множеств основных примеров дизъюнктов множества S есть

$$S'=\{\sim P(g(b))\vee Q(f(g(b)),\ g(b)),\ g(b),\ g(b),\ \sim Q(f(g(b)),\ g(b))\}.$$

6.2.5. Применение теоремы Эрбрана

Второй вариант теоремы Эрбрана предполагает существование процедуры опровержения, т. е. если надо доказать невыполнимость множества дизъюнктов и мы имеем машинную процедуру, которая умеет успешно порождать множества S_1', ..., S_n' ... основных примеров дизъюнктов из S и успешно устанавливать их невыполнимость, то эта процедура (в соответствии с теоремой Эрбрана) укажет нам такое конечное N, что S_N' невыполнимо. Имеет место машинная программа Гилмора, составленная в соответствии с этой процедурой, которая порождает множества S_0', S_1', ..., S_n' ..., где S_i' — множество всех основных примеров, полученных заменой переменных в S константами из H_i, которое является множеством констант i-го уровня для S. Каждая S_i' есть конъюнкция основных дизъюнктов, для которой легко доказать ее невыполнимость. У Гилмора использовался мультипликативный метод, в котором каждое порожденное множество S_i' приводилось к дизъюнктивной нормальной форме. После этого любая конъюнкция в дизъюнктивной нормальной форме, которая содержит контрарные пары, исключалась. При наличии некоторых пустых S_i' невыполнимость S_i считается установленной.

Пример 6.17. Рассмотрим

$$S=\{P(x),\ \sim P(a)\};$$

$$H_0=\{a\};$$

$$S_0'=P(a)\wedge\sim P(a)=T.$$

Тем самым доказано, что S невыполнимо. Здесь используется обозначение True (T) для формулы, которая всегда истинна и обозначение false (f) для формулы, которая всегда ложна.

Пример 6.18. Рассмотрим

$$S=\{P(a),\ \sim P(x)\vee Q(f(x)),\ \sim Q(f(a))\};$$

$$H_0=\{a\};$$

$$S_0'=P(a)\wedge(\sim P(a)\vee Q(f(a)))\wedge\sim Q(f(a))=(P(a)\wedge\sim P(a)\wedge\sim Q(f(a)))\vee$$

$$\vee(P(a)\wedge Q(f(a))\wedge Q(f(a))=T\vee T=T.$$

Тем самым доказано, что S невыполнимо.

Рис. 6.8. Состав метода Дэвиса я Патема

Следует заметить, что мультипликативный метод Гилмора неэффективен. Так, для малого множества, состоящего из 10 двухлитеральных основных дизъюнктов, существует 2^{10} конъюнкций. Поэтому был предложен более эффективный метод Дэвиса и Патема.

Метод Дэвиса и Патема. Пусть S — множество дизъюнктов. Метод состоит из четырех правил (рис. 6.8).

I. Правило тавтологии. Вычеркнем все тавтологичные основные дизъюнкты из S. Оставшееся множество S' невыполнимо тогда и только тогда, когда и S невыполнимо.

II. Правило однолитеральных дизъюнктов. Если существует единичный основной дизъюнкт L в S, то S' получается из S вычеркиванием тех основных дизъюнктов в S, которые содержат L. Если S' пусто, то S выполнимо. В противном случае строится множество S'' исключением из S вхождений $\sim L$. Множество S'' невыполнимо тогда и только тогда, когда и S невыполнимо. Отметим, что если $\sim L$ — единичный основной дизъюнкт, то при вычеркивании $\sim L$ он превратится в T.

III. Правило чистых литералов. Назовем литерал L в основном дизъюнкте из S чистым в S' тогда и только тогда, когда литерал $\sim L$ не появится ни в каком основном дизъюнкте из S. Если литерал L чистый в S, то вычеркнем все основные дизъюнкты, содержащие L. Оставшееся множество S' невыполнимо тогда и только тогда, когда и S невыполнимо.

IV. Правило расщепления. Если множество S можно представить в виде

$$(A_1 \vee L) \wedge \ . \ . \ . \wedge (A_m \vee L) \wedge (B_1 \vee \sim L) \wedge \ . \ . \ . \wedge (B_n \vee L) \wedge R,$$

где A_i, B_i и R свободны от L и $\sim L$, то получим множества (называемые *множествами расщепления*)

$$S_1 = A_1 \wedge \ . \ . \ . \wedge A_m \wedge R \quad \text{и} \quad S_2 = B_1 \wedge \ . \ . \ . \wedge B_n \wedge R.$$

S невыполнимо тогда и только тогда, когда $(S_1 \vee S_2)$ невыполнимо, т. е. и S_1 и S_2 невыполнимы.

Можно показать, что вышеприведенные правила обратимы, т. е. если первоначальное множество S невыполнимо, то оставшееся после применения одного из правил множество все еще невыполнимо, и наоборот [61].

Эти правила очень важны, что будет показано в дальнейшем, и имеют более широкую область применения. Приведем примеры использования этих правил.

Пример 6.19. Требуется показать, что

$$S = (P \vee Q \vee \sim R) \wedge (P \vee \sim Q) \wedge \sim P \wedge R \wedge U$$

невыполнимо:

1) $(P \vee Q \vee \sim R) \wedge (P \vee \sim Q) \wedge \sim P \wedge R \wedge U$;

2) $(Q \vee \sim R) \wedge (\sim Q) \wedge R \wedge U$ — на базе правила II с $\sim Q$;

3) $\sim R \wedge R \wedge U$ — на базе правила II с $\sim Q$;

4) $T \vee U$ — на базе правила II с $\sim R$.

Так как последняя формула содержит пустой дизъюнкт T, то S невыполнимо.

Пример 6.20. Показать, что $S = (P \vee Q) \wedge \sim Q \wedge (\sim P \vee Q \vee \sim R)$ выполнимо.

1) $(P \vee Q) \wedge \sim Q \wedge (\sim P \vee Q \vee \sim R)$;

2) $P \wedge (\sim P \vee \sim R)$ — на базе правила II с $\sim Q$;

3) $\sim R$ — на базе правила II с P;

4) F — на базе правила II с R.

Мы получили в конце пустое множество F. Последнее множество всегда выполнимо, поэтому S выполнимо.

Пример 6.21. Показать, что $S = (P \vee \sim Q) \wedge (\sim P \vee Q) \wedge (Q \vee \sim R) \wedge (\sim Q \vee \vee \sim R)$ выполнимо:

1) $(P \vee \sim Q) \wedge (\sim P \vee Q) \wedge (Q \vee \sim R) \wedge (\sim Q \vee \sim R)$;

2) $\sim (Q \wedge (Q \vee \sim R) \wedge (\sim Q \vee \sim R)) \vee (Q \wedge (Q \vee \sim R) \wedge (\sim Q \vee \sim R)$ — на базе правила IV с P;

3) $\sim R \vee \sim R$ — на базе правила II с $\sim Q$ и Q;

4) $F \vee F$ — на базе правила II с $\sim R$.

Так как оба множества расщепления выполнимы, то и S выполнимо.

Пример 6.22. Показать, что $S = (P \vee Q) \wedge (R \vee \sim Q) \wedge (R \vee Q) \wedge (R \vee \sim Q)$ выполнимо:

1) $(P \vee Q) \wedge (P \vee \sim Q) \wedge (R \vee Q) \wedge (R \vee \sim Q)$;

2) $(R \vee Q) \wedge (R \vee \sim Q)$ — на базе правила III с P;

3) F — на базе правила III с R.

Таким образом, S выполнимо.

Вышеизложенный метод более эффективен, чем мультипликативный метод. Метод применим к любой формуле в логике высказываний, которую требуется представить в конъюнктивной нормальной форме. Можно сравнить эти четыре правила с шестью правилами Генцена в секвенциальном исчислении.

Метод резолюций. Рассмотренная теорема Эрбрана и процедуры поиска вывода требуют порождения множеств S_1', S_2' ... основных примеров дизъюнктов, причем последовательность растет экспоненциально. Для пояснения рассмотрим пример

$$S=\{P(x,\ g(x),\ y,\ h(x,\ y),\ z,\ k(x,\ y,\ z)),$$
$$\sim P(u,\ v,\ e(v),\ w,\ f(v,\ w),\ x)\}.$$

Так как

$$H_0=\{a\};$$
$$H_1=\{a,\ g(a),\ h(a,\ a),\ k(a,\ a,\ a),\ e(a),\ f(a,\ a)\};$$

. . . .

то число элементов в S_0', S_1' ... есть 2, 15, 12 ... соответственно, причем самое первое невыполнимое множество $S_5'H_0H_5'$ имеет порядок 10^{64} элементов. Следовательно, S_5' имеет порядок 10^{256} элементов. При современном уровне вычислительной техники невозможно разместить в памяти S_5', не говоря уже о проверке выполнимости. Во избежание порождения множеств основных примеров (в методе Эрбрана) предлагается метод резолюций Робинсона, который может применяться к любому множеству дизъюнктов (не обязательно основных) в целях проверки его невыполнимости.

Основная идея метода резольвенты состоит в том, чтобы проверить, содержит ли S пустой дизъюнкт T. Если да, то S невыполнимо, если нет, то проверяется следующее положение: может ли быть получен T из S. По теореме Эрбрана (вариант I) проверка получения T эквивалентна подсчету числа узлов замкнутого семантического дерева для S (рис. 6.9). По теореме 6.3 S невыполнимо тогда и только тогда, когда существует конечное замкнутое семантическое дерево T для S. Ясно, что S содержит I тогда и только тогда, когда T состоит только из одного узла — корневого. Если S не содержит T, то T должно содержать более чем один узел. Однако если мы сможем свести число узлов в T к одному, то в конце концов T обязательно появится. В этом и состоит метод резолюций. Его можно рассматривать как специальное правило вывода, используемое для порождения из S новых дизъюнктов. Если добавить эти новые дизъюнкты в S, то некоторые узлы в исходном T окажутся опровергающими узлами. Поэтому число узлов в T может быть уменьшено, и пустой дизъюнкт T будет в итоге получен.

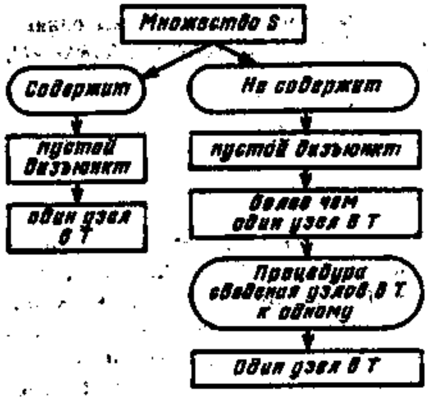


Рис. 6.9. Пояснение к методу резольвенты

6.2.6. Метод резолюций для исчисления высказываний

Можно рассматривать метод резолюций как обобщение *правила однолитеральных дизъюнктов* Дэвиса и Патема (см. п. 6.2.5).

Рассмотрим следующие дизъюнкты:

$$C_1: P;$$
$$C_2: \sim P\vee Q.$$

Используя правило однолитеральных дизъюнктов, из C_1 и C_2 получаем дизъюнкт

$$C_3: Q.$$

Правило однолитеральных дизъюнктов требуется нам, чтобы вначале определить, имеются ли контрарная пара литералов (например, P) в C_1 и литералы (например, $\sim P$) в C_2, затем вычеркнуть эту пару из C_1 и C_2, чтобы получить дизъюнкт C_3, который есть Q.

Обобщая вышеуказанное и применяя его к любой паре дизъюнктов (не обязательно единичных), получаем следующее правило, которое назовем *правилом резолюций*: для любых двух дизъюнктов C_1 и C_2, если существуют литерал H_1 в C_1, который контрарен литералу L_2 в C_2, то, вычеркнув L_1 и L_2 из C_1 и C_2 соответственно, построим дизъюнкцию оставшихся дизъюнктов. Построенный дизъюнкт есть резольвента C_1 и C_2.

Пример 6.23. Рассмотрим дизъюнкты

$$C_1: P\vee R;$$
$$C_2: \sim P\vee Q.$$

Дизъюнкт C_1 имеет литерал, который контрарен литералу $\sim P$ в C_2. Следовательно, вычеркивая P и $\sim P$ из C_1 и C_2 соответственно и составляя дизъюнкцию оставшихся дизъюнктов R и Q, получаем резольвенту $R\vee Q$.

Пример 6.24. Рассмотрим дизъюнкты

$$C_1: \sim P\vee Q\vee R;$$
$$C_2: \sim Q\vee S.$$

Резольвента C_1 и C_2 есть $\sim P\vee R\vee S$.

Пример 6.25. Рассмотрим дизъюнкты

$$C_1: \sim P\vee Q;$$
$$C_2: \sim P\vee R.$$

Так как не существует никакого литерала в C_1, которая контрарна какому-нибудь литералу в C_2, то не существует никакой резольвенты C_1 и C_2.

Теорема 6.5. Пусть даны два дизъюнкта C_1 и C_2. Тогда резольвента C дизъюнктов C_1 и C_2 есть логическое следствие C_1 и C_2.

Доказательство. Пусть C_1, C_2 и C означают следующее:

$$C_1 = L \vee C_1'; \qquad C_2 = \sim L \vee C_2 \quad \text{и} \quad C = C_1' \vee C_2',$$

где C_1' и C_2 — дизъюнкции литералов. Допустим, что C_1 и C_2 истинны в интерпретации I. Нужно доказать, что резольвента C дизъюнктов C_1 и C_2 также истинна в I. Для доказательства отметим, что L или $\sim L$ ложно в I. Допустим, что L ложно в I. Тогда C_1 не должен быть единичным дизъюнктом, иначе C_1 был бы ложен в I. Следовательно, C_1' должен быть истинен в I. Таким образом, резольвента C_1, т. е. $C_1' \vee C_2'$, истинна в I. Аналогично можно показать, что если $\sim L$ ложно в I, то C_2' должен быть истинен в I. Следовательно, $C_1' \vee C_2'$ должна быть истинна в I, что и требовалось доказать.

Если имеются два единичных дизъюнкта, то их резольвента, если она существует, есть пустой дизъюнкт T. Для невыполнимого множества дизъюнктов применением правила резолюций можно породить T.

Определение. Пусть S — множество дизъюнктов. *Резолютивный вывод* C из S есть такая последовательность $C_1, C_2, \ldots, C_k$ дизъюнктов, что каждый C_i или принадлежит S, или является резольвентой дизъюнктов, предшествующих C_i и $C_k = C$. Вывод T из S называется *опровержением* (или *доказательством невыполнимости* S).

Дизъюнкт может быть *выведен* или *получен* из S, если существует вывод C из S. Дадим несколько поясняющих примеров.

Пример 6.26. Рассмотрим множество

$$\left.\begin{array}{l} 1)\ \sim P \vee Q, \\ 2)\ \sim Q, \\ 3)\ P \end{array}\right\} S.$$

Из 1 и 2 получаем резольвенту

4) $\sim P$.

Из 4 и 3 можно получить резольвенту

5) T.

Так как T получается из S применением правила резолюции, то согласно теореме 6.5 T является логическим следствием S. Следовательно, S невыполнимо.

Пример 6.27. Для множества

$$\left.\begin{array}{l} 1)\ P \vee Q \\ 2)\ \sim P \vee Q \\ 3)\ P \vee \sim Q \\ 4)\ \sim P \vee \sim Q \end{array}\right\} S$$

породим следующие резольвенты:

5) Q из 1 и 2.

6) $\sim Q$ из 3 и 4.

7) T из 5 и 6.

Так как получен T, то S невыполнимо. Этот вывод представлен деревом на рис. 6.10, которое называется *деревом вывода*.

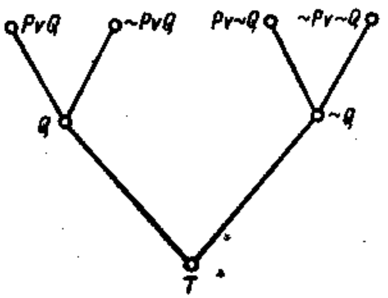


Рис. 6.10. Пояснение к примеру 6.27

Далее можно доказать полноту метода резолюций для установления невыполнимости множества дизъюнктов, т. е. множество S невыполнимо тогда и только тогда, когда существует резолютивный вывод пустого дизъюнкта T.

6.3. СЕКВЕНЦИАЛЬНОЕ ИСЧИСЛЕНИЕ

6.3.1. Общие сведения

Изложим генценовскую формализацию исчисления предикатов первого порядка, обозначаемую как LK [62].

При попытках автоматизации решения творческих задач (создания механизмов формирования решений в системах искусственного интеллекта) возникают различные дедуктивные системы (исчисления). Моделирование таких исчислений в рамках классической логики (CL или C — классическое исчисление предикатов) возможно, но может оказаться неадекватным исходной задаче, очень громоздким построением, превращающим простые прикладные задачи в практически нерешаемые. Поэтому математическая логика в последние годы характеризуется особенностью, связанной с нецелесообразностью сведения различных исчислений к логическим, а тем более к исчислению C. Более общая постановка проблемы автоматизации доказательств связана с проблемой *поиска вывода* в самых разнообразных исчислениях, к примеру интуиционистское исчисление предикатов и вообще дедуктивные системы нелогического типа.

Исчисление — это способ задания множества путем указания исходных элементов (*аксиом*) и *правил вывода*, каждое из которых описывает, как строить новые элементы из исходных, уже построенных. Список, каждый элемент которого или является аксиомой, или получен из предшествующих элементов списка по одному из правил вывода, называется *выводом* в исчислении (а его члены — *выводимыми*).

Допустим, имеется класс K творческих задач. В соответствии с особенностями этого класса подбираем исчисление C_k, аксиомами которого являются задачи с заранее известным решением, а правила вида

$$S_1, S_2, \ldots, S_m \vdash S_0 \quad (m \geqslant 0) \quad (A)$$

гарантируют решаемость задачи S_0 в случае решаемости задач $S_1, S_2, \ldots, S_m$. Теперь любая задача из класса K имеет решение тогда, когда она «выводима» в построенном исчислении. Считается, что найти решение задачи — это найти один из возможных ее выводов; найти «оптимальное» (экономное) решение — значит, найти короткий вывод.

Пример 6.28. Построить *исчисление для нахождения первообразных* в интегральном исчислении. Здесь аксиомами этого исчисления являются табличные интегралы. Правила соответствуют интегрированиям подстановкой и по частям (а также стандартным преобразованиям элементарных функций, вынесению множителей, интегрированию сумм). Этот пример дает некоторое представление творческого процесса как процесса вывода, при этом допускается сколько угодно неформальный процесс поиска, лишь бы результат этого процесса мог быть предъявлен достаточно формально. В связи с этим в математической логике появился раздел — теория поиска вывода, которая занимается определением по исчислению и объекту в языке исчисления структуры возможных выводов этого объекта или выявлением по гипотезе структуры ее возможных доказательств, при этом представляют интерес средства сужения класса возможных выводов и выявления дедуктивных методов, достаточных для доказательства данной гипотезы.

6.3.2. Основные понятия

При формализации логики прежде всего следует формализовать язык.

Определение. Формальный язык первого порядка состоит из следующих символов (знаков).

1. Константы:

1.1) индивидные

$$k_0, k_1, \ldots, k_j \ldots \quad (j=0, 1, 2 \ldots);$$

1.2) функциональные (i-местные)

$$f_0^i, f_1^i, \ldots, f_j^i \ldots (i=1, 2 \ldots; j=0, 1, 2 \ldots);$$

1.3) предикатные (i-местные)

$$R_0^i, R_1^i, \ldots, R_j^i \ldots (i=0, 1, 2 \ldots; j=0, 1, 2 \ldots)$$

2. Переменные:

2.1) свободные

$$a_0, a_1, a_2, \ldots, a_j \ldots (j=0, 1, 2 \ldots);$$

2.2) связанные

$$x_0, x_1, \ldots, x_j \ldots (j=0, 1, 2 \ldots).$$

3. Логические символы:

$\urcorner$ (не) (иначе $\sim$) $\wedge$ (и), или $\supset$ (влечет) (иначе $\Rightarrow$) $\forall$ (для всех) и $\exists$ (существует).

Первые четыре символа называются пропозициональными связками, а последние два — кванторами.

4. Вспомогательные символы:

левая скобка (, правая скобка) и запятая. Будем говорить, что задан некоторый язык первого порядка L, если заданы все его константы. Существенно, что каждое множество переменных бесконечно и существует, по крайней мере, одна предикатная константа. Другие множества могут иметь любую мощность и, в частности, могут быть пустыми. Верхние индексы в символах из 1.2 и 1.3 будем опускать, а символы из 1 и 2 будем употреблять и как метасимволы, и как формальные символы. Буквы g, $h \ldots$ будут использоваться в качестве символов для функциональных констант, буквы a, b, $c \ldots$ — для свободных переменных, а буквы x, y, $z \ldots$ — для связанных переменных. Всякая конечная последовательность символов (из языка L) называется выражением.

Определение. Термы определяются индуктивно (рекурсивно) следующим образом:

1) всякая индивидная константа есть терм;

2) всякая свободная переменная есть терм;

3) если f^i есть i-местная функциональная константа и $t_1, \ldots, t_i$ — термы, то $f^i(t_1, \ldots, t_i)$ также есть терм;

4) термами являются только те выражения, которые получаются согласно пунктам 1—3. Термы часто обозначаются буквами t, s, $t_1 \ldots$

Пример 6.29. Предложения «$x+1$ больше x» и «Отец Джона любит Джона» на формальном языке логики первого порядка запишутся в следующем виде:

БОЛЬШЕ (плюс (x, 1), x) и ЛЮБИТ (отец (Джон), Джон). Эти два выражения являются атомами логики первого порядка, причем БОЛЬШЕ и ЛЮБИТ — предикатные символы; x — переменная; Джон — индивидный символ или константа; отец и плюс — функциональные символы. Само по себе слово *терм* имеет смысл конечный и означает «то, что подставляется» вместо переменной в понятие «подстановка».

Определение. Если R_i есть i-местная предикатная константа и $t_1, \ldots, t_i$ термы, то выражение $R^i(t_1, \ldots, t_i)$ называется атомарной формулой. Формулы и их внешние логические символы определяются индуктивно следующим образом:

1) всякая атомарная формула есть формула, она не имеет внешних логических символов;

2) если A и B — формулы, то $(\urcorner A)$, $(A \wedge B)$, $(A \vee B)$ и $(A \supset B)$ — тоже формулы. Их внешними логическими символами являются $\urcorner$, $\wedge$, $\vee$, $\supset$ соответственно;

3) если A — формула, a — свободная переменная и X — связанная переменная, не входящая в A, то формулами будут $\forall x A'$, $\exists x A'$, где A' — выражение, полученное из A подстановкой X вместо переменной a в каждом ее вхождении в A. Их внешними логическими символами являются $\forall$ и $\exists$ соответственно;

4) формулами являются только те выражения, которые получены согласно пунктам 1—3.

Далее буквы A, B, C, ..., F, G ... будут метапеременными, обозначающими формулы. Формула без свободных переменных, называется замкнутой формулой или предложением. Формула, при определении которой не используется п. 3, называется бескванторной.

Выражение A' в п. 3 называется областью действия кванторов $\forall x$ и $\exists x$ соответственно. Иногда терм или формула языка L будет называться L-термом или L-формулой. В дальнейшем будут опускаться скобки всякий раз, когда смысл ясен и без них. В частности, будут опускаться внешние скобки. Введем следующее соглашение о приоритете связок: связка $\urcorner$ имеет приоритет перед каждой из связок $\wedge$ и $\vee$, а последние имеют преимущество перед связкой $\supset$. Таким образом, $\urcorner A \wedge B$ есть сокращение для $((\urcorner A) \wedge B)$, $A \wedge B \supset C \vee D$ — сокращение для $(A \wedge B) \supset (C \vee D)$, $\urcorner \urcorner A$ — сокращение для $\urcorner(\urcorner A)$. Через $A \equiv B$ будет обозначаться формула $(A \supset B) \wedge (B \supset A)$.

Определение. Пусть A — некоторое выражение, $\tau_1, \ldots, \tau_n$ — различные исходные символы и $\sigma_1, \ldots, \sigma_n$ — любые выражения. Через

$$\left(A \frac{\tau_1, \ldots, \tau_n}{\sigma_1, \ldots, \sigma_n}\right)$$

обозначим выражение, полученное из A заменой всех вхождений каждого из символов $\tau_1, \ldots, \tau_n$ на $\sigma_1, \ldots, \sigma_n$ соответственно, причем все эти символы заменяются одновременно. Такая операция называется (одновременным) замещением символов $\tau_1, \ldots$ ..., τ_n выражениями $\sigma_1, \ldots, \sigma_n$ в A, при этом не обязательно символы $\tau_1, \ldots, \tau_n$ должны входить в A.

Предложение. 1) Если ни один из символов $\tau_1, \ldots, \tau_n$ не входит в A, то выражение

$$\left(A \frac{\tau_1, \ldots, \tau_n}{\sigma_1, \ldots, \sigma_n}\right)$$

совпадает с A.

2) Если $\sigma_1, \ldots, \sigma_n$ — различные исходные символы, то выражение

$$\left(\left(A \frac{\tau_1, \ldots, \tau_n}{\sigma_1, \ldots, \sigma_n}\right) \frac{\sigma_1, \ldots, \sigma_n}{\theta_1, \ldots, \theta_n}\right)$$

совпадает с выражением

$$\left(A \frac{\tau_1, \ldots, \tau_n}{\theta_1, \ldots, \theta_n}\right).$$

Определение. 1) Пусть A — некоторая формула и $t_1, \ldots, t_n$ — термы. Если существуют формула B и n различных свободных переменных $b_1, \ldots, b_n$, таких, что A совпадает с

$$\left(B \frac{b_1, \ldots, b_n}{t_1, \ldots, t_n}\right),$$

то для всякого $i (1 \leqslant i \leqslant n)$ вхождение терма t_i, полученное в результате этого замещения, называется *отмеченным в A*, и этот факт можно выразить, записав формулу B как $B(b_1, \ldots, b_n)$, а формулу A как $B(t_1, \ldots, t_n)$. Конечно, A может содержать и другие вхождения терма t_i. Это происходит тогда, когда формула B содержит этот терм.

2) Будем говорить, что терм t *полностью отмечен* в формуле A (или *всякое вхождение t в A отмечено*), если каждое вхождение в A получено таким замещением (из некоторой формулы B при $n=1$ и $t=t_1$).

Следует заметить, что формула B и свободные переменные, по которым формула A может быть получена с помощью замещения, определяются неоднозначно; отмеченные вхождения термов в A задаются относительно конкретной формулы и конкретных свободных переменных.

Предложение. Если $A(a)$ — формула (в которой a отмечено необязательно полностью) и x — связанная переменная, не входящая в $A(a)$, то $\forall x A(x)$ и $\exists x A(x)$ — также формулы.

Пусть заглавные греческие буквы Γ, Δ, Π, Λ, Γ_0 Γ_1 ... обозначают конечные (возможные пустые) последовательности формул, разделенных запятыми. Чтобы описать исчисление секвенций, необходимо ввести вспомогательный символ $\rightarrow$.

Определение. Выражение вида $\Gamma \rightarrow \Delta$, где Γ и Δ — произвольные конечные последовательности формул, называется *секвенцией*, при этом Γ и Δ называются соответственно *антецедентом* и *сукцедентом* этой секвенции, а каждая формула в последовательности Γ или Δ называется секвенциальной формулой. Интуитивно секвенция $A_1, \ldots, A_m \rightarrow B_1, \ldots, B_n$ (где $m, n \geqslant 1$) означает следующее: если $A_1 \wedge \ldots \wedge A_m$, то $B_1 \vee \ldots \vee B_n$. При $m \geqslant 1$ секвенция $A_1, \ldots, A_m \rightarrow$ означает, что $A_1 \wedge \ldots \wedge A_m$ приводит к противоречию. При $n \geqslant 1$ секвенция $\rightarrow B_1, \ldots, B_n$ означает, что в рассматриваемой системе имеется противоречие. Секвенция будет обозначаться буквой S (с нижними индексами или без них). Секвенции не содержат свободных переменных.

6.3.3. Правила вывода

Определение. Всякую фигуру вида S_1/S или S_1S_2/S_2, где S_1, S_2 и S_3 — секвенции, будем называть *непосредственным выводом*. Секвенции S_1 и S_2 называются верхними секвенциями, а S — *нижней секвенцией* этого непосредственного вывода. Иначе, $S_1 \vdash S$; $S_1S_2 \vdash S$.

Интуитивно это означает, что если утверждается секвенция S_1 (утверждаются секвенции S_1 и S_2), то из нее (соответственно из них) можно вывести секвенцию S.

Напомним, что вывод в исчислении предикатов использует следующие правила вывода:

$$\frac{A,\ A\supset B}{B};\qquad \frac{A}{\forall xA},$$

где A, B — произвольные формулы; x — произвольная переменная. Первая формула называется «модус понենс» (modus ponens), вторая — *правилом обобщения*. Ограничимся рассмотрением непосредственных выводов, получаемых применением одного из следующих правил вывода (рис. 6.11), где A, B, C, D, $F(a)$ обозначают формулы.

1. *Структурные правила*:

1.1) ослабление

слева $\dfrac{\Gamma\to\Delta}{D,\ \Gamma\to\Delta}$; справа $\dfrac{\Gamma\to\Delta}{\Gamma\to\Delta,\ D}$;

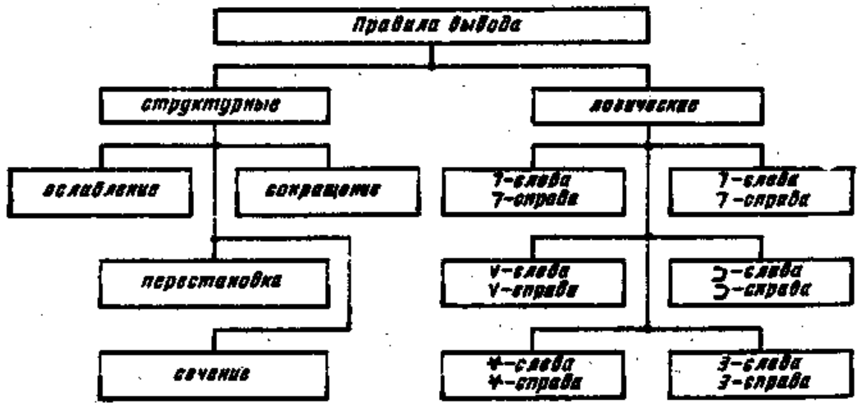


Рис. 6.11. Классификация правил вывода секвенциального исчисления

1.2) сокращение

слева $\dfrac{D,\ D,\ \Gamma\to D}{D,\ \Gamma\to\Delta}$; справа $\dfrac{\Gamma\to\Delta,\ D,\ D}{\Gamma\to\Delta,\ D}$;

1.3) перестановка

слева $\dfrac{\Gamma,\ C,\ D,\ \Pi\to\Delta}{\Gamma,\ D,\ C,\ \Pi\to\Delta}$; справа $\dfrac{\Gamma\to\Delta,\ C,\ D,\ \Lambda}{\Gamma\to\Delta,\ D,\ C,\ \Lambda}$.

Эти три правила вывода будем называть *слабыми* правилами, а все остальные — *сильными* правилами;

1.4) сечение

$$\frac{\Gamma\to\Delta,\ D\qquad D,\ \Pi\to\Lambda}{\Gamma,\ \Pi\to\Delta,\ \Lambda}.$$

Формула D называется *высекаемой формулой* этого правила. Это правило вывода является центральным в исчислении секвенций, так как связано с теоремой об устранении (правила) сечений. Теорема утверждает, что это правило допустимо в исчислении секвенций, т. е. если в исчислении секвенций выводимы посылки $\Gamma\to\Delta$, D и $D, \Pi\to\Lambda$ этого правила, то необходима и выводимая секвенция Γ, $\Pi\to\Delta$, Λ, которая является заключением рассмотренного правила. Теорема об устранении сечений утверждает, что все применения правила сечения можно затем устранить и получить вывод без сечений. Упрощенный вариант правила сечений имеет вид

$$\frac{\Gamma\to\Delta,\ D\qquad D,\ \Gamma\to\Delta}{\Gamma\to\Delta}.$$

2. *Логические правила*:

2.1) $\neg$-слева $\dfrac{\Gamma\to\Delta,\ D}{\neg D,\ \Gamma\to D}$; $\neg$-справа $\dfrac{D,\ \Gamma\to\Delta}{\Gamma\to\Delta,\ \neg D}$;

формулы D и $\neg D$ называются, соответственно, боковой и главной формулами этого правила вывода;

2.2) $\wedge$-слева $\dfrac{C,\ \Gamma\to\Delta}{C\wedge D,\ \Gamma\to\Delta}$ и $\dfrac{D,\ \Gamma\to\Delta}{C\wedge D,\ \Gamma\to\Delta}$;

$\wedge$-справа $\dfrac{\Gamma\to\Delta,\ C\qquad \Gamma\to\Delta,\ D}{\Gamma\to\Delta,\ C\wedge D}$;

формулы C и D называются *боковыми* формулами, а формула $C\wedge D$ — главной формулой этого правила;

2.3) $\vee$-слева $\dfrac{C,\ \Gamma\to\Delta D,\ \Gamma\to\Delta}{C\vee D,\ \Gamma\to\Delta}$;

$$\bigvee\text{-справа}\ \frac{\Gamma\to\Delta,\ C}{\Gamma\to\Delta,\ C\bigvee D}\quad \text{и}\quad \frac{\Gamma\to\Delta,\ D}{\Gamma\to\Delta,\ C\bigvee D};$$

формулы C и D называются *боковыми* формулами, а формула $C\bigvee D$ — главной формулой этого правила;

$$2.4)\ \supset\text{-слева}\ \frac{\Gamma\to\Delta,\ C\qquad D,\ \Pi\to\Lambda}{C\supset D,\ \Gamma,\ \Pi\to\Delta,\ \Lambda};$$

$$\supset\text{-справа}\ \frac{C,\ \Gamma\to\Delta,\ D}{\Gamma\to\Delta,\ C\supset D};$$

формулы C и D называются *боковыми* формулами, а формула $C\supset D$ — главной формулой этого правила.

Правила вывода 2.1—2.4 называются *пропозициональными*;

$$2.5)\ \forall\text{-слева}\ \frac{F(t),\ \Gamma\to\Delta}{\forall\ xF(x),\ \Gamma\to\Delta};$$

$$\forall\text{-справа}\ \frac{\Gamma\to\Delta,\ F(a)}{\Gamma\to\Delta,\ \forall xF(x)},$$

где t — произвольный терм и a не входит в нижнюю секвенцию. Формулы $F(t)$ и $F(a)$ называются *боковыми*, а $\forall xF(x)$ — *главной* формулой. Переменная a в правиле $\forall$-справа называется *собственной переменной* этого правила. Заметим, что в правиле $\forall$-справа все вхождения переменной a в $F(a)$ являются отмеченными. В правиле $\forall$-слева $F(t)$ и $F(x)$ означают

$$\left(F(a)\frac{a}{t}\right)\quad \text{и}\quad \left(F(a)\frac{a}{x}\right)$$

соответственно (для некоторой свободной переменной a), поэтому не всякий терм t в $F(t)$ обязательно отмечен;

$$2.6)\ \exists\text{-слева:}\ \frac{F(a),\ \Gamma\to\Delta}{\exists\ xF(x),\ \Gamma\to\Delta};$$

$$\exists\text{-справа:}\ \frac{F\to\Delta,\ F(t)}{\Gamma\to\Delta,\ \exists xF(x)},$$

где a не входит в нижнюю секвенцию и t — произвольный терм. Формулы $F(a)$ и $F(t)$ называются *боковыми*, а формула $\exists xF(x)$ — *главной*. Переменная a в правиле $\exists$-слева называется собственной переменной этого правила. Отметим, что в правиле $\exists$-слева переменная a полностью отмечена, тогда как в правиле $\exists$-справа не всякое t обязательно отмечено [и снова $F(t)$ есть $\left(F(a)\frac{a}{t}\right)$ для некоторого a].

Правила вывода 2.5 и 2.6 называются *кванторными*. Условие, что собственная переменная не должна входить в нижнюю секвенцию в правилах $\forall$-справа и $\exists$-слева, будем называть ограничением на собственную переменную для этих правил.

Всякая секвенция вида $A\to A$ называется начальной секвенцией или аксиомой.

Ввиду важности правил вывода Генцена [54, 67, 68] приведем другую их запись, при этом в рассмотрение введем всевозможные секвенции вида (которые играют роль аксиом)

$$\Gamma_1,\ \ F,\ \ \Gamma_2\Rightarrow\Gamma_3,\ \ F,\ \ \Gamma_4,$$

где F — произвольная формула. Правила вывода разделяются на шесть групп (см. рис. 6.11):

$$1.\ \left.\begin{array}{l}\Gamma_1,\ \Gamma_2\Rightarrow F\Gamma_3\\ \vdash\Gamma_1,\ \sim F,\ \Gamma_2\Rightarrow\Gamma_3\end{array}\right\}(\sim\Rightarrow),\quad \left.\begin{array}{l}\Gamma_1F\Rightarrow\Gamma_2,\ \Gamma_3\\ \Gamma_1\Rightarrow\Gamma_2,\ \sim F,\ \Gamma_3\end{array}\right\}(\Rightarrow\sim);$$

$$2.\ \left.\begin{array}{l}\Gamma_1,\ F_1,\ F_2\Gamma_3\Rightarrow\Gamma_3\\ \vdash\Gamma_1,\ F_1\bigwedge F_2,\ \Gamma_2\Rightarrow\Gamma_3\end{array}\right\}(\bigwedge\Rightarrow),\quad \left.\begin{array}{l}\Gamma_1\Rightarrow\Gamma_2,\ F_1,\ F_2,\ \Gamma_3\\ \vdash\Gamma_1\Rightarrow\Gamma_2,\ F_1\bigvee F_2,\ \Gamma_3\end{array}\right\}(\Rightarrow\bigvee);$$

$$3.\ \left.\begin{array}{l}\Gamma_1,\ F_1,\ \Gamma_2\Rightarrow\Gamma_3\\ \Gamma_1,\ F_2,\ \Gamma_2\Rightarrow\Gamma_3\\ \vdash\Gamma_1,\ F_1\bigvee F_2,\ \Gamma_2\Rightarrow\Gamma_3\end{array}\right\}(\bigvee\Rightarrow),\quad \left.\begin{array}{l}\Gamma_1\Rightarrow\Gamma_2,\ F_2,\ \Gamma_3\\ \Gamma_1\Rightarrow\Gamma_2,\ F_2,\ \Gamma_3\\ \vdash\Gamma_1\Rightarrow\Gamma_2,\ F_1\bigwedge F_2,\ \Gamma_3\end{array}\right\}(\Rightarrow\bigwedge);$$

$$4.\ \left.\begin{array}{l}\Gamma_1,\ F\theta,\ (\forall x)F,\ \Gamma_2\Rightarrow\Gamma_3\\ \vdash\Gamma_2,\ (\forall x)F,\ \Gamma_2\Rightarrow\Gamma_3\end{array}\right\}(\forall\Rightarrow),\quad \left.\begin{array}{l}\Gamma_1\Rightarrow\Gamma_2,F\theta,(\exists x)F,\Gamma_3\\ \vdash\Gamma_1\Rightarrow\Gamma_2,\ (\exists x)F,\ \Gamma_3\end{array}\right\}(\exists\Leftarrow)$$

(здесь θ — подстановка вида $\{t/x\}$, где t — произвольный основной терм);

$$5.\ \left.\begin{array}{l}\Gamma_1,\ F\theta,\ \Gamma_2\Rightarrow\Gamma_3\\ \vdash\Gamma_1,\ (\exists x)F,\ \Gamma_2\Rightarrow\Gamma_3\end{array}\right\}(\exists\Rightarrow),\quad \left.\begin{array}{l}\Gamma_1\Rightarrow\Gamma_2,\ F\theta,\ \Gamma_3\\ \vdash\Gamma_1\Rightarrow\Gamma_2,(\forall x)F,\ \Gamma_3\end{array}\right\}(\Rightarrow\forall)$$

(здесь θ — подстановка $\{a/x\}$ и a — функциональная константа, не встречающаяся в заключении правила);

$$6.\ \left.\begin{array}{l}\Gamma_1\Rightarrow\Gamma_2,\ F,\ \Gamma_3;\ \Gamma_4,\ F,\ \Gamma_5\Rightarrow\Gamma_6\\ \vdash\Gamma_1,\ \Gamma_4,\ \Gamma_5\Rightarrow\Gamma_2,\ \Gamma_3,\ \Gamma_6\end{array}\right\}\ \text{(правило сечения)}.$$

Можно доказать, что с помощью последнего правила построенное исчисление эквивалентно любой другой формализации классического исчисления предикатов первого порядка.

Последнее правило занимает особое место, которое будем называть *проблемой устранения сечений*. Чтобы определить по данному объекту S, выводим ли он в данном исчислении C, воспользуемся идеей поиска «снизу вверх». Так, по S определяем его возможные *непосредственные* предшественники (т. е. объекты, из которых S может быть выведен одним применением какого-либо правила исчисления C). Затем по каждому из получившихся объектов-предшественников, не являющемуся аксиомой исчисления, определяем свое множество непосредственных предшественников и т. д. (вплоть до построения окончательного вывода). Дерево, которое возникает в ходе такого процесса, называется *деревом поиска вывода*, оно превращается в *дерево вывода* в тот момент, когда все его «листья» оказались аксиомами исчисления C.

Если посмотреть в этой связи на приведенный вариант исчисления C, то можно установить, что все его правила хорошо приспособлены для нахождения непосредственных составляющих, за исключением правила сечения. Фундаментальный результат Генцена гарантирует *сохранение множества выводимых при удалении сечения из числа правил*. Правило называется допустимым, если его добавление не меняет множества выводимых объектов; результат Генцена означает допустимость сечения в исчислении с правилами $(\sim\Rightarrow)\div(\Rightarrow\forall)$. Правило сечения аналогично правилу modus ponens в методе резольвенты $A,\ A\to B\vdash B$. Заметим, что для секвенций стандартного типа (того самого, приведение к которому предполагается методом резолюций) можно из второй формы теоремы Эрбрана получить следующий частный случай теоремы Генцена [61].

Теорема 6.6. Множество дизъюнктов S невыполнимо тогда и только тогда, когда секвенция $S\Rightarrow$ может быть выведена в исчислении, единственными правилами которого являются

$$(\forall\Rightarrow),\quad (\vee\Rightarrow),\quad (\wedge\Rightarrow),\quad (\sim\Rightarrow).$$

S интерпретируется как конъюнкция всех дизъюнктов из S, где каждая переменная считается управляемой квантором всеобщности. Теорема доказывается путем процесса подъема «снизу вверх» с применением одного лишь правила $(\forall\Rightarrow)$ от $S\Rightarrow$ к секвенции следующего вида:

$$S,\ S_1,\ .\ .\ .\ ,S'_N\Rightarrow.$$

Теорема доказывает, что в состав $S\cup S_1\cup\ldots\cup S_N$ входит любой основной дизъюнкт из упомянутого в теореме Эрбрана множества S'. Противоречивость секвенции (Б) может быть установлена уже средствами одного лишь исчисления высказываний, для чего можно продолжать подъем, пользуясь правилами $(\vee\Rightarrow)$ и $(\wedge\Rightarrow)$. Процесс ветвится, но на каждой из ветвей мы должны прийти к секвенции вида $\Gamma_1,\ A,\ \Gamma_2,\ \sim A,\ \Gamma_3\Rightarrow$ или $\Gamma_1,\ \sim A,\ \Gamma_2,\ A,\ \Gamma_3\Rightarrow$. Однократное контрприменение правила $(\sim\Rightarrow)$ дает нам теперь аксиому. Нетрудно убедиться, что, несколько изменив вид аксиом, можно ограничиться односторонним («слева от стрелки») вариантом секвенциального исчисления, т. е. использовать односторонний двойственный («справа от стрелки») вариант такого исчисления.

Можно сказать, секвенциальное исчисление имеет более широкую область применения, чем теорема Эрбрана, связанная с методом резолюций, и в состоянии описывать даже логические системы выше первого порядка.

К настоящему времени для подавляющего большинства важнейших логических и логико-математических исчислений доказана устранимость правила сечений. При этом чаще всего удается построить исчисления со следующим *свойством подформульности:* в дереве вывода каждой формулы участвуют только ее подформулы. Для классического и интуиционистского исчислений (с равенством и без) устранение правила сечения, облегчая поиск вывода, может привести к резкому удлинению искомого вывода.

с-правила. Поиску «снизу вверх» противоречит наличие таких правил, у которых при фиксированном S_0 возможно бесконечное число различных наборов посылок $S_1, \ldots, S_m$. Такие правила будем называть *правилами типа сечения* или *с-правилами.* Проблема устранения *с*-правил актуальна для самых различных исчислений, в том числе и для нелогических. Так, в примере на исчисление первообразных можно говорить об устранении *с*-правил интегрирования произвольной подстановкой путем замены его разрешением производить подстановку лишь основных элементарных функций, через суперпозицию которых моделируется применение исходного правила.

Другим примером может служить «исчисление задач на построение циркулем и линейкой». Здесь вместо бесконечного числа разнообразных пополнений чертежа прямой или окружностью можно ограничиться проведением прямых через уже намеченные чертежом точки (их конечное число) и проведением окружностей с центрами в таких точках и с радиусами, равными расстояниям между такими точками. Возможность по любому «выводу» — построению указать «вывод», удовлетворяющий сформулированному ограничению, и имеет характер математической теоремы об устранимости *с*-правила. Иначе, если одно и то же выражение имеет место вследствие левой верхней части и в посылке правой верхней части правила вывода, то это выражение может быть исключено, «высечено» без изменения семантики вывода (аналогично тавтологиям $A\vee\sim A$ или вычеркиванию одинаковых по значению, но разных по знаку литералам в методе резольвенты). Более того, обычные логические исчисления после устранения сечений продолжают содержать *с*-правила. Действительно, таковы в исчислении *с*-правила $(\forall\Rightarrow)$ и $(\Rightarrow\exists)$, контрприменение которых допускает бесконечный перебор при варьировании терма t [фиктивность перебора при контрприменениях $(\Rightarrow\forall)$ и $(\exists\Rightarrow)$ очевидна]. В результате именно незнание термов выявляется в качестве основной трудности в организации поиска логического вывода. Из возможных путей преодоления этой трудности рассмотрим сначала наиболее естественный.

Правила $(\forall\Rightarrow)$ и $(\Rightarrow\exists)$ получили в логике название *минус-правил*, а перестройка выводов, приводящая к устранению бесконечного перебора значений t,— *минус-нормализации*. Можно показать, что без уменьшения множества вводимых в C секвенций можно ограничиться лишь такими применениями минус-правил, при которых в роли t выступают лишь термы, явно входящие в секвенцию-заключение правил (при отсутствии в заключении подходящих термов в роли t выступает одна и та же для всех применений фиксированная константа). Возможны и еще более жесткие ограничения области значений t, а также перенесение идеи минус-нормализации на другие логические исчисления [62].

В итоге из важнейших логических исчислений оказались устраненными все c-правила. Эта же идея может применяться и для нелогических исчислений. Именно такой характер имеет упомянутое выше устранение c-правила интегрирования подстановкой. При всей внешней эффективности минус-нормализации целесообразность ее использования во многих случаях весьма сомнительна. Действительно, если при контрприменении минус-правила мы «на самом деле» должны подставить терм, не входящий в заключение свободно, то, соблюдая ограничения, мы лишь оттянем получение искомого вывода. В этой ситуации ограничение перебора является фиктивным и лишь удлиняет поиск и ухудшает качество результата.

Таким образом, с устранением c-правил удлиняется результирующий вывод. Вместе с тем привлекательной их особенностью часто является возможность полного сохранения языка исходного исчисления, т. е. строгая эквивалентность исчислений без правил и их исходных вариантов. Но подобные методы неуниверсальны, так как с использованием весьма простых наборов дедуктивных средств можно строить исчисления, устранение c-правил из которых невозможно без изменения языка [61].

Однако стоит допустить некоторое хорошо интерпретируемое изменение языка и вида выводимых в исчислении объектов, как оказывается возможным полное устранение c-правил из любых исчислений при полном сохранении качества результирующего вывода.

6.3.4. Понятие вывода

Теперь введем понятие вывода, т. е. формального доказательства, в LK [62].

Определение. Выводом P (в LK) или LK-выводом называется дерево секвенций, удовлетворяющее следующим условиям:

1) самые верхние секвенции в P являются начальными секвенциями;

2) каждая секвенция в P, кроме самой нижней, является верхней секвенцией некоторого непосредственного вывода, нижняя секвенция которого тоже принадлежит P.

Выводы в исчислении секвенций будем записывать в виде деревьев. Высота вывода есть количество секвенций в наиболее длинной ветви вывода. Как правило, не будут различаться формулы и секвенции, отличающиеся лишь переименованием переменных, т. е. если имеется вывод с нижней секвенцией S, то этот вывод считается автоматически и выводом любой секвенции, которая получается из S переименованием связанных переменных.

Выводы в исчислении секвенций обладают замечательным свойством: секвенции, расположенные выше всякого правила вывода, состоят лишь из подформул формул в секвенций, расположенных в заключении правила вывода. Иначе, при рассмотрении правил вывода «снизу вверх» формулы в секвенциях лишь дробятся, никакие посторонние формулы не появляются. В этом и состоит так называемое свойство подформульности исчисления секвенций. Это обстоятельство в значительной степени облегчает поиск вывода данной секвенции. Основная цель данного пункта — показать, что поиск в исчислении предикатов эквивалентен выводу в исчислении секвенций, но при втором условии вывод обладает как бы бóльшим автоматизмом.

При рассмотрении выводов в LK будут использоваться понятия и соглашения, приведенные ниже.

Определение. Из предыдущего следует, что каждый вывод P имеет единственную секвенцию, расположенную ниже всех остальных. Она будет называться заключительной секвенцией вывода P. Вывод с заключительной S называется также выводом, оканчивающимся секвенцией S, или выводом секвенции S. Секвенция S называется выводимой в LK или LK-выводимой, если существует ее LK-вывод. Формула A называется LK-выводимой (или теоремой в LK), если секвенция $\rightarrow A$ является LK-выводимой. Приставка LK в выражениях LK-вывод и LK-выводима будет часто опускаться. Вывод, не содержащий правила сечения, называется выводом без сечений. Вводится специальный символ (рис. 6.12, *а*), который будет обозначать часть вывода. Например, фигуры на рис. 6.12, *б*, *в* обозначают соответственно некоторый вывод секвенции S и какой-то вывод секвенции S из секвенций S_1 и S_2. Выводы обычно обозначаются буквами P, Q... Запись типа $P(a)$ означает, что все вхождения переменной a в P отмечены. Такое обозначение целесообразно только в том случае, когда предполагается заменить эти вхождения переменной a некоторым другим термом.

Рис. 6.12. Пояснение к обозначению «часть вывода»

В таком случае $P(t)$ есть результат замены всех вхождений a в $P(a)$ термом t.

Рассмотрим некоторые слегка измененные правила вывода. Схема

$$J\ \frac{\Gamma\to\Delta, A \quad \Pi\to\Lambda B}{\Gamma, \Pi\to\Delta, \Lambda, A\wedge B}$$

не является правилом вывода системы LK. Тем не менее из двух верхних секвенций можно вывести в LK нижнюю секвенцию, используя несколько структурных правил и правило $\wedge$-справа:

$$\wedge\text{-справа}\ \frac{\dfrac{\dfrac{\Gamma\to\Delta, A}{\begin{array}{l}\text{несколько ослаблений}\\ \text{и перестановок}\end{array}}}{\Gamma, \Pi\to\Delta, \Lambda, A} \qquad \dfrac{\dfrac{\Pi\to\Lambda, B}{\begin{array}{l}\text{несколько ослаблений}\\ \text{и перестановок}\end{array}}}{\Gamma, \Pi\to\Delta, \Lambda, B}}{\Gamma, \Pi\to\Delta, \Lambda, A\wedge B}.$$

Наоборот, из секвенций $\Gamma\to\Delta, A$ и $\Gamma\to\Delta, B$ можно вывести секвенции $\Gamma\to\Delta, A\wedge B$, используя несколько структурных правил и некоторый пример схемы J:

$$\dfrac{\dfrac{J\ \dfrac{\Gamma\to\Delta, A \quad \Gamma\to\Delta, B}{\Gamma, \Gamma\to\Delta, \Delta, A\wedge B}}{\text{несколько сокращений и перестановок}}}{\Gamma\to\Delta, A\wedge B.}$$

Таким образом, можно рассматривать J как сокращение для (*). В таком случае будем использовать обозначение

$$\frac{\Gamma\to\Delta, A \quad \Pi\to\Lambda, B}{\overline{\Gamma, \Pi\to\Delta, \Lambda, A\wedge B}}.$$

В дальнейшем двойной чертой, как и в этом примере, будем указывать на сокращение некоторого числа шагов. Другое замечание, которое мы хотим здесь сделать, состоит в том, что ограничение на употребление связанных переменных мешает появлению таких нежелательных выводов, как, например

$$\dfrac{\dfrac{\dfrac{A(a), \quad B(b)\to A(a)\wedge B(b)}{A(a), \quad B(b)\to\exists x(A(x)\wedge B(b))}}{A(a), \quad B(b)\to\exists x\exists x(A(x)\wedge B(x))}}{\exists xA(x), \quad \exists xB(x)\to\exists x\exists x(A(x)\wedge B(x)).}$$

В нашей системе этого никогда не произойдет, так как выражение $\exists x\exists x(A(x)\wedge B(x))$ не является формулой.

Бескванторный фрагмент системы LK, т. е. ее подсистема, не содержащая кванторов, называется исчислением высказываний.

Пример 6.30. Следующие фигуры являются LK-выводами:

1)

$$\begin{array}{rl}
\neg\text{-справа} & \underline{A\to A}\\
\vee\text{-справа} & \underline{\to A, \neg A}\\
\text{перестановка справа} & \underline{\to A, A\vee\neg A}\\
\vee\text{-справа} & \underline{\to A\vee\neg A, A}\\
\text{сокращение справа} & \underline{\to A\vee\neg A, A\vee\neg A}\\
 & \quad\to A\vee\neg A
\end{array}$$

2) Пусть свободная переменная a полностью отмечена в $F(a)$;

$$\begin{array}{rl}
\exists\text{-справа} & \underline{F(a)\to F(a)}\\
\neg\text{-справа} & \underline{F(a)\to\exists xF(x)}\\
\forall\text{-справа} & \underline{\to\exists xF(x), \neg F(a)}\\
\neg\text{-слева} & \underline{\to\exists(x)F(x), \forall y\neg F(y)}\\
\supset\text{-справа} & \underline{\neg\forall y\neg F(y)\to\exists xF(x)}\\
 & \to\neg\forall y\neg F(y)\supset\exists xF(x)
\end{array}$$

Следует заметить, что нижняя секвенция применения правила $\forall$-справа не содержит собственной переменной a.

Определение. 1) Формулой, термом или логическим символом в данном выводе (в секвенции или в формуле) называется формула, терм или логический символ, рассматриваемый вместе с местом, которое он занимает в этом выводе (соответственно в секвенции или в формуле).

2) Последовательность секвенций в выводе называется нитью (вывода P), если выполняются следующие условия:

а) последовательность начинается с начальной секвенции и заканчивается заключительной секвенцией;

б) всякая секвенция в этой последовательности, кроме последней, является верхней секвенцией некоторого непосредственного вывода, и за ней непосредственно следует нижняя секвенция этого непосредственного вывода.

3) Пусть S_1—S_3 — секвенции в выводе P. Мы говорим, что S_1 расположена выше S_2 или что S_2 расположена ниже S_1 (в P), если существует нить, содержащая как S_1, так и S_2, и в этой нити S_1 появляется раньше, чем S_2. Если S_1 расположена выше S_2 и S_3 расположена выше S_3, то говорим, что S_2 расположена между S_1 и S_2.

4) Некоторый непосредственный вывод в P расположен ниже секвенции S (из P), если его нижняя секвенция расположена S.

5) Пусть P — некоторый вывод. Всякая часть P, которая сама является выводом, называется подвыводом вывода P. Это понятие можно определить следующим образом. Если S — произвольная секвенция в выводе P, то часть этого вывода, которая состоит из секвенции S и всех секвенций в P, расположенных выше S, называется подвыводом вывода P (с заключительной секвенцией S).

6) Пусть P_0 — некоторый вывод, изображенный на рис. 6.13, где (✱) обозначает часть вывода P_0, расположенную под $\Gamma \to \theta$. Если Q есть некоторый вывод, оканчивающийся секвенцией Γ, $D \to \theta$, то копией вывода P_0 относительно Q назовем вывод P (рис. 6.14), где (✱✱) отличается от (✱) только тем, что всякой секвенции (✱) вида $\Pi \to \Lambda$ в (✱✱) соответствует секвенция вида Π, $D \to \Lambda$. Иначе, P получается из P_0 заменой подвывода, оканчивающегося секвенцией $\Gamma \to \theta$, на Q и добавлением формулы D к антецеденту каждой секвенции (из ✱). Аналогичным образом можно определить и копию с дополнительной формулой в сукцеденте. Можно распространить это определение также и на случай нескольких дополнительных формул. Точно копию вывода P_0 относительно Q можно определить индукцией по числу применений правил вывода в (✱). Но мы это делать не будем, так как интуитивно это и так понятно.

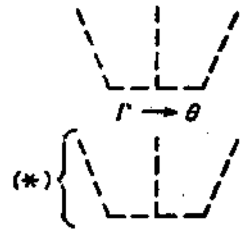


Рис. 6.13. Пояснение к выводу P_0

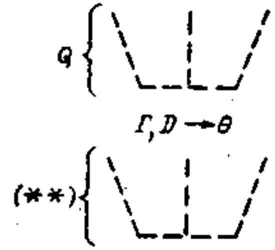


Рис. 6.14. Пояснение к выводу P_0 (продолжение)

7) Пусть $S(a)$ [или $F(a) \to \Delta(a)$] обозначает секвенцию вида $A_1(a), \ldots, A_m(a) \to B_1(a), \ldots, B_n(a)$. Тогда $S(t)$ [или $\Gamma(t) \to \Delta(t)$] обозначает секвенцию.

Аналогично предыдущему можно определить следующее: терм t полностью отмечен в $S(t)$ [или в $\Gamma(t) \to \Delta(t)$]. Для доказательства основного свойства выводимости, а именно того факта, что выводимость сохраняется при подстановке термов вместо свободных переменных, сначала приведем несколько определений и лемм.

Определение. Вывод в LK называется регулярным, если, во-первых, в нем все свободные переменные отличны друг от друга и, во-вторых, любая свободная переменная, входящая в некоторую секвенцию S этого вывода в качестве собственной, входит кроме нее только в те секвенции, которые расположены выше S.

Лемма. 1) Пусть $\Gamma(a) \to \Delta(a)$ — некоторая выводимая (в LK) секвенция, в которой переменная a полностью отмечена, и пусть $P(a)$ — вывод секвенции $\Gamma(a) \to \Delta(a)$. Пусть b — свободная переменная, не входящая в $P(a)$. Тогда дерево, полученное из $P(a)$ замещением каждого вхождения a в $P(a)$ на b, является выводом с заключительной секвенцией $\Gamma(b) \to \Delta(b)$.

2) Для произвольного LK-вывода существует регулярный вывод с той же заключительной секвенцией. Более того, требуемый вывод получается из первоначального простым переименованием (в подходящих местах) свободных переменных.

Доказательство. 1) При доказательстве используется индукция по числу непосредственных выводов в $P(a)$. Если $P(a)$ состоит только из начальной секвенции $A(a) \to A(a)$, то $P(b)$ состоит из секвенции $A(b) \to A(b)$, которая тоже является начальной.

Допустим, что наше утверждение верно для выводов, содержащих не более n непосредственных выводов, и предположим, что $P(a)$ содержит $n+1$ непосредственный вывод. Рассмотрим только случай, когда последний непосредственный вывод, скажем I, представляет собой применение правила $\forall$-справа (остальные случаи разбираются аналогично). Допустим, что собственной переменной этого применения является a и $P(a)$ имеет вид, представленный на рис. 6.15, где $Q(a)$ — подвывод вывода $P(a)$, оканчивающийся секвенцией $\Gamma \to \Lambda$, $A(a)$. Напомним, что a не входит ни в Γ, ни в Λ, ни в $A(x)$. По предположению индукции результат замещения всех a в $Q(a)$ на b есть вывод, заключительная секвенция которого суть $\Gamma \to \Lambda$, $A(x)$.

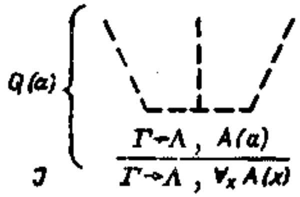


Рис. 6.15. Пояснение к лемме о регулярном выводе; правило $\forall$-справа и a — собственная переменная

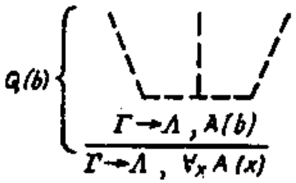


Рис. 6.16. Пояснение к правилу $\forall$-справа с использованием переменной b как собственной

Ни Γ, ни Λ не содержат b. Следовательно, можно в этой секвенции применить правило $\forall$-справа, используя переменную b в качестве собственной (рис. 6.16).

Это значит, что $P(b)$ есть вывод, оканчивающийся секвенцией $\Gamma\to\Lambda,\ \forall x A(x)$. Если a не является собственной переменной применения аппликации J, то $P(a)$ имеет вид, показанный на рис. 6.17.

По предположению индукции результат замещения всех a в $Q(a)$ на b является выводом с заключительной секвенцией $\Gamma(b)\to\Lambda(b),\ A(b, c)$.

2) Индукция по числу l применений правил $\forall$-справа и $\exists$-слева в данном выводе P представлена на рис. 6.18, где P_i — подвывод вывода P (рис. 6.19) и I_i — самое нижнее применение правил $\forall$-справа и $\exists$-слева в P $(i=1, \dots, k)$ [т. е. в той части вывода, которая обозначена через (✱), нет применений правил $\forall$-справа и $\exists$-слева].

Рассмотрим лишь случай, когда I_i является применением правила $\forall$-справа. Тогда вывод P_i содержит меньшее количество применений правила $\forall$-справа или $\exists$-слева, чем P, и поэтому по предположению индукции существует регулярный вывод P_i секвенции $\Gamma_i\to\Delta_i,\ F_i(b_i)$. Заметим, что ни одна свободная переменная из $\Gamma_i\to\Delta_i,\ F_i(b_i)$ (включая и b_i) не использовалась в P_i в качестве собственной переменной. Пусть $c_{i,1}, \dots, c_{i,m_i}$ — все свободные переменные вывода P_i', $i=1, \dots, k$.

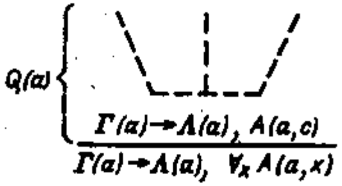


Рис. 6.17. Пояснение к правилу $\forall$-справа с использованием переменной a как собственной

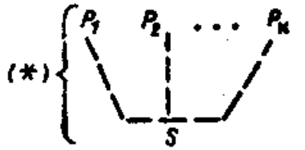


Рис. 6.18. Пояснение к индукции по числу l

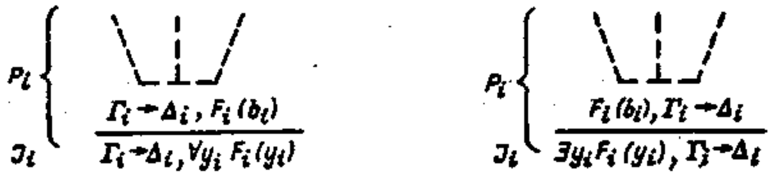


Рис. 6.19. Пояснение к подвыводу P_i вывода P

Пусть d_{ij} $(i=1, \dots, k;\ j=0,\ 1, \dots, m_i)$ — попарно различные свободные переменные, не входящие ни в один из выводов P, $P_1', \dots, P_k'$. Заменив в P_i' переменные b_i, $C_{i,1}, \dots, C_{i,m_i}$ на $d_{i,0}, d_{i,1}, \dots, d_{i,m_i}$ соответственно $(i=1, \dots, k)$. Получим, таким образом, регулярные выводы $P_1'', \dots, P_k''$. Наконец, вывод P' определим из рис. 6.20, где (✱) — такой же вывод, как и в P. Тем самым лемма доказана. В дальнейшем будем считать, что имеет место регулярный вывод. Аналогичным образом можно доказать следующую лемму.

Лемма. Пусть t — произвольный терм. Далее, пусть $\Gamma(a)\to\Delta(a)$ — некоторая выводимая (в LK) секвенция, в которой переменная a полностью отмечена, и пусть $P(a)$ — вывод, оканчивающийся секвенцией $\Gamma(a)\to\Delta(a)$, в котором всякая собственная переменная отлична от a и не содержится в t. Тогда $P(t)$ (результат замещения всех a в $P(a)$ термом t) является выводом с заключительной секвенцией $\Gamma(t)\to\Delta(t)$.

Лемма. Пусть t — произвольный терм, $\Gamma(a)\to\Delta(a)$ — выводимая (в LK) секвенция, в которой a полностью отмечена, и $P(a)$ — вывод секвенции $\Gamma(a)\to\Delta(a)$. Пусть $P'(a)$ — вывод, полученный из $P(a)$ некоторым переименованием собственных переменных (причем различные переменные заменяются не обязательно различными), таким, что в $P'(a)$ всякая собственная переменная отлична от a и не содержится в t. Тогда $P'(t)$ является выводом секвенции $\Gamma(t)\to\Delta(t)$.

Доказательство. При доказательстве используется индукция по числу собственных переменных в $P(a)$, которые совпадают с a или содержатся в t с использованием лемм.

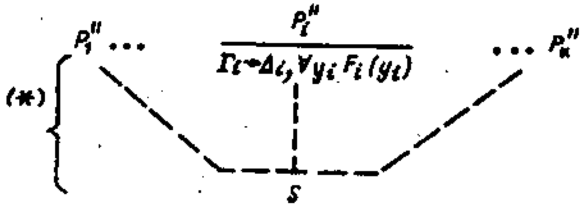


Рис. 6.20. Определение вывода P'

Перепишем часть утверждения первой леммы следующим образом.

Предложение. Пусть t — произвольный терм и $S(a)$ — выводимая (в LK) секвенция, в которой a полностью отмечена. Тогда секвенция $S(t)$ тоже выводима.

Предложение. Если некоторая секвенция выводима, то она имеет такой вывод, в котором все начальные секвенции состоят из атомарных формул. Кроме того, если некоторая секвенция выводима без сечений, то она имеет вывод указанного вида без сечений.

Доказательство. Достаточно показать, что для произвольной формулы A секвенция $A\to A$ выводима без сечений из начальных секвенций, состоящих из атомарных формул, а это легко доказать индукцией по сложности формулы.

Определение. Будем считать, что формулы A и B являются алфавитными вариантами друг друга, если для некоторых $x_1, \ldots, x_n$; $y_1, \ldots, y_n$ формула

$$\left(A\,\frac{x_1, \ldots, x_n}{z_1, \ldots, z_n}\right)$$

совпадает с формулой

$$\left(B\,\frac{y_1, \ldots, y_n}{z_1, \ldots, z_n}\right),$$

где $z_1, \ldots, z_n$ — различные связанные переменные, не входящие ни в A, ни в B, т. е. если A и B различаются только тем, что у них разный набор связанных переменных. В таком случае будем писать $A \sim B$.

Можно доказать, что отношение $A \sim B$ является отношением эквивалентности. Интуитивно ясно, что замена связанных переменных в формуле не изменяет ее смысла. Индукцией по числу логических символов в A можно доказать, что если $A \sim B$, то формула $A \equiv B$ выводима без сечений. Поэтому будем часто отождествлять два алфавитных варианта без специальных оговорок.

6.4. Синтаксическое введение в модальную логику. Модальное исчисление высказываний

Оставшаяся часть этой главы будет посвящена рассмотрению неклассических логик двух типов — модальных и интенсиональных, которые, как в дальнейшем убедимся, тесно связаны между

собой. Идея применения аппарата интенсиональных логик в задачах представления знаний в системах искусственного интеллекта начинает завоевывать все более прочные позиции наряду с такими признанными концепциями, как теория фреймов в различных ее вариантах, семантические сети [30] и т. д.

Наша цель состоит именно в том, чтобы показать все удобства применения моделей, основанных на модальной логике, как дающих возможность более адекватного отображения предметной области (ПО). Это приводит к необходимости серьезного подхода к созданию соответствующих систем представления знаний (СПЗ), основанных на таком математическом аппарате.

В этом параграфе рассмотрены общий подход к формальным системам модальной логики и несколько таких *синтаксических* систем [74]. Однако для инженера знаний более важным является вопрос о возможных интерпретациях или *моделях* таких систем, т. е. *семантика*. В настоящее время существует большое число различных точек зрения на интерпретацию модальных систем [74].

В большинстве своем семантические системы связаны с такими именами, как С. Крипке, Р. Монтегю, Я. Хинтикка и Д. Скотт (см. рис. 6.2). Рассмотрим три подхода к семантике, предложенных первыми тремя авторами. *Структуры* Крипке, рассматриваемые в § 6.5, позволяют проследить связь синтаксических систем с интерпретациями. Далее рассмотрены *прагматический язык* модальной логики Монтегю и его интерпретация, а также расширения, приводящие к *интенсиональному языку*, подмножеством формул которого по существу является множество формул модальной логики.

Весьма важно при семантическом рассмотрении модальных систем обсудить понятие возможных миров или точек соотнесения.

Один из подходов их моделирования — *модельные множества*, предложенные Хинтиккой, которые рассмотрены в § 6.8. В заключение дадим обзор взглядов Д. Скотта на «точки соотнесения» и интенсиональные системы и тем самым подведем черту всему обсуждению.

6.4.1. Исчисление предикатов как отправной пункт для построения модального исчисления высказываний

Модальная логика возникла в результате расширения исчисления предикатов первого порядка. В дальнейшем, чтобы избежать путаницы, будем подразумевать исчисление со следующими постулатами.

1. Аксиомы классического исчисления высказываний (Рассела—Бернайса):

1.1. $\vdash (p \vee p) \rightarrow p.$

1.2. $\vdash p \rightarrow (p \vee q).$

1.3. $\vdash (p \vee q) \rightarrow (q \vee p).$

1.4. $\vdash (p \rightarrow q) \rightarrow [(r \vee p) \rightarrow (r \vee q)].$

2. Правила вывода:

2.1. Правило отделения

$$\frac{\vdash P \vdash P \rightarrow Q}{\vdash Q}.$$

2.2. Правило обобщения: если X несвободно в P, то

$$\frac{\vdash}{\vdash \forall XP}.$$

3. Аксиомы:

3.1. $\vdash \forall x(P \rightarrow Q) \rightarrow (\forall xP \rightarrow \forall xQ).$

3.2. Пусть X/Y есть результат подстановки Y вместо каждого свободного вхождения переменной X в P. Тогда, если Y не окажется связанной в P на тех местах, где переменная X была свободна,

$$\vdash \forall xP \rightarrow (x/Y)\,P.$$

3.3. $\vdash \forall x \forall yP \rightarrow \forall Y \forall xP.$

4. Правила подстановки в нашей системе со схемами аксиом не требуются.

5. Определения:

5.1 $\exists xP \rightleftarrows \neg \forall x \neg P,$

5.2. $P \vec{x} Q \rightleftarrows \forall x\,(P \rightarrow Q).$

5.3. $P \overset{\leftrightarrow}{x} Q \rightleftarrows \forall x\,(P \leftrightarrow Q).$

Приведем теперь (без вывода) основные теоремы и выводимые правила предложенного исчисления.

1. Отрицание:

1.1. $\vdash \neg \forall xP \leftrightarrow \exists x \neg P.$

1.2. $\vdash \forall xP \leftrightarrow \neg \exists x \neg P.$

1.3. $\vdash \neg \exists xP \leftrightarrow \forall x \neg P.$

1.4. $\vdash \exists xP \leftrightarrow \neg \forall x \neg P.$

2. Подчинение (Y не связана в P на тех местах, где x свободна)

2.1. $\vdash \forall xP \rightarrow (x/Y)P.$

2.2. $\vdash (x/Y)P \rightarrow \exists xP.$

2.3. $\vdash \forall xP \rightarrow \exists xP.$

3. Дистрибутивность кванторов по отношению к & и $\vee$

3.1. $\vdash \forall x(P \mathbin{\&} Q) \leftrightarrow (\forall xP \mathbin{\&} \forall xQ).$

3.2. $\vdash (\forall xP \vee \forall xQ) \rightarrow \forall x(P \vee Q).$

3.3. $\vdash \exists x(P \vee Q) \leftrightarrow (\exists xP \vee \exists xQ).$

3.4. $\vdash \exists x(P \mathbin{\&} Q) \rightarrow (\exists xP \mathbin{\&} \exists xQ).$

3.5. $\vdash \exists x(P \mathbin{\&} Q) \rightarrow \exists xP.$

4. Дистрибутивность кванторов по отношению $\rightarrow$ и $\leftrightarrow$

4.1. $\vdash \forall x(P \rightarrow Q) \rightarrow (\forall xP \rightarrow \forall xQ).$

4.2. $\vdash [\forall x(P \rightarrow Q) \mathbin{\&} \forall xP] \rightarrow \forall xQ.$

4.3. $\vdash \forall x(P \rightarrow Q) \rightarrow (\exists xP \rightarrow \exists xQ).$

4.4. $\vdash [\forall x(P \rightarrow Q) \mathbin{\&} \exists xP] \rightarrow \exists xQ.$

4.5. $\vdash \forall x(P \leftrightarrow Q) \rightarrow (\forall xP \leftrightarrow \forall xQ).$

4.6. $\vdash \forall x(P \leftrightarrow Q) \rightarrow (\exists xP \leftrightarrow \exists xQ).$

5. Правила дедукции

5.1. $\vdash \forall x(P \rightarrow Q)/\vdash \forall xP \rightarrow \forall xQ.$

5.2. $\vdash \forall x(P \rightarrow Q)/\vdash \exists xP \rightarrow \exists xQ.$

5.3. $\vdash \forall x(P \leftrightarrow Q)/\vdash \forall xP \leftrightarrow \forall xQ.$

5.4. $\vdash \forall x(P \leftrightarrow Q)/\vdash \exists xP \leftrightarrow \exists xQ.$

Связки $\rightarrow$ и $\leftrightarrow$ будем в дальнейшем называть материальными связками (импликации и эквивалентности соответственно), а $\vec{x}$ и $\overset{\leftrightarrow}{x}$ — формальными связками.

В дальнейшем, при построении модального исчисления, возьмем те же самые переменные $p, q, r \ldots$, те же самые связки $\neg$, &, $\vee$, $\rightarrow$, $\leftrightarrow$, те же аксиомы и правила. Кроме того, в язык добавятся символы $\Box$ (необходимость) и $\Diamond$ (возможность), постулаты для которых в дальнейшем будут сформулированы.

6.4.2. Основная идея перехода к модальному исчислению высказываний

В теории предикатов первого порядка всякое предложение (формула свободных переменных) — это утверждение о некотором определенном факте. Модальная логика в отличие от него будет рассматривать утверждения при некоторых «обстоятельствах».

Очевидно, что для обозначения «обстоятельства» или «случая» надо ввести соответствующую переменную t, следуя [72, 77]. Это переменная особого рода, отличная от индивидных переменных. Фразу «Событие p происходит в случае «запишем так: p, t. Утверждение «Событие p необходимо» запишем как $\forall tpt$, «Событие p возможно — как $\exists tpt$.

Перед тем как переходить собственно к модальному исчислению высказываний, рассмотрим некоторый промежуточный язык, в котором в каждый предикат введена дополнительная переменная t, а модальности выражены посредством кванторов. Приведем более строгое построение.

6.4.3. Правила образования промежуточного языка

Задан бесконечный список пропозициональных переменных $p, q, r, s \ldots$, которые, следуя [77], будем обозначать $P, Q, R, S \ldots$, а переменную t — через T.

Предложения промежуточного языка:

1. Выражения вида PT — предложения. Однако ни P, ни T в отдельности предложениями не являются.

2. Если M и N — предложения, то $\neg M$, $M \& N$, $M \vee N$, $M \to N$, $M \leftrightarrow N$ — тоже предложения.

3. Если M — предложение, то $\forall\ TM$, $\exists TM$ — тоже предложения.

Таким образом, наш язык не допускает предметных переменных: на базе исчисления *предикатов* первого порядка строим модальное исчисление *высказываний*.

Постулаты вспомогательного исчисления формулируются «параллельно» постулатам исчисления предикатов первого порядка. Однако при этом необходимо учесть следующее. Аналога аксиомы 3.3 не существует. Правило обобщения, параллельное правилу 2.2, существенно. Можно доказать, что $Q'TQTPT$ строго эквивалентно сводится к $QTPT$, где Q — один из кванторов $\forall$ или $\exists$.

Очевидно, что если некоторое выражение исчисления предикатов первого порядка доказуемо, то соответствующее ему выражение промежуточного языка тоже доказуемо.

Теперь перейдем к модальному исчислению высказываний посредством следующей замены:

$$pt,\quad qt,\quad rt,\ .\ .\ .\ ,\ \forall t,\quad \exists t, \vec{t},\quad \overleftrightarrow{t}$$
$$\text{на}\quad p,\ q,\ t,\ .\ .\ .\ ,\ \Box,\ \Diamond,\ \Rightarrow,\ \Leftrightarrow.$$

Эта замена ввиду одно-однозначного соответствия сохраняет доказуемость.

6.4.4. Постулаты и основные теоремы модального исчисления высказываний

Постулаты

1. Все аксиомы (1.1—1.4).

2. Правило подстановки: всякая доказуемая формула остается доказуемой, если вместо пропозициональной переменной подставить некоторую формулу.

3. Правило отделения 2.1.

4. Определения символов &, $\to$, $\leftrightarrow$.

5. Правило $\dfrac{\vdash p}{\vdash \Box p}$.

6. Аксиомы для модальности.

6.1. $\vdash, \Box(p \to q) \to (\Box p \to \Box q)$.

6.2. Если P — собственная модельность, т. е. предложение, содержащее непустой набор $\Box$ и $\Diamond$, то $\vdash p \to \Box p$.

6.3. $\vdash \Box p \to \Box p$.

6.4. Если P — формула исчисления предикатов первого порядка, то P — формула в модальном исчислении высказываний. Если P, Q — формулы, то $\neg P$, $P \vee Q$, $\Box P$ — тоже формулы.

7. Определения:

7.1. $\Diamond P \rightleftarrows \neg \Box \neg P$.
7.2. $P \Rightarrow Q \rightleftarrows \Box (P \to Q)$.
7.3. $P \Leftrightarrow Q \rightleftarrows \Box (P \leftrightarrow)$.

Основные теоремы модального исчисления высказываний

Приведем выводимые формулы и правила, соответствующие теоремам и правилам исчисления предикатов первого порядка. Доказательство их легко может быть проведено.

1. Отрицание:

1.1. $\vdash \neg \Box p \leftrightarrow \Diamond \neg p$.
1.2. $\vdash \Box p \leftrightarrow \neg \Diamond \neg p$.
1.3. $\vdash \neg \Diamond p \leftrightarrow \Box \neg p$,
1.4. $\vdash \Diamond p \leftrightarrow \neg \Box \neg p$.

2. Субординация:

2.1. $\vdash \Box p \to p$.

2.2. $\vdash p \to \Diamond p$.

2.3. $\vdash \Box p \to \Diamond p$.

3. Дистрибутивность модальностей относительно & и $\vee$:

3.1. $\vdash \Box (p \,\&\, q) \leftrightarrow (\Box p \,\&\, \Box q)$.

3.2. $\vdash (\Box p \vee \Box q) \to \Box (p \vee q)$.

3.3. $\vdash \Diamond (p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q)$.

3.4. $\vdash \Diamond (p \,\&\, p) \to (\Diamond p \,\&\, \Diamond q)$.

3.5. $\vdash \Diamond (p \,\&\, q) \to \Diamond p$.

4. Дистрибутивность модальностей относительно условных связок:

4.1. $\vdash (p \Rightarrow q) \to (\Box p \to \Box q)$.

4.2. $\vdash [(p \Rightarrow q) \,\&\, \Box p] \to \Box q)$.

4.3. $\vdash (p \Rightarrow q) \to (\Diamond p \to \Diamond q)$.

4.4. $\vdash [(p \Rightarrow q) \,\&\, \Diamond p] \to \Diamond q$.

4.5. $\vdash (p \Leftrightarrow q) \to (\Diamond p \leftrightarrow \Diamond q)$.

5. Правила дедукции:

5.1. $\vdash P \Rightarrow Q / \vdash \Box P \to \Box Q$.

5.2. $\vdash P \Rightarrow Q / \Diamond P \to \Diamond Q$.

5.3. $\vdash P \Leftrightarrow Q / \vdash \Box P \leftrightarrow \Box Q$.

5.4. $\vdash P \Leftrightarrow Q / \vdash \Diamond P \leftrightarrow \Diamond Q$.

В дальнейшем символы $\Rightarrow$ и $\Leftrightarrow$ будем называть, следуя [72], *строгими* условными связками, а символы $\to$ и $\leftrightarrow$ — по-прежнему *материальными* связками.

Рассмотренное нами модальное исчисление высказываний удобно в эвристическом смысле благодаря тому, что оно соответствует исчислению предикатов первого порядка. Однако «ввиду своей силы» оно является в некотором смысле «тривиальной системой» [72]. Интересно было бы рассматривать более слабые системы с менее узкими условиями, накладываемыми на модальности и строгие связки.

Очевидный путь — в замене ряда (независимых) аксиом на более слабые. В результате получим такие важные в модальной логике системы, как *S1*, *S4* и *S5*, которые понадобятся нам для дальнейшего, преимущественно семантического изложения.

6.4.5. Система *S1*

1. Аксиомы:

1.1. $\vdash p \& q \Rightarrow p$.

1.2. $\vdash p \& q \Rightarrow q \& p$.

1.3. $\vdash [(p \& q) \& r] \Rightarrow [p \& (q \& r)]$.

1.4. $\vdash p \Rightarrow p \& p$.

1.5. $\vdash [(p \Rightarrow q) \& (q \Rightarrow r)] \Rightarrow (p \Rightarrow r)$.

1.6 $\vdash p \Rightarrow \Diamond p$.

Заметим, что последняя аксиома независима от предыдущих, в чем легко можно убедиться при помощи таблиц истинности.

2. Правило образования:
если P и Q — формулы, то $\neg P, \Diamond P, P \& Q$ — тоже формулы.

3. Правило подстановки:
доказуемое предложение остается доказуемым, если в него вместо пропозициональных переменных подставить некоторые формулы.

4. Правило соединения:
если $\vdash P$ и если $\vdash Q$, то $\vdash P \& Q$.

5. Правило отделения для $\Rightarrow$:
если $\vdash P$ и если $\vdash P \Rightarrow Q$, то $\vdash Q$.

6. Правило замены строго эквивалентных:
если $\vdash P \Rightarrow Q$, то доказуемая формула остается доказуемой, если в ней некоторые вхождения P заменить на Q.

7. Определения:
7.1. $P \vee Q \rightleftarrows \neg(\neg P \& \neg Q)$.
7.2. $P \to Q \rightleftarrows \neg (P \& \neg Q)$.
7.3. $P \leftrightarrow Q \rightleftarrows (P \to Q) \& (Q \to P)$.
7.4. $P \Rightarrow Q \rightleftarrows \neg \Diamond (P \& \neg Q)$.
7.5. $P \Leftrightarrow Q \rightleftarrows (P \Rightarrow Q) \& (Q \Rightarrow P)$.
7.6. $\Box P \rightleftarrows \neg \Diamond \neg P$.

Можно показать, что система аксиом *S1* непротиворечива.

Условимся о дальнейшем употреблении термина «модальный» [72, 75]:

1. Пропозициональные переменные являются модальными выражениями (*модальными предложениями*).

2. Если P и Q — модальные предложения, то $\neg P$, $P \& Q$, $\Diamond Q$ — модальные предложения.

3. Каждое выражение, строго эквивалентное по определению модальному предложению, само является модальным предложением.

Модальность — это последовательность символов $\neg$, $\Diamond$ или любое выражение, которое может заменить такую последовательность.

Степень модальности — число символов $\Diamond$, содержащихся в модальности, состоящей только из $\neg$ и $\Diamond$ и эквивалентной по определению рассматриваемой модальности.

Собственная модальность — модальность степени выше чем нуль.

Утвердительная модальность — модальность, в которой число символов $\neg$ равно нулю или четно. Отрицательная модальность — модальность, в которой число символов $\neg$ нечетно.

Назовем $\Diamond$ знаком *возможности*, а $\Box$ — знаком *необходимости.*

6.4.6. Система *S4*

Весьма важной является также система *S4*. Характерна она тем, что почти все теоремы, доказуемые в следующей системе *S5*, доказуемы также и в ней.

Постулаты S4

Аксиомы *S4* — это аксиомы *S1* с добавлением

$$\vdash \Diamond\Diamond P \Rightarrow \Diamond P$$

или двойственной ей аксиомы: $\vdash \Box P \Rightarrow \Box\Box P$.

Введение этой аксиомы ведет к тому, что становятся доказуемыми формулы

$$\vdash \Diamond\Diamond P \Leftrightarrow \Diamond P \quad \text{и} \quad \vdash \Box\Box P \Leftrightarrow \Box P.$$

Таким образом, «возможно» в этой системе отождествляется с «возможно, что возможно».

Заметим, что введенная аксиома независима от постулатов *S*1 и постулаты *S4* непротиворечивы.

6.4.7. Система *S5*

Постулаты *S*5 определяются как постулаты *S1* с добавленной к ним аксиомой

$$\vdash \Diamond P \Rightarrow \Box\Diamond P.$$

Данная аксиома независима от постулатов *S1* и не противоречит им.

В *S*5 выводим эквивалентности

$$\vdash \Box\Diamond P \Leftrightarrow \Diamond P \quad \text{и} \quad \vdash \Diamond\Box P \Leftrightarrow \Box P.$$

В *S5* можно вывести формулу $\vdash \Diamond\Diamond P \Rightarrow \Diamond P$, поэтому система содержит систему *S4*.

Итак, мы рассмотрели различные виды модального исчисления высказываний первого порядка, не касаясь модального исчисления предикатов первого порядка, которое может быть получено из первого путем введения соответствующих постулатов

для формул, содержащих квантификацию по предметным переменным.

Далее наше изложение будет носить, главным образом, семантический характер; в частности, будут рассмотрены модели приведенных систем.

6.5. ВЗАИМОСВЯЗЬ СИНТАКСИЧЕСКИХ СИСТЕМ С СЕМАНТИЧЕСКИМИ. СТРУКТУРЫ КРИПКЕ

В этом параграфе рассмотрим один из способов интерпретации модальных систем при помощи структур, предложенный С. Крипке [77]. Достоинство этого метода состоит в том, что четко прослеживается связь между системами аксиом и интерпретирующими отношениями, т. е. между синтаксисом и семантикой.

Сделаем несколько замечаний об используемом языке. Модальное исчисление высказываний по-прежнему задается бесконечным списком пропозициональных переменных, комбинируя которые при помощи связок &, $\neg$ и $\Box$, можно получать правильно построенные формулы, которые при необходимости будут обозначаться метапеременными *A*, *B*, *C*... Модальное исчисление высказываний будет называться нормальным, если оно содержит следующие аксиомы и правила вывода:

*A*1. $\Box A \to A$.

*A*3. $\Box(A \to B) \supset (\Box A \to \Box B)$.

*R*1. Если $\vdash A$ и $\vdash A \to B$, то $\vdash B$.

*R*2. Если $\vdash A$, то $\vdash \Box A$ (в ненормальных системах правило R_2 не выполняется).

Система, содержащая только приведенные аксиомы и правила, называется системой *M* (системой *T*, системой Фейса фон Вригта).

Система,, эквивалентная *S*4, получается из *M* добавлением схемы аксиом *A*4. $\Box A \to \Box\Box A$.

Брауэрова система получается из *M* добавлением брауэровой схемы аксиом $A \to \Box\Diamond A$.

Система, эквивалентная *S*5, может быть определена как система *M* с добавленной схемой аксиом *A*2. $\neg\Box A \to \Box\neg\Box A$.

Отметим, что система *S*4 с добавленной брауэровской схемой эквивалентна *S*5. В дальнейшем под *S*4, *S*5 будем понимать рассмотренные здесь системы.

6.5.1. Модельные структуры

Нормальной модельной структурой [72] называется упорядоченная тройка *G*, *K*, *R*, где *K* — некоторое непустое множество, элементы которого могут рассматриваться как «точки соотнесения» или «обстоятельства», в которых высказывание истинно или ложно; $G \in K$ — некоторый выделенный элемент этого множества, иначе «действительный мир», тогда как другие элементы — «возможные миры»; R — отношение, определенное на множестве K, на которое наложено единственное требование рефлексивность.

Моделью формулы A системы M, $S4$, $S5$ называется двухместная функция $\Phi(P, H)$, соответствующая данной M-, $S4$-, $S5$-модельной структуре. Первая переменная пробегает множество атомарных подформул формулы A, а переменная H пробегает элементы множества K. Функция принимает значения на множестве $\{T, F\}$.

Для данной модели Φ, соответствующей модельной структуре (G, K, R), определим некоторое однозначное расширени функции Φ так, чтобы получить оценку [т. е. значение $\Phi(B, H)$] для каждой (быть может, неатомарной) подформулы B формулы A.

Значение $\Phi(B, H)$ определяется индукцией по числу логических связок (i).

Если B — атомарная формула, то оценка B уже определена в модели ii.

Пусть $\Phi(B, H)$ и $\Phi(C, H)$ определены. Тогда, если $\Phi(B, H) = \Phi(C, H) = T$, то $\Phi(B \& C, H) = T$, в противном случае $\Phi(B \& C, H) = F$. Если $\Phi(B, H) = T$, то $\Phi(\neg B, H) = F$, в противном случае $\Phi(\neg B, H) = T$. (iii). Пусть $\Phi(B, H') = T$ для всех H' из K, таких, что HRH', тогда полагаем $\Phi(\Box B, H) = T$ В противном случае $\Phi(\Box B, H) = F$.

Формула A называется истинной в модели Φ, связанно с модельной структурой (G, K, R), если $\Phi(A, G) = T$, и ложно в противном случае. Формула A называется общезначимо если она истинна во всех своих моделях, и выполнимой, ес она истинна хотя бы в одной из моделей.

6.5.2. Свойства отношения R

Теперь установим связь между свойствами модельных и си таксических систем.

Отношение R интуитивно может пониматься следующим образом: если для двух данных миров H_1, $H_2 \in K H_1 R H_2$, то H «возможен» относительно H_1 (зависит от H_1). Отсюда следуе требование рефлексивности отношения R. Каждый мир возможен относительно самого себя; всякое высказывание, *истинное* в H, является также *возможным* в H. Формула A оценивается как необходимая в мире H, если она истинна в каждом мире, возможном относительно H. Наоборот, A возможна в некотором мире, если существует мир, возможный относительно данного, в котором она истинна.

Выясним, при каком условии отношение R будет обладать свойством транзитивности. Требуется, чтобы из $H_1 R H_2$ и $H_2 R H_3$ следовало $H_1 R H_3$. Если $H_2 R H_3$, то каждая формула A, истинная в H_3, возможна в H_2. Но $H_1 R H_2$, значит, формула $\Diamond A$, истинная в H_2, возможна в H_1, т. е. в H_1 истинна формула $\Diamond \Diamond A$. Но $H_1 R H_3$, значит, если A истинна в H_3, то A_3 возможна в H_1, т. е. $\Diamond A$ истинна в H_1. Это приводит к требованию дополнительной аксиомы: $\Diamond \Diamond A \rightarrow \Diamond A$ «возможно, что, возможно, A означает, что A возможно», а это — аксиома, редукции $S4$. В соответствии с этим будем называть модельную структуру с транзитивным отношением R $S4$-модельной структурой.

Аналогично можно показать, что брауэрова аксиома приводит к симметричности отношения R. Действительно, пусть $A \rightarrow \Box \Diamond A$ и имеет место $H_1 R H_2$. Установим, что $H_2 R H_1$. Это будет означать, что всякое утверждение A, истинное в H_1, возможно в H_2. Но если A истинно в H_1, то и $\Box \Diamond A$ истинно в H_1, т. е. $\Diamond A$ необходимо в H_1 и $\Diamond A$ истинно в H_2, что и требовалось доказать. Будем называть в связи с этим модельную структуру с симметричным отношением R брауэровой модельной структурой.

Отказаться от отношения R — это то же самое, что принять его существование для каждой пары элементов множества K. Это значит, что каждое возможное высказывание является необходимо возможным, т. е. это приводит к характеристической аксиоме $S5$. К той же аксиоме придем, считая R отношением эквивалентности. Поэтому будем называть модельную структуру с таким отношением $S5$-модельной структурой. Модельная структура с отношением R, на которое не наложено особых требований, называется просто M-модельной структурой.

Определение. Модельная структура (G, K, R) называется *связной*, если для всех $H \in K$ имеет место GR^*H, где R^* — транзитивное замыкание отношения R.

Определение. Модель называется связной, если она определена на связной модельной структуре.

Определение. Модель Φ является моделью формулы A тогда и только тогда, когда A истинна в Φ, в противном случае Φ является контрмоделью для A.

Предложение. Каждая выполнимая формула имеет связную модель.

Пусть A выполняется в модели $\Phi(P, H)$, определенной на модельной структуре (G, K, R);

$$K' \rightleftharpoons \{H / H \leftarrow K, GR^*H\};$$

R' — ограничение отношения R^* на множестве K', а модель $\Phi'(P, H)$ есть модель $\Phi(P, H)$, ограниченная условием $H \in K'$. Мы получили модельную структуру (G, K', R') и в ней связную модель Φ'. Остается показать, что $\Phi'(B, H) = \Phi(B, H)$ для всех $H \in R'$ и любой подформулы B формулы A, тогда Φ' и будет искомой моделью. Случай, когда B атомарна, тривиален. Пусть $B = C \& D$ или $B \mathrel{\hat{=}} > C$. В этом случае проверка утверждения также очевидна (в соответствии с определением модели для B), при условии, что утверждение уже доказано для C и D.

Остается рассмотреть случай $B \mathrel{\hat{=}} \Box C$, а утверждение для C доказано. Если $H \in K$, то из $HR'H'$ следует $H' \in K'$, следовательно, HRH' (т. е. для $H \in K'$ $HRH' \equiv HR'H'$). Но $\Phi(C, H') = \Phi'(C, H')$. Тогда:

1) $\Phi(\Box C, H) = T$ тогда и только тогда, когда для всякого $H' \in K$, $HRH' \Phi(C, H') = T$;

2) $\Phi'(\Box C, H) = T$ тогда и только тогда, когда для всякого $H' \in K'$, $HR'H' \Phi'(C, H') = T$.

Отсюда, если положить $H \in K'$, то результаты 1 и 2 эквивалентны, т. е. $\Phi(\Box C, H) = T$ тогда и только тогда, когда $\Phi'(\Box C, H) = T$, т. е. в этом случае $\Phi(\Box C, H) = \Phi'(\Box C, H)$, что и требовалось.

Доказанное предложение приводит к тому, что можно было бы ограничиться рассмотрением только связных моделей. Очевидно, что если в связной модели R есть отношение эквивалентности, то любые два мира связаны с этим отношением; этим объясняется то, что модельная структура с таким отношением есть $S5$-модельная структура.

Дадим еще несколько определений, относящихся к связным моделям.

Определение. Тройка (G, K, S), где K — некоторое множество, $G \in K$, а S — отношение на K, называется деревом, а G — его корнем, если:

1) не существует $H \in K$, HSG;

2) для каждого $H \in K$, кроме G, существует единственное $H' : H'SH$;

3) для каждого $H \in K$, GS^*H.

Если HSH', то H называется предшественником.

Определение. M-модельная структура (G, K, R) называется *древовидной*, если существует такое отношение S', что (G, K, S) является деревом, а R — наименьшее рефлексивное отношение, содержащее S (рефлексивное отношение, «порожденное» S'). Аналогично определяется брауэрова, $S4$-, $S5$-модельная структура с условием, что R — наименьшее рефлексивное и симметричное, рефлексивное и транзитивное отношение и отношение эквивалентности соответственно.

Очевидно, что *каждая древовидная модельная структура связна.*

В $S5$ каждая конечная или счетная связная модельная структура является древовидной; это, однако, может не выполняться для $S4$.

Определение. Модель, ассоциированная с древовидной модельной структурой, называется *древовидной моделью.*

Можно показать, что семантическая теория не потеряет общности, если рассматривать только древовидные модели. Кроме того, модельные структуры допускают наглядное диаграммное представление.

6.5.3. Семантические таблицы

Установим теперь критерий общезначимости формулы A в рассматриваемых системах. Для этого нам придется ввести дополнительное понятие *семантической таблицы.* Для установления общезначимости формулы A будем пытаться найти контрмодель для этой формулы, если такой контрмодели не существует, то A общезначима.

Пусть формула A сведена к виду

$$A_1 \& \ . \ . \ . \ \& A_n \rightarrow B_1 \vee \ . \ . \ . \ \vee B_n.$$

Очевидно, что во всякой контрмодели A формулы $A_1, \ldots, A_m$ должны быть ложными, а $B_1, \ldots, B_n$ — истинными. Чтобы попытаться найти такую модель, поместим $A_1, \ldots, A_m$ в левый столбец, а $B_1, \ldots, B_m$ — в правый столбец главной таблицы. (На каждом шаге будем иметь дело с системой альтернативных таблиц, упорядоченных отношением R, одна из которых выбрана в качестве главной, а другие — вспомогательные.)

К любой таблице — главной или вспомогательной — на каждом шаге будут применяться следующие правила:

Nl. Если $\rceil A$ в левом столбце, то помещаем A в правый столбец.

Nr. Если $\rceil A$ в правом столбце, то помещаем A в левый столбец.

Λl. Если $A \& B$ в левом столбце, то помещаем A и B в левый столбец.

Λr. Если $A \& B$ в правом столбце некоторой таблицы t то имеются две возможности: расширить таблицу t, поместив в правый столбец либо A, либо B.

В этом случае на следующем шаге у нас будут уже две альтернативные таблицы t_1 и t_2. В последнем случае следует отметить, что, так как таблица t содержалась в некотором множестве φ, упорядоченном рефлексивным отношением R, то на следующем шаге получим два множества φ_1 и φ_2, упорядоченных рефлексивными отношениями R_1 и R_2, которые совпадают с R, за исключением того, что t заменена на t_1 (t_2).

Заметим, что перечисленные правила напоминают структурные правила из п. 6.5.1.

Yl. Если $\Box A$ оказывается в левом столбце таблицы t, то для каждой таблицы t', такой, что tRt', помещаем A в левом столбце таблицы t'.

Yr. Если $\Box A$ стоит в правом столбце таблицы t, то составляем новую таблицу t' с A в правом столбце, такую, что tRt'.

На отношение R при построении семантических таблиц накладываются те же самые требования, что и на отношение R в случае брауэровых $S4$-, $S5$-модельных структур (G, K, R).

Определение. Таблица называется замкнутой, если некоторая формула A оказывается в обоих столбцах таблицы. Множество φ таблиц замкнуто, если некоторая таблица в нем замкнута. Система альтернативных множеств замкнута, если каждое множество в ней замкнуто.

Определение. *Конструкцией* (построением) для формулы A будем называть построение, начинающееся с помещения A в правый столбец главной таблицы данной конструкции.

Определение. Конструкция называется *замкнутой*, если на какой-то стадии в ней появляется замкнутая система альтернативных множеств.

Наложим также некоторые требования на применение правил N, Λ, $Y(l, r)$. Правила нельзя применять к формуле, находящейся в замкнутом альтернативном множестве. Правило нельзя применять, если оно «излишне» т. е. если множество уже содержит «результат» применения этого правила.

Приведем пример построения конструкции семантических таблиц для формулы A в $S4$-структуре.

Пусть имеется формула $\Box(A\&B)\rightarrow\Box(\Box A\&\Box B)$. Для нее строится таблица:

$$\begin{array}{r|c|c|c}
 & \Box(A\&B) & \Box(\Box A\&\Box B) & \Box A\&\Box B \\
Yl & A\&B & & Yr \\
\Lambda l & A & \rightarrow & \\
\Lambda l & B & {}_{t_1} & {}_{t_2}
\end{array}$$

(Рядом с формулами написаны правила, по которым они получены.) В результате применения Yr к правой части получена таблица t_2. Стрелка означает, что t_1Rt_2.

Правило Λr теперь доказывает две возможности:

$$\begin{array}{c|c|c}
\Box(A\&B) & \Box(\Box A\&\Box B) & \Box A\&\Box B \\
A\&B & \rightarrow & \Box A \\
A & & \\
B & {}_{t_1} & {}_{t_{21}}
\end{array}$$

или

$$\begin{array}{c|c|c}
\Box(A\&B) & \Box\Box A\&\Box B & \Box A\&\Box B \\
A\&B & \rightarrow & \Box B \\
A & & \\
B & {}_{t_1} & {}_{t_{22}}
\end{array}$$

Продолжим рассмотрение первого варианта. Приложив правило Yr, получим t_3 с A в правом столбце. Так как t_1Rt_{21} и $\Box(A\&B)$ слева в t_1, то поместим $A\&B$ слева в t_{21}, а ввиду транзитивности R, также поместим $A\&B$ слева в t_3. Далее, по Λl в левые части этих таблиц поместим A и B:

$$\begin{array}{c|cc|cc|c}
\Box(A\&B) & \Box(\Box A\&\Box B) & A\&P_0 & \Box A\&\Box B & A\&B & A \\
A\&B & & A & \Box A & A & \\
A & & B & & B & \\
B & {}_{t_1} & & {}_{t_{21}} & & {}_{t_3}
\end{array}$$

Построение замкнулось, так как A оказалось в обеих частях t_3.

Предложение. Не может быть $S4$-модели, в которой $\Box(A\&B)$ было бы истинно, а $\Box(\Box A\&\Box B)$ — ложно.

В такой модели каждой t_i $(i=1, 2, 3)$ будет соответствовать мир H_i $(G=H_1)$, такой, что для каждого C имеем $\Phi(C, H_i)=T(F)$, если C слева (справа) в таблице t_i.

Мы выбрали одну из альтернатив так, что A оказалось справа и слева, и получили противоречия — одновременно

$$\Phi(A,\ H_3)=T \quad \text{и} \quad \Phi(A,\ H_3)=F.$$

Заметим, что если отношение R нетранзитивно, то мы не смогли бы получить замыкание. Построение не изменилось бы, если бы стрелку рассматривать как симметричное отношение. Поэтому формула $\Box(A\&B)\rightarrow\Box(\Box A\&\Box B)$ общезначима в $S4$, но не в M и не в брауэровой системе.

6.5.4. Результаты общезначимости

Приведем (без доказательства) результаты, показывающие связь свойств семантических таблиц со свойствами моделей.

Предложение. Построение для формулы A замыкается тогда и только тогда, когда A общезначима для каждой из четырех рассматриваемых систем (M, брауэрова, $S4$ и $S5$) или для любых других систем, в которых на отношения R для таблиц и моделей накладываются одни и те же ограничения.

Эта теорема сводится к следующим двум леммам.

Лемма. Если построение для формулы A замыкается, то формула A общезначима.

Наметим схему доказательства этой леммы. Предполагается от противного, что A — необщезначимая формула. Далее индукцией по n доказывается, что для каждого n на n-м шаге конструкции найдется альтернативное множество φ этой конструкции и отображение α, преобразующее таблицы из множества φ в элементы K со следующим свойством: если t является таблицей из φ, $H=\alpha(t)$, а B есть любая формула, встречающаяся в левом (правом) столбце t, то $\Phi(B, H)=T$. Далее если t_1 и t_2 принадлежат φ $H_1=\alpha(t_1)$, а $H_2=\alpha(t_2)$, то t_1Rt_2 влечет H_1kH_2.

Это утверждение очевидно при $n=1$. Предположим, что результат получен для n-го шага: имеется альтернативное множество φ и отображение α с требуемыми свойствами. Переходим к $(n+1)$-му шагу. Этот шаг получается из предыдущего по одному из ранее приведенных правил. Необходимо последовательно просмотреть все правила. Так, для правила Λl $B\&C$ оказывается в t слева и согласно гипотезе для $H=\alpha(t)$ имеем $\Phi(B\&C, H)=T$. Следовательно, $\Phi(B, H)=\Phi(C, H)=T$, и когда Λl требует поместить B и C в левый столбец таблицы, свойства α сохраняются. Тем самым правило Λl оправдано. Аналогично можно оправдать правила Nl в Nr, а также Λr.

Несколько особый случай с правилом Yr. Это правило заставляет ввести таблицу t', такую, что tRt' и B стоит справа в t', если в t $\Box B$ стоит справа $\Phi(\Box B, H)=F$. Но если $\Phi(\Box B, H)=F$, то существует $H': HRH'$ и $\Phi(A, H')=F$.

Тогда на $(n+1)$-м шаге положим $\alpha(t_i)=H'$ и удовлетворим тем самым всем требованиям.

Аналогичным образом поступим и с правилом Yl.

Лемма (без доказательства). Если построение для A не замыкается, то формула A необщезначима.

6.5.5. Полнота и разрешимость

В заключение приведем результаты, устанавливающие связь между доказуемостью формул в системах M, $S4$, $S5$ и брауэровой и общезначимостью, а также разрешимостью метода, использующего семантические таблицы.

Определение. Ассоциированная форма таблицы t на некотором шаге есть $A_1\&\ldots\&A_m\&\neg B_1\&\ldots\&\neg B_n$, где $A_1,\ldots, A_m$ — формулы, встречающиеся в t слева на этом шаге, $B_1,\ldots,B_n$ — формулы, встречающиеся справа.

Определение. Ранг таблицы для дерева таблиц, упорядоченного соотношением S на данном шаге, определяется так:

1) таблица t имеет в дереве ранг 0, если не существует таблицы t', такой, что tSt';

2) в противном случае, пусть $t_1,\ldots,t_n$ — все такие таблицы, что tSt_i, $i=1, \ldots, n$. Тогда $k(t)=\mathrm{Max}\{\mathrm{rank}(t_i)\}+1$.

Определение. Характеристическая форма таблицы t на некотором шаге определяется индукцией по рангу t:

1) $\mathrm{rank}(t)=0$—x, Φ — это просто ассоциированная формула;

2) $\mathrm{rank}(t)>0$. Пусть $t_1,\ldots,t_n$ — таблицы t_i, такие, что tSt_i. Для каждой t_i, такой, что $\mathrm{rank}(t_i)<\mathrm{rank}(t)B_i$, есть x, Φ, t_i. Пусть A—$a_1\Phi_1$ — таблицы t. Тогда x, Φ, t есть $A\&\Diamond B_1\&\Diamond B_2\&\ldots\&\Diamond B_n$. Характеристическая формула дерева — характеристическая формула его главной таблицы.

Лемма. Если A_0 является характеристической формулой начальной стадии конструкции, а B_0 — характеристической формулой на любом шаге, то $\vdash A_0\to B_0$.

Предложение. Если формула A общезначима (в M, $S4$, $S5$, брауэровой системе), то A доказуема (в соответствующей системе).

Приведем также результат о разрешимости метода семантических таблиц.

Предложение. Пусть A есть формула степени m в M или брауэровой системе. Тогда построение для A завершается. Оно замыкается, и тогда формула A общезначима либо приводит к конечному дереву контрмодели для A, в которой каждая ветвь дерева имеет длину $\leqslant m$. Следовательно, каждая формула доказуема либо имеет контрмодель с конечным деревом.

Тот же результат не сохраняется для $S4$, если не наложить дополнительного требования: никакое правило не должно применяться к формуле, входящей в повторенную таблицу.

6.6. ИНТЕНСИОНАЛЬНАЯ ЛОГИКА МОНТЕГЮ

Рассмотрим подробно «прагматическую» систему Р. Монтегю [72]. Под *прагматикой* понимается ветвь философии языка, обращающая внимание на возможные ситуации использования лингвистических выражений и объектов и на тех, кто их использует. Прагматика должна следовать образцу *семантики*, т. е. использовать такие понятия, как истина и выполнимость, но с учетом ситуации использования. Рассмотрим построение прагматического языка.

6.6.1. Прагматический язык

Символы прагматического языка L

1. Логические константы $\neg$, &, $\vee$, $\to$, $\leftrightarrow$, $\vee$, $\wedge$, $=$, E («неверно», «и», «или», «если…, то…», «если и только если», «некоторые», «все», «равно», «существует»).

2. Разделители (,), [,].
3. Индивидные переменные $v_0, \ldots, v_k \ldots$
4. Индивидные константы.
5. n-местные предикатные константы для каждого натурального n.
6. Операторы.

Индивиды будут рассматриваться как некоторые возможные объекты; тот факт, что индивид x *действительный*, будет обозначаться как $E[x]$. Операторы [74] понимаются как одноместные, это «необходимо», «случайно», «обычно», «будет так, что...» и т. д.

Формулы прагматического языка L

Множество формул L — наименьшее множество Г такое, что:

а) Г содержит все выражения:

$$E[\zeta], \quad \zeta=\eta, \quad p[\zeta_0, \ . \ . \ . \ , \zeta_{n-1}],$$

где ζ, ζ_0, ζ_{n-1} — индивидные константы или индивидные переменные, а p — n-местная предикатная константа.

б) Г замкнуто относительно применения логических связок;

в) выражения $\bigvee$ и φ, $\bigwedge$ и φ принадлежат Г, если u — индивидная переменная, а φ принадлежит Г;

г) $N\varphi$ принадлежит Г, если N есть оператор, а φ принадлежит Г.

6.6.2. Интерпретация прагматического языка

Неформально под *интерпретацией* L будем понимать следующее.

1. Центральным понятием *интерпретации* являются «возможные ситуации использования» выражений языка или *точки соотнесения*, следуя терминологии Д. Скотта (§ 6.7). Поэтому мы должны определить некоторое множество, элементами которого будут точки соотнесения.

2. Для каждой точки соотнесения определим множество объектов A_i, «существующих» по отношению к i.

3. Необходимо определить *смысл* или *интенсионал* каждой предикатной и индивидной константы нашего языка. Для этого каждой константе приписывается в каждой точке соотнесения i ее значение или *экстенсионал*. Так, для предикатных констант — это множество объектов, на которых предикат выполняется, для индивидных констант — значение индивида в данный момент. Например, экстенсионал «Нынешний король Франции» — личность, рассматриваемая как король Франции bi. Все множество таких личностей в каждой i будет интенсионалом данной константы.

4. Интерпретация каждого оператора N определяется как отношение между точками соотнесения и множествами точек соотнесения.

Обозначение. $\langle U_0, \ldots, U_{n-1}\rangle$-отношение есть подмножество множества $U_0 x \ldots U_{n-1}$; $\langle I, U_0, \ldots, U_{n-1}\rangle$*-предикат* есть функция, определенная на множестве I со значениями из множества всех $\langle U_0, \ldots, U_{n-1}\rangle$-отношений. Термин «отношение» соответствует экстенсионалу, а термин «предикат» — интенсионалу. При $n=0$ имеем дело с пустыми отношениями, Λ-отношениями. Единственные Λ-отношения — это пустое множество Λ и единичное множество $\{\Lambda\}$, обозначаемые как F и T. Тогда соответствующие $\{I\}$-предикаты будут *суждениями*, указанием на точку соотнесения.

Определим теперь понятие интерпретации более строго.

Возможная интерпретация для прагматического языка L есть тройка $\langle A, F, R\rangle$, такая, что:

1) A есть некоторая функция;

2) для каждого i из области определения A A_i есть множество [здесь и далее для любой функции AA_i $A_i \rightleftarrows A(i)$ — значение функции];

3) F — функция, область определения которой — множество предикатных и индивидных констант L;

4) если c — индивидная константа L, то F_c — функция с той же областью определения, что и A, причем для каждого j из области определения $F_c(j) \times \in \bigcup_{i \in DA} A_i$;

5) если p—n-местная предикатная константа L, то F_p—n-местный $\boldsymbol{D}A$-предикат на элементах $\bigcup_{i \in DA} A_i$. Здесь $\boldsymbol{D}A$ — область определения A;

6) R — функция, область определения которой есть множество операторов L;

7) если $N \in \boldsymbol{D}R$, то R_N есть $<\boldsymbol{D}_A, S\boldsymbol{D}_A>$-отношение, где $S\boldsymbol{D}A$ — степень множества (множество всех подмножеств) множества $\boldsymbol{D}A$.

Заметим, что мы приняли в качестве значений для константы не только индивиды, существующие в данной точке соотнесения (см. п. 4), но также любые возможные индивиды.

Условимся фразу «Предложение (т. е. формула без свободных переменных) φ истинно в точке соотнесения i при интерпретации $\mathfrak{A}$» обозначать «φ истинно i, $\mathfrak{A}$», а фразу «Возможный индивид x выполняет формулу φ в точке соотнесения i при интерпретации $\mathfrak{A}$» обозначать «x вып i, $\mathfrak{A}$ φ».

6.6.3. Неформальное понятие прагматической истины и выполнимости

Пусть $\mathfrak{A}$ — возможная интерпретация $\langle A, F, R\rangle$ языка L; $i \in \boldsymbol{D}A$, $x \in \underset{j \in DA}{U} A$, а P — *двухместная* (например) предикатная константа языка L.

Тогда можно предложить следующие *неформальные* критерии истины и выполнимости.

1) $P[c, d]$ истинно v, $\mathfrak{A}$, если и только если $\langle F_c(i),\ F_d(i)\rangle \in F_p(i)$;

2) x вып i, $\mathfrak{A}\, p[C, U]$, если и только если $\langle F_c(i),\ x\rangle \in F_p(i)$;

3) x вып i, $\mathfrak{A}\ C=U$, если и только если $F_c(i)$ совпадает с x;

4) x вып i, $\mathfrak{A}\ E\ (u)$, если и только если $x \in A_i$;

5) если φ есть предложение L, то $\neg\varphi$ истинно i, $\mathfrak{A}$, если и только если φ не истинно i, $\mathfrak{A}$;

6) если φ и ψ есть предложения L, то $(\varphi\&\varphi)$ истинно i, $\mathfrak{A}$, если и только если оба φ и ψ истинны i, $\mathfrak{A}$;

7) если φ есть формула L с единственной свободной переменной u, то V и φ истинны i, $\mathfrak{A}$, если и только если имеется $y \in \underset{j\in DA}{U} A_j$, такой, что у вып i, $\mathfrak{A}\,\varphi$;

8) если φ есть предложение L и N — оператор L, то $N\varphi$ истинно i, $\mathfrak{A}$, если и только если $\langle i, \{j/j \in DA$ и φ истинно i, $\mathfrak{A}\}\rangle \in R_N$.

В качестве примера к п. 8 примем, что $D\,A$ есть множество действительных чисел (например, моментов времени), а R_N — множество пар $\langle i,\ j\rangle$, таких, что $i \in DA$, $J \subseteq DA$ и существует $j \in J : j < i$ (иначе N будет интерпретироваться как «было так, что...»). Тогда $N\varphi$ истинно i, $\mathfrak{A}$, если и только если существует $j \in i$, такое, что φ истинно j, $\mathfrak{A}$. Точно так же можно рассматривать любые другие временные операторы, например будущее время.

Заметим также, что согласно п. 7 квантификация производится по всем возможным индивидам, а не только по действительным индивидам данной точки соотнесения.

6.6.4. Экстенсионалы и интенсионалы

Теперь перейдем к формальному определению экстенсионала. Полагаем при этом, что возможная интерпретация $\mathfrak{A} \rightleftharpoons \langle A, F, R\rangle$, $U \rightleftharpoons \bigcup_{j\in DA} A_j$, U^{ω} есть множество всех *бесконечных* (счетных) последовательностей, состоящих из членов U, $i \in DA$, а n — некоторое натуральное число.

Определение. Пусть ζ — индивидная переменная или индивидная константа L. Тогда *экстенсионал* ζ в i при интерпретации $\mathfrak{A}(E_x t_i,\ \mathfrak{A}\zeta)$ есть *функция* H с областью определения U^{ω}:

1) если ζ — переменная v_n и $x \in U^{\omega}$, то $H(x) = U_n$;

2) если ζ — индивидная константа и $x \in U^{\omega}$, то $H(x) = F_{\zeta}(i)$.

Иными словами, для каждой последовательности возможных значений экстенсионал «выделяет» единственное значение, а в случае константы это значение не зависит от такой последовательности (рис. 6.21).

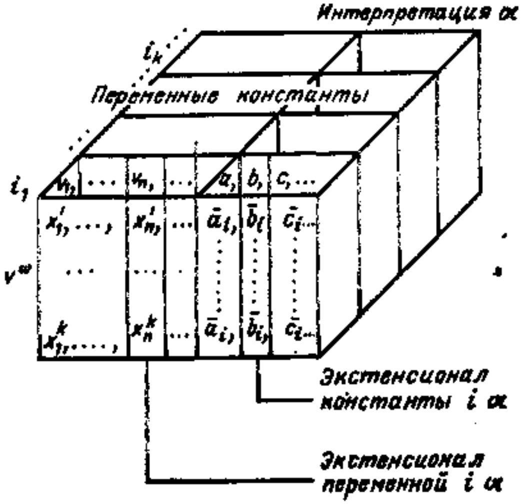


Рис. 6.21. К понятию экстенсионала переменной и константы

Определение. Введем экстенсионал формулы при интерпретации $\mathfrak{A}$.

1. Пусть ζ — индивидная константа или переменная L. Тогда $E_x t_i \mathfrak{A}(E[\zeta]) \rightleftharpoons \{x : x \in U^{\omega}$ и $(E_x t_i,\ \mathfrak{A}(\zeta))(x) \in A_i\}$.

2. Если ζ и η — индивидные константы или индивидные переменные L, то $E_x t_i,\ \mathfrak{A}(\zeta=\eta) \rightleftharpoons \{x : x \in U^{\omega}\ (E_x t_i \mathfrak{A}(\zeta))(x)$ тот же, самый, что и $(E_x t_i,\ \mathfrak{A}(\eta))(x)\}$.

3. Если p — n-местная предикатная константа L и $\zeta_0, \dots, \zeta_{n-1}$ — индивидные константы или индивидные переменные L, то $E_x t_i,\ \mathfrak{A}(P[\zeta_0, \dots, \zeta_{n-1}]) \rightleftharpoons \{x : x \in U^{\omega}$ и $<(E_x t_i,\ \mathfrak{A}(\zeta_0))(x), \dots, (E_x t_i,\ \mathfrak{A}(\zeta_{n-1}))(x) > \in F_p(i)\}$.

4. Если φ и ψ — формулы L, то $E_x t_i,\ \mathfrak{A}(\neg\varphi)$ есть $U^{\omega} - E_x t_i,\ \mathfrak{A}(\varphi)$; $E_x t_i,\ \mathfrak{A}(\varphi\&\psi)$ есть $E_x t_i \mathfrak{A}(\varphi) \cap E_x t_i,\ \mathfrak{A}(\psi)$.

Аналогично экстенсионал формулы вводится и для других пропозициональных связок.

5. Если φ — формула L, то $E_x t_i,\ \mathfrak{A}(\bigvee v_n \varphi) \rightleftharpoons \{x/x \in U^{\omega}$ и для некоторого $y \in U < x_0, \dots, x_{n-1},\ y,\ x_{n+1}, \dots > \in E_x t_i,\ \mathfrak{A}(\varphi)\}$. Аналогично определяется $E_x t_i,\ \mathfrak{A}(\bigwedge v_n,\ \varphi)$.

6. Если φ — формула L и N — оператор L, то $E_x t_i,\ \mathfrak{A}(N\varphi) \rightleftharpoons \{x/x \in U^{\omega}$ и $<i,\ \{j/j \in DA$ и $x \in E_x t_j,\ \mathfrak{A}(\varphi)\} > \in R_N\}$.

Определение. Если φ — предложение L, то φ *истинно* $i\mathfrak{A}$, если и только если $E_x t_i,\ \mathfrak{A}\varphi = U^{\omega}$.

Определение. Если φ — формула L, единственная свободная переменная которой v_n, то у вып i, $\mathfrak{A}\varphi$, если и только если существует $x \in E_x t_i,\ \mathfrak{A}(\varphi)$, такой, что $x_n = y$.

Таким образом, экстенсионал формулы в некоторой точке соотнесения есть множество последовательностей, «выполняющих» эту формулу, а экстенсионал индивидной константы или переменной в данной точке соотнесения есть функция, приписывающая возможный индивид каждой последовательности из U^ω (рис. 6.21).

Перейдем к понятию *интенсионала* для прагматики.

Определение. Пусть φ — индивидная константа, формула или индивидная переменная L. Тогда *интенсионал* ее в интерпретации $\mathfrak{A}(I_n t\,\mathfrak{A}(\varphi))$ есть функция H, определенная на $\mathbf{D}A$, такая, что для каждого $i \in \mathbf{D}A$ $H(i) = E_x t_i, \mathfrak{A}(\varphi)$.

Определение. Предложение φ есть *логическое следствие* множества предложений Γ, если и только если для каждого прагматического языка L и для всех $\mathfrak{A} \rightleftarrows <A, F, R>$ и $i \in \mathbf{D}A$; $F \cup \{\varphi\}$ есть множество предложений L, и для каждого $\psi \in F$, если ψ истинно $i\,\mathfrak{A}$, то φ истинно, i, $\mathfrak{A}$.

Предложение *логически истинно*, если оно есть логическое следствие пустого множества.

Предложение φ *логически эквивалентно* предложению ψ, если и только если предложение $(\varphi \leftrightarrow \psi)$ логически истинно.

Заметим, что при наших определениях в силу п. 6 *экстенсионал* формулы есть функция экстенсионала и интенсионалов. Этот принцип значительно шире, чем принцип зависимости экстенсионала формулы только от экстенсионалов частей [74].

Теперь перейдем к интенсиональной логике, рассмотрев предварительно интенсиональный язык.

6.6.5. Интенсиональный язык

Символы интенсионального языка

1. Логические константы прагматических языков.
2. Разделители (,) [,].
3. Индивидные переменные $v_0, \ldots, v_n \ldots$
4. Индивидные константы.
5. n-местные предикатные переменные $G_{0,n}, \ldots, G_{k,n} \ldots$ для каждого натурального n.
6. Предикатные константы типа S для каждой конечной последовательности S целых чисел $\geqslant -1$.
7. Оператор $\Box$ — «необходимо, что...».
8. Оператор дискрипции Т — «единственный такой, что...». Этот оператор рассматривается вместе с символами 1 и 7 как логическая константа.

Итак, как видно из этого определения, существенное отличие интенсионального языка от прагматического заключается уже в допущении предикатных переменных. Вводится типизация предикатных констант, тип указывает грамматическую категорию последовательностей ее аргументов, причем -1 соответствует индивидным символам, а неотрицательное число n — n-местной предикатной переменной.

Множество формул интенсионального языка

Множество формул языка L есть наименьшее множество Γ, такое, что:

1) Γ содержит выражения $E[\zeta]$;

$$\zeta = \eta; \qquad G[\zeta_0, \ldots, \zeta_{n-1}],$$

где ζ, η, $\zeta_0, \ldots, \zeta_{n-1}$ есть индивидные константы L, если индивидные переменные, и G есть n-местная предикатная переменная. Множество Γ содержит также выражения $P[\zeta_0, \ldots, \zeta_{n-1}]$, где P есть предикатная константа L типа $<S_0, \ldots, S_{n-1}>$, и для каждого $i<n$ либо $S_i \geqslant 0$ и ζ_i есть S_i-местная предикатная переменная, либо $S_i = -1$ и ζ_i есть индивидная константа L или индивидная переменная;

2) Γ замкнуто относительно применения логических связок;

3) Λ и φ и $\bigvee$ и φ принадлежат Γ, если φ принадлежат Γ и u есть индивидное или предикатное переменное;

4) $\Box\varphi$ принадлежит Γ, если φ принадлежит Γ;

5) если φ, ψ принадлежат Γ, а G есть предикатная переменная, то Γ содержит также результат *замены* в φ всех вхождений G, которые не следуют непосредственно за $\bigvee$, $\bigwedge$ или T на $TG\psi$.

Определение. Под *термом* понимается индивидная константа, индивидная переменная или выражение $TG\varphi$, где G — предикатная переменная, а φ — формула L.

Определение. Возможная интерпретация для интенсионального языка L есть пара $<A, F>$, удовлетворяющая пунктам 1—4 из определения интерпретации для прагматического языка и, кроме того, выполняющая условия 5′: если p — предикатная константа L типа $<S_0, \ldots, S_{n-1}>$, то F_p есть $<\mathbf{D}A, U_0, \ldots, U_{n-1}>$-предикат, где для каждого $i<n$ либо $S_i = -1$ и $U_i \rightleftarrows \underset{i \in \mathbf{D}A}{\subset} A_i$, либо $S_i \geqslant 0$ и U_i есть множество всех S_i — местных $\mathbf{D}A$ предикатов, заданных на элементах $\bigcup_{i \in \mathbf{D}A} A_i$.

Нетрудно заметить, что условие 5 определения для прагматического языка есть частный случай определения (5′) для интенсионального языка именно тогда, когда $S_0 = \ldots = S_{n-1} = -1$.

Теперь по аналогии с прагматическим языком неформально дадим понятие истинности и выполнимости.

6.6.6. Неформальное понятие интенсиональной истинности и выполнимости

Пусть $\mathfrak{A} \rightleftarrows <A, F>$ — возможная интерпретация интенсионального языка L, $i \in \mathbf{D}A$, $U_j \rightleftarrows \bigcup_{\in \mathbf{D}A} A_j$, $x \in U$, пусть p — предикатная константа языка L, c и d — индивидные константы, а u — индивидная переменная.

1) Пункты 1—7 критериев *прагматической* истины и выполнимости сохраняются;

8′) если φ — формула L, единственной свободной переменной которой является n-местная предикатная переменная G, то $\bigvee G\varphi$ истинно i, $\mathfrak{A}$, если и только если имеется n-местный $\mathbf{D}A$-предикат x на членах U, такой, что x вып $i\mathfrak{A}\varphi$;

9′) если G есть n-местная предикатная переменная, P — предикатная константа L типа $<n>$ и x — n-местный DA-предикат на членах U, то x вып i, $\mathfrak{A}P[G]$, если и только если $<x'> \in$ $\in F_p(i)$;

10′) если φ есть предложение L, то $\Box\varphi$ истинно i, $\mathfrak{A}$, если и только если φ истинно j, $\mathfrak{A}$ для всех $j \in \mathbf{D}A$;

11′) если G — n-местная предикатная переменная, $\mathscr{P}$ — предикатная константа L типа $<n>$ и φ — формула LG в качестве единственной свободной переменной, то $\mathscr{P}[TG\varphi]$ истинно i $\mathfrak{A}$, если и только если либо существует в точности один n-местный DA-предикат x на членах U, такой, что X вып $i\mathfrak{A}\varphi$, и этот предикат принадлежит $F_{\mathscr{P}}(i)$, либо неверно, что существует в точности один такой предикат и пустой предикат (т. е. $DAx\{\wedge\}$), принадлежащий $F_{\mathscr{P}}(i)$;

12′) если G есть 0-местная предикатная переменная и X есть $<DA>$-предикат [т. е. функция, определенная на множестве DA со значениями «истина» и «ложь» (не путать с DA-предикатом!)], то X вып $i\mathfrak{A}G$ [ ], если и только если пустая последовательность есть член $x(i)$, (т. е. $x(i)=\{\Lambda\}$).

Определение. Пусть $\mathfrak{A} \rightleftharpoons <A,\ F>$ — возможная интерпретация интенсионального языка L, а $U \rightleftharpoons \bigcup\limits_{i \in DA} A_i$. *Индивидным концептом* в $<A,\ F>$ называется функция, определенная на DA и принимающая значения из области U.

Индивидные концепты можно рассматривать как $<DA,$ $U>$-предикаты, выполняющие формулу

$$\Box U_u \Lambda_v (G[v] \leftrightarrow v = u).$$

6.6.7. Связь интенсиональной логики с другими логиками

В интенсиональных языках в отличие от прагматических языков имеется единственный случай, когда интенсионалы некоторых (косвенных) компонентов приходится рассматривать для определения экстенсионала всего выражения — именно интенсионалы компонентов, находящихся в области действия оператора $\Box$. Встает вопрос: можно ли ввести другие операторы прагматических языков в интенсиональные языки?

Пусть L — некоторый *прагматический* язык и $\mathfrak{A} \rightleftharpoons <A,\ F,$ $R>$ — некоторая возможная интерпретация для него. Пусть операторам N языка L взаимно однозначно соответствуют предикатные константы N' типа $<0>$. Пусть L — *интенсиональный* язык, все индивидные и предикатные константы которого содержатся в L вместе с символами N', соответствующими операторам N языка L. Пусть F' таково, что $<A,\ F'>$ — возможная интерпретация для интенсионального языка L', $F \leqslant F'$, а для каждого оператора N прагматического языка L и для каждого $i \in DAF'N'(i)$ — это $\{<U>/U$ есть $<DA>$-предикат $<i,\ \{j/j \in \mathrm{D}A$ и $U'(j)=\{\Lambda\}\}> \in R_N\}$.

Тогда легко доказывается: если φ — предложение L и если φ' получается из φ заменой каждой подформулы вида $N\psi$, где N есть оператор L, а ψ — формула L на подформулу

$$\bigvee G(\Box(G[\quad] \leftrightarrow \psi) \,\&\, N'[G])$$

и $i \in DA$, то φ истинно в i при *прагматической* интерпретации $<A,\ F,\ R>$, если и только если φ истинно в i при интенсиональной интерпретации $<A,\ F'>$.

Итак, рассматривая одноместные модальности, мы свели прагматику к интенсиональной логике. Интенсиональная логика, в свою очередь, может быть лишь частично сведена к прагматике.

Пусть L — интенсиональный язык, все предикатные константы которого имеют тип $<0>$ или $<S_0, \ldots, S_{n-1}>$, где $S_p = -1$ для всех $p < n$, и $\mathfrak{A} = <A,\ F>$ — интерпретация L. Пусть предикатные константы $\mathscr{P}$ языка L типа $<0>$ взаимно однозначно отображаются на операторы $\mathscr{P}'$ и пусть N — оператор, не входящий в их число. Пусть L' — прагматический язык с теми же индивидными константами, что и L, и теми же предикатными константами, не имеющими типа $<0>$, операторы которого — N и $\mathscr{P}''$, где $\mathscr{P}'$ — предикатная константа типа $<0>$. Пусть $<A,\ F',\ R>$ — возможная интерпретация L', $F \leqslant$ $\leqslant F'$, $R_N \rightleftharpoons \{<i,\ J>/i \in DA$ и $J = DA\}$ и для каждого предиката $\mathscr{P}: R_{\mathscr{P}} \rightleftharpoons \{<i,\ J>/i \in DA$ существует $\psi \in F_{\mathscr{P}}(i): J=\{j/j \in DA$ и $Y(j)=\{\Lambda\}\}\}$.

Тогда легко показать, что если $i \in DA$, φ — предложение L и φ' получается из φ заменой каждой подформулы $\mathscr{P}[\&\psi]$, где ψ — формула L на $\mathscr{P}'\psi$ и φ' есть предложение L, то φ истинно i, $<A,\ F>$ (т. е. при *интенсиональной интерпретации*), если и только если φ истинно i, $<A,\ F',\ R>$ (при *прагматической интерпретации*).

Соотношение между рассмотренными системами приблизительно следующее. Под *модальной логикой* понимается та часть интенсиональной логики, в которой не рассматриваются формулы с предикатными переменными. Стало быть, интенсиональная логика есть *модальная логика второго порядка*. Прагматика в некотором смысле в ней содержится: она может рассматриваться как первопорядковая часть интенсиональной логики.

Обращение к системам второго порядка потребовалось для применения интенсиональной логики к философским проблемам. Так, очевидно, что одноместные модальности в основном совпадают со свойствами суждений, и в интенсиональной логике могут рассматриваться благодаря применению предикатных переменных таких, как «убеждения».

Так, оператор «убежден» рассмотрим как предметную константу $\mathscr{B}$ типа $<-1, 0>$. Пусть $\mathfrak{A} \rightleftharpoons <A, F>$ — возможная интерпретация L. Тогда DA — множество возможных миров, $F_i(i)$ — экстенсионал константы с b_i, $<DA>$-предикат есть суждение в философском смысле, а интенсионал предложения — суждение, выражаемое этим предложением. Константа $\mathscr{B}$ есть сокращение для «убежден», и $F_\beta\ (i)$ есть множество пар $<X, U>$, таких, что x верит в суждение U в возможном мире i.

6.7. ИНТЕНСИОНАЛЬНАЯ ЛОГИКА СКОТТА С ТОЧКИ ЗРЕНИЯ СИСТЕМ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ

В заключение на примере рассмотрим, каким образом описанные семантические системы могут быть применены в задачах представления знаний. Попутно будет описана семантическая система, предложенная Д. Скоттом [75], и в ее терминах описана предметная область «поставщики и поставки». Система Скотта возникла, по-видимому, в результате глубокого анализа, в частности, трех описанных здесь систем, поэтому рассмотрение ее естественным образом подведет черту всему кругу обсуждавшихся нами проблем.

Интерес к интенсиональным логикам с точки зрения систем представления знания и, в частности, к системе Д. Скотта как к наиболее простой и удобной связан в первую очередь с возможностью учета точек зрения пользователей на предметную область, моменты времени, возможные миры и т. д. чего, конечно, не позволяют достичь традиционные системы представления знаний. Таким образом, применение этого математического аппарата дает возможность более адекватного отображения предметной области и перехода на уровне пользователя к языку более высокого уровня, т. е. еще на одну ступеньку приблизиться к естественному языку в системах представления знаний.

Возможность рассматривать высказывания, содержащие модальности типа «необходимо» и «возможно», позволит охватывать высказываниями состояния базы данных в различных точках соотнесения.

Рассмотрим предметную область «поставщики и поставки» — пример, распространенный в литературе по теории баз данных. База данных, отображающая данную предметную область, приводится на рис. 6.22.

Поставщики

Номер поставщика	Фамилия	Статус	Город
01	Смит	1	Нью-Йорк
02	Джонс	2	Лондон

Детали

Номер детали	Название	Цвет	Всего
3102	Гайка	Красный	1,5
3101	Болт	Зеленый	0,5

Поставки

Номер поставщика	Номер детали	Количество
01	3101	100
02	3102	100
01	3102	200

а)

Поставщики

Номер поставщика	Фамилия	Статус	Город
01	Смит	2	Сан-Франциско
03	Браун	1	Лос-Анджелес

Детали

Номер детали	Название	Цвет	Масса
3101	Болт	Зеленый	1,5
3103	Шайба	Синий	0,05

Поставки

Номер поставщика	Номер детали	Количество
01	3101	200
03	3101	100
03	3103	200

б)

Рис. 6.22. База данных «поставщики и поставки»:

а — состояние базы данных в точке соотнесения «понедельник»; *б* — состояние базы данных в точке соотнесения «вторник»

Предположим, что график поставок изменяется в зависимости от дня недели. Нам хотелось бы иметь в базе данных отображение этого графика и получить возможность оценивать высказывания, относящиеся к поставкам в различные дни недели, чего не допускают традиционные системы. Для простоты рассмотрим состояния БД, показанные на рис. 6.22, *а* и *б*.

При построении семантики будем следовать обозначениям и терминологии, принятой Д. Скоттом [74].

Индивиды. В нашей семантике рассматриваются три домена объектов (индивидов): возможные $|D|$, виртуальные $|V|$ и действительные $|A|$.

Множество *возможных индивидов* $|D|$ — множество всех объектов отображаемой предметной области. В нашем случае это все *поставщики* и все *детали*. Каким образом в БД должна храниться информация о возможных индивидах, рассмотрим ниже. Индивиды могут идентифицироваться некоторыми ключевыми атрибутами, например номером.

Кроме того, есть возможность ввести в рассмотрение «для увеличения регулярности в языке» некоторые «идеальные» объекты, подобные $\pm\infty$ в теории действительных чисел, которые вкупе с множеством D образуют область *виртуальных индивидов* $|V|$. Виртуальные объекты могут быть выведены из возможных с помощью некоторой процедуры (например, усреднения). В нашем случае можно, например, ввести объект «крепеж» в качестве обобщения для «гайка» и «болт», ввести для этого объекта некоторый номер, взять среднюю массу соответствующих возможных объектов и т. д. Важно лишь, чтобы для таких объектов было отмечено, что они имеют *другую природу*, нежели возможные.

Прежде чем переходить к действительным индивидам, введем новый домен — множество *I точек соотнесения* (индексов, возможных миров и т. д.). В любом выражении языка будет делаться ссылка на индекс. Одно и то же выражение оказывается истинным в одной точке соотнесения и ложным в другой.

Структура точек соотнесения может быть достаточно сложна. Так, для некоторого $i \in I$ допустимо

$$i = (\omega,\ t,\ p,\ a\ .\ .\ .),$$

где ω — возможный мир; t — время; p — трехмерное положение в мире x, y; a — субъект и т. д.

Для простоты предположим, что I интерпретируется как множество дискретных моментов времени или *дней недели*. Точки соотнесения будут в дальнейшем выступать в качестве элементов атрибутов некоторых специальных отношений.

Теперь каждому $i \in I$ ставим в соответствие некоторое множество *действительных индивидов* $A_i \subseteq D$, относящихся к данной точке соотнесения. Потребуем, чтобы $\bigcup A_i = D$. Теперь имеем $A_i \subseteq D \subseteq V$.

Введем в соответствии с этим два вида кванторов: $\forall x$ — переменная x пробегает по всем областям; $\forall . x$ — переменная пробегает лишь по некоторому множеству действительных индивидов A_i, причем в выражении языка $\forall . x \mathscr{P}(x)$ должно присутствовать явное указание на i. (Аналогично вводятся и кванторы существования.)

Положим $A_{пн}$ = {Смит, Джонс, гайка, болт} и т. д.

Теперь в каждом из дней недели можно оценивать истинность, например, высказывания «Поставщики находятся только в Нью-Йорке и Лондоне».

Перейдем к *структуре индивидов*. Введем новое множество *характеристик* индивидов относительно всех точек соотнесения. В нашем случае характеристики — это элементы декартовых произведений: Номер поставщика⊗Фамилия⊗Статус⊗Город и Номер детали⊗Название⊗Цвет⊗Масса.

Индивид a рассматривается как функция из множества точек соотнесения b множество характеристик:

$$a \in S^I,$$

где S^I — множество всех таких функций. Итак, в каждой точке соотнесения индивид характеризуется своими свойствами $a_i \in S$. Типичный пример функции-индивида — изменение статуса или города-поставщика в зависимости от момента времени.

Пропорциональные концепты. Истинность любого высказывания будем рассматривать лишь относительно некоторой точки соотнесения. В соответствии с этим каждое утверждение Φ сопоставим с функцией $\|\Phi\|$ (*значение* Φ в данной интерпретации), такой, что $\|\Phi\|_i = 1$ тогда и только тогда, когда Φ истинно *относительно* i.

Обозначив множество значений истинности $2=\{0,1\}$, видим, что $\|\Phi\|$ имеет тип

$$\|\Phi\|\in 2^{I}.$$

Такая функция называется пропозициональным концептом, так как Ф — замкнутая пропозициональная формула.

Пример пропозиционального концепта: $\|$*Смит поставляет гайки*$\|$ — истинно в точке соотнесения «понедельник» и ложно в точке соотнесения «вторник».

В языке допустимы также пропозициональные связки. Например, для конъюнкции имеем следующую семантику:

$$(\&)\,\|\Phi\,\&\,\psi\|_i=1\,\langle\Leftrightarrow\rangle\,\|\Phi\|_i=1 \quad \text{и} \quad \|\psi\|_i=1.$$

Теперь можно более строго сформулировать определение кванторов:

$$(\forall)\,\|\forall x\,\Phi(x)\|_i=1\,\langle\Leftrightarrow\rangle\,\|\Phi(\overline{a})\|_i=1 \quad \text{для} \quad \text{всех} \quad a\in\Delta;$$

$$(\forall.)\,\|\forall.\ x\,\Phi(x)\|_i=1\,\langle\Leftrightarrow\rangle\,\|\Phi(\overline{a})\|_i=1 \text{ для всех } a\in A_i,$$

где $\overline{a}$ — обозначение константы языка, соответствующей индивиду a в области D.

Индивидные концепты. Любому терму σ ставится в соответствие частично определенная функция $\|\sigma\|$, называемая *индивидным концептом* и имеющая тип $\|\sigma\|\in V^{(I)}$ (скобки используются для указания на частично определенную функцию). В каждой точке соотнесения значением функции является некоторый индивид $\|\sigma\|_i$. Индивидная константа является частным случаем терма, причем $\|\overline{a}\|_i=a$ в этом случае — значение индивидного концепта, которое не зависит от точки соотнесения. Примером индивидного концепта в нашем случае может быть $\|$*Поставщик с самым большим счетом в банке*$\|$. $\|\sigma\|_i$ рассматривается как смысл (интенсионал), а $\|\sigma\|$ — как значение (экстенсионал). Основной принцип семантики состоит в следующем: интенсионал целого выражения есть функция интенсионалов его частей.

Частным случаем индивидных концептов являются так называемые *дескрипции*. Семантика оператора дескрипции:

$$(I)\,\|Ix\,\Phi(x)\|_i=a\,\langle\Leftrightarrow\rangle\,\{a\}=\{b\in D\,|\,\|\Phi(\overline{b})\|_i=1\}.$$

(Буква I введена для обозначения индивида.)
Этот оператор указывает на тот единственный в каждой точке соотнесения индивид, который обладает некоторым свойством Ф. Для поддержания его, однако, не требуется вводить специальное отношение. Например: *Поставщик с самым высоким статусом.*

Отношения. Наряду с обычными (экстенсиональными) отношениями между индивидами вводятся более сложные — интенсиональные отношения — отношения между индивидными концептами, так что допустимы выражения вида $R(\sigma,\tau)$, где

$$\|R\|\in(2^{I})\,V^{(I)}\times V^{(I)}$$

(иначе $\|R(\sigma,\tau)\|=\|R\|(\|\sigma\|,\ \|\tau\|)$ в соответствии с основным принципом нашей семантики).

Пусть R — отношение ПОСТАВЛЯТЬ, тогда в качестве σ и τ можно поставить термы: σ-«Поставщик с самым большим счетом в банке» τ-«Самая дорогая деталь». Данные выражения будут истинными в одних точках соотнесения и ложными в других.

Допустимы и экстенсиональные отношения, семантика которых следующая:

$$(\overline{R})\,\|\overline{R}(\sigma,\ \tau)\|_i=1\,\langle\Leftrightarrow\rangle\,(\|\tau\|_i,\ \ \|\sigma\|_i)\in R_i$$

— константное отношение $(\hat{R})\,\|\tilde{R}(\sigma,\tau)\|_i=1<\Leftrightarrow>(\|\tau\|_i,$ $\|\sigma\|_i)\in R_i$ — переменное отношение. Так, если в качестве σ и τ брать индивиды, такие, как «Смит» и «гайка» соответственно, то получим обычное отношение ПОСТАВКИ, зависящее, однако, от точки соотнесения.

Интенсиональные операторы. Интенсиональные операторы позволяют выражать высказывания, относящиеся к различным точкам соотнесения. Наиболее распространенный — *оператор необходимости* $(\Box)\,\|\Box\Phi\|_i=1$ и $\|\Phi\|_j=1$ для всех $j\in$I. Пример его использования: *«Каждый день Смит поставляет болты»*. Аналогично можно ввести *оператор возможности.*

Более общим является *оператор альтернативности.* Пусть ρ — некоторое бинарное отношение, заданное на множестве I (отношение альтернативности). Тогда
$(\hat{\rho})\,\|\hat{\rho}\Phi\|_i=1<\Leftrightarrow>\|\Phi\|_j=1$ для всех $j\in I$, таких, что $i\rho j$. Пусть ρ — отношение *«быть завтрашним днем»*, тогда с помощью оператора $\hat{\rho}$ может быть, например, выражено утверждение *«Завтра Браун будет поставлять шайбы»*, истинное в точке соотнесения *«понедельник»*.

6.8. МОДЕЛЬНЫЕ МНОЖЕСТВА ХИНТИККИ

Рассмотрим кратко один из способов моделирования «возможных миров» при помощи *модельных множеств*, предложенных Хинтикки [60].

Для *немодальной* логики некоторое множество формул λ *выполнимо*, если и только если существует некоторое описание состояния, в котором справедливы все члены λ. Определим

[60] μ как множество формул, истинных в некотором отдельном описании состояния:

* (C.1) если p есть атомарная формула или равенство, то $p \in \mu$ либо $\neg p \in \mu$, но не то и другое вместе;

(C.2) если p — атомарная формула или равенство и если все свободные индивидные переменные p встречаются в других формулах μ, то либо $p \in \mu$, либо $\neg p \in \mu$;

* (C.3) если p — атомарная формула или равенство, если q отличается от p лишь тем, что a и b заменили друг друга в одном или нескольких местах, если $p \in \mu$ и если $a = b \in \mu$, то $q \in \mu$;

* (C.4) неверно, что $\neg (a = a) \in \mu$;

* (C.5) если $(p \& q) \in \mu$, то $p \in \mu$ и $q \in \mu$;

(C.6) если $p \in \mu$ и $q \in \mu$, то $(p \& q) \in \mu$;

* (C.7) если $(p \vee q) \in \mu$, то $p \in \mu$ или $q \in \mu$ (или оба);

(C.8) если $p \in \mu$ или $q \in \mu$ и если все свободные индивидные переменные $(p \vee q)$ встречаются в других формулах μ, то $(p \vee q) \in \mu$;

* (C.9) если $(Ex) p \in \mu$, то $p(a/x) \in \mu$, по крайней мере, для одной свободной индивидной переменной a;

(C.10) если $p(a/x) \in \mu$, по крайней мере, для одной свободной индивидной переменной a, то $(Ex) p \in \mu$;

* (C.11) если $(\bigcup x) p \in \mu$ и если b встречается, по крайней мере, в одной формуле μ, то $p(b/x) \in \mu$;

(C.12) если $p(b/x) \in \mu$ для каждой свободной индивидной переменной, которая встречается в формулах μ, то $(\bigcup x) p \in \mu$.

Были приняты следующие соглашения:

1) $p, q \ldots$ — произвольные формулы;

2) $a, b \ldots$ — произвольные свободные индивидные переменные;

3) x, y — произвольные *связанные* индивидные переменные;

4) $p(a/x)$ — результат замены в формуле p переменной x на a;

5) можно исключить все пропозициональные связки, кроме $\neg$, & и $\vee$;

6) все формулы могут быть приведены к такому виду, что знаки $\neg$ стоят только перед атомарными формулами или равенствами;

7) знак $\in$ есть сокращение для *«является членом»*.

Можно принять дополнительное условие:

(C.$\bigcup$) если $(\bigcup x) p \in \mu$, то $p(a/x) \in \mu$, по крайней мере, для одной свободной индивидной переменной a. Тем самым, исключим пустоту универсума рассуждения.

Некоторое множество формул *выполнимо*, если и только если оно может быть *включено* в множество, удовлетворяющее условиям C.1—C.12. Множество, удовлетворяющее *всем* условиям C.1—C.12, называется *полным онисанаем состояния.*

Часть условий C.1—C.12 можно исключить, оставив только условия, помеченные *.

Определение. Множество формул, удовлетворяющее условиям C.1, C.3—C.5, C.7, C.9, C.11, называется *модельным множеством*. Введем для этих условий обозначения (C. $\neg$), (C. $=$), (C. само$\neq$), (C.&), (C.$\vee$), (C.E) и (C.$\bigcup$) соответственно.

Предложение. Множество формул выполнимо, если и только если его можно включить в некоторое модельное множество μ.

Доказательство. 1) Доказательство «только если» тривиально. 2) Докажем «если».

Определение. Модельное множество μ называется *максимальным*, если не существует более широкого модельного множества $\nu \supset \mu$, такого, что каждая свободная индивидная переменная, встречающаяся в формулах ν, уже встречалась в формулах μ.

Лемма. Каждое модельное множество может быть включено в максимальное модельное множество.

Лемма. Каждое максимальное модельное множество является полным описанием состояния.

Лемма. Доказывается проверкой то, что если одно из условий C.2, C.6, C.8, C.10, C.12 не выполняется μ, то к μ можно добавить новую формулу так, чтобы получить более широкое модельное множество.

Из лемм следует, что любое множество формул, которое можно включить в модельное множество, можно включить и в максимальное модельное множество, т. е. в полное описание состояния.

Можно сказать, что модельное множество является как бы *частичным описанием* возможного состояния дел или «возможного мира». Из этого, однако, не следует, что данное состояние дел «действительно возможно» [60].

Для множества формул, содержащих модальные операторы, уже нельзя обойтись рассмотрением только одного модельного множества. При рассмотрении возможности и необходимости надо рассматривать состояние дел, отличное от действительного. Будем поэтому рассматривать множества модельных множеств или *модельные системы.*

Пусть Ω — модельная система, а M обозначает «возможно». Тогда, если $Mp \in \mu \in \Omega$, то p, быть может, неистинное в μ, должно быть истинно в другом состоянии дел или «альтернативе μ». Поэтому требуем: (C. M.*). Если $Mp \in \mu$, то в Ω существует, по крайней мере, одна альтернатива ν к μ, такая что $p \in \nu$.

Пусть N означает «необходимость». Тогда аналогично потребуем: (C. N). Если $Np \in \mu$, то $p \in \mu$. Этого, однако, недостаточно. Потребуем, чтобы p было истинно во всех «положениях дел»: (*С. N^+*). Если $Np \in \mu \in \Omega$ и если $\nu \in \Omega$ является альтернативой μ, то $p \in \nu$.

Определение. *Модельная система* есть пара $\langle \Omega, R \rangle$, первый член которой есть множество модельных множеств, каждое из которых выполняет условие (*С. N*). Второй член—двухместное отношение *альтернативности*. Кроме того, Ω и R должны удовлетворять условиям (С. M^*) и (С. N^+).

Определение. Множество λ формулы выполнимо, если и только если существует модельная система $\langle \Omega, R \rangle$, такая, что для некоторого $\mu \in \Omega$ имеет место $\mu \geqslant \lambda$.

Рассмотренная здесь семантическая система эквивалентна синтаксической системе M. Под эквивалентностью здесь понимается то, что если некоторая формула доказуема в M, она общезначима в семантической системе. Наложив требование транзитивности на отношение альтернативности, как это показано в п. 6.4.6, получим систему, эквивалентную $S4$. Потребовав симметричность R, получим семантический аналог синтаксической системы Брауэра (см. п. 6.4.1). Требование транзитивности и симметричности R приводит, как и в п. 6.4.7, к системе $S5$.

Глава 7

ТЕОРИЯ ТИПОВ

7.1. ОБЩИЕ СВЕДЕНИЯ

Существует несколько вариантов теории типов и пояснения по поводу причин ее возникновения. С позиций пользователя-кибернетика или пользователя-специалиста по системам искусственного интеллекта это прежде всего необходимо для построения или математического описания теории представления знаний (так же как и λ-исчисление) в виде фреймов. Иначе говоря, на теорию типов можно смотреть как на еще один шаг (или средство) для описания семантики, математического моделирования смысла информации. В различных кибернетических моделях, фреймах, семантических сетях, системах программирования появились не очень строго обоснованные понятия типов данных, в том числе абстрактные типы данных, в которых как бы различные значения данных, родственные друг другу переменные объединяются семантическим понятием типа. В качестве примера можно привести фрейм системы CSDL [36, 37]. Автор этой модели фреймов Руссопулос специально ввел в рассмотрение такой вид фрейма, чтобы использовать математический аппарат теории типов и других разделов математики, связанных с ним. В прямоугольнике (рис. 7.1, *а*) размещается предикат, описывающий действие с помощью глагола «загружать». Дуги связывают вершину-предикат с *переменными x, y, z* и несут семантическую нагрузку в виде ролевых падежей: объект, агент и т. д. Самое интересное для нас расположено вне прямоугольника, в овалах размещают *типы* (данных, переменных и т. д.): «преподаватель», «семестр», «лекции». Каждая переменная дугой с меткой отнесена к своему типу. Поэтому в описании модели CSDL очень часто встречаются слова «переменная X типа», «преподаватель». Стоит поменять типы (данных) «преподаватель» на «судно», «семестр» на «год», «лекции» на «зерно», как получится фрейм «загружать» (рис. 7.1, *б*).

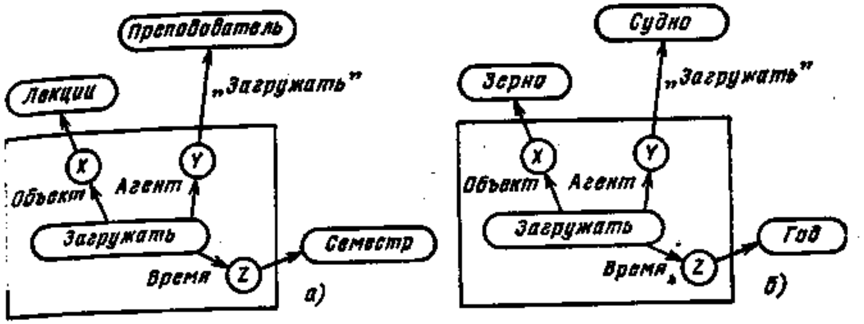


Рис. 7.1. Пояснение к понятию типа во фреймах:
а — «преподаватель загружается лекциями в течение семестра»; *б* — «судно загружается зерном в течение года»

В математике имеются в настоящее время два направления развития (или применения) теории типов. Первое — это в теории доказательств для построения исчисления предикатов выше первого порядка. Поэтому упрощенно можно считать, что теория типов — это теория предикатов от предикатов. И второе направление — для построения вариантов λ-исчисления с типами (типовое λ-исчисление) типовой комбинаторный логики, причем эти построения тесно связаны с теорией категорий и, в частности, с декартово-замкнутой категорией (ДЗК). Ниже рассмотрены оба направления, которые в общем случае не совпадают.

В соответствии с индуктивным методом изложения можно упрощенно считать, что теория типов — это исчисление предикатов высшего порядка, выше первого (или нулевого порядка). И прежде всего, теория типов отличается от исчисления предикатов первого порядка наличием предикатных (функциональных) символов выше первого порядка и всего, что с этим связано. Поэтому целесообразно вначале рассмотреть теорию

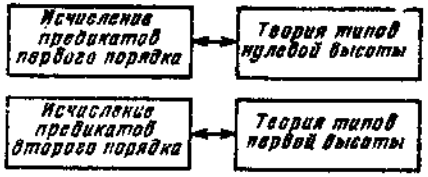


Рис. 7.2. Исчисление предиктов первого, второго и т. д. порядков и теория типов нулевой, первой и т. д. высоты

типов второго порядка (или первого) как продолжение исчисления предикатов первого порядка и расширение секвенциального исчисления. В дальнейшем термин «порядок» будет заменен на термин «тип», а первый, второй и т. д. порядок — на нулевую, первую и т. д. высоту типа (рис. 7.2) [62].

И одно из основных «занятий» теории типов — это переходы от низших типов к высшим и наоборот (рис. 7.3).

7.2. ИСЧИСЛЕНИЕ ПРЕДИКАТОВ ВТОРОГО ПОРЯДКА

Определение. Язык предикатов второго порядка (язык второго порядка) получается из языка исчисления предикатов первого порядка добавлением следующих символов двух типов переменных второго порядка [62]:

свободных i-местных переменных ($i=0, 1, 2\ldots$) $\alpha^i_0, \alpha^i_1, \ldots, \alpha^i_j\ldots$ ($j=0, 1, 2\ldots$);

связанных i-местных переменных ($i=0, 1, 2\ldots$) $\varphi^i_0, \varphi^i_1, \ldots, \varphi^i_j\ldots$ ($j=0, 1, 2\ldots$).

Переменные ($a_0, a_1, \ldots, x_0, x_1\ldots$), определенные в гл. 6, будем называть *переменными первого порядка,* чтобы отличить их от переменных второго порядка.

Термы определяются так же, как и в гл. 6. Для формальных переменных и для метапеременных будем также использовать одни и те же символы; буквы α, β, γ (с индексами и без них) будут использоваться для обозначения свободных переменных второго порядка, а буквы φ, ψ, χ — для обозначения связанных переменных второго порядка. Верхний индекс i у переменных α^i_j и φ^i_j будет обычно опускаться.

Определение. Формулы языка второго порядка определяются так же, как и § 6.3, с учетом следующих изменений: если R^i — некоторая i-местная предикатная константа или свободная переменная второго порядка, а $t_1, \ldots, t_i$ — термы, то $R^i(t_1, \ldots, t_i)$ — атомарная формула. В определении § 6.3 вместо «a — свободная переменная» и «X — связанная переменная» следует читать соответственно «a — свободная переменная первого порядка» и «X — связанная переменная первого порядка».

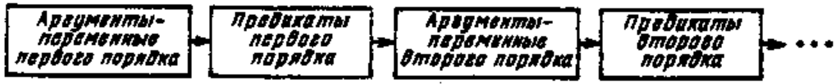


Рис. 7.3. Пояснение к системе типов как системе предиктов второго и более порядков

Рис. 7.4. Пояснение к понятию системы типов

Кроме того, к п. 3 определения добавляется пункт $3'$: если A — формула, α — свободная переменная второго порядка и φ — связанная переменная второго порядка, не входящая в A и имеющая то же число аргументов, что и α, то $\forall\varphi A'$ и $\exists\varphi A'$ — тоже формулы, где A' есть выражение, полученное из A заменой α на φ в каждом вхождении α в A. Внешними логическими символами формул $\forall\varphi A'$ и $\exists\varphi A'$ являются соответственно и $\forall$ и $\exists$.

Напомним, что речь идет об исчислении предикатов второго порядка, т. е. предикатов второго порядка, аргументами которых являются предикаты первого порядка.

Бескванторные формулы и замкнутые формулы (т. е. предложения) определяются так же, как и ранее.

Замещение символов и понятия отмеченных и полностью отмеченных вхождений данных символов определяется так же, как и в § 6.3. Таким образом, из $F(\alpha)$ получаем $F(R)$, заменив отмеченное вхождение α на R. Понятие алфавитного варианта определяется таким же образом, как и § 6.3 (только теперь связанные переменные должны замещаться другими связанными переменными того же порядка, а для переменных второго порядка, кроме того, с тем же числом аргументов). Секвенция — это выражение вида $\Gamma\to\Delta$, где Γ и Δ — конечные последовательности формул языка, который обозначим через L_2.

Вначале определим некоторую систему второго порядка, не содержащую никаких «аксиом выделения» и являющуюся просто системой LK (классической логикой, исчислением предикатов первого порядка) с переменными второго порядка. Так как эта система является основной среди систем второго порядка, назовем ее *базисным исчислением* для систем второго порядка и будем сокращенно обозначать через BC (рис. 7.4).

Определение. Формулы и секвенции системы BC — это формулы и секвенции языка L_2.

Заметим, что по мере развития теории типов мы все больше будем отрываться от предикатов, секвенциального исчисления и, «наладив» свойства переменных, заниматься дедуктивной теорией типов.

Правила вывода для системы BC такие же, как и для системы LK, т. е. к десяти правилам вывода § 6.3 необходимо добавить следующие правила.

Правила второго порядка для ∀:

∀-слева

$$\frac{F(R),\ \Gamma\to\Delta}{\forall\varphi F(\varphi),\ \Gamma\to\Delta},$$

где R — произвольная предикатная константа или свободная переменная второго порядка и φ имеет такое же число аргументных мест, что и R;

∀-справа

$$\frac{\Gamma\to\Delta,\ F(\alpha)}{\Gamma\to\Delta,\ \forall\varphi F(\varphi)},$$

где α — свободная переменная второго порядка, полностью отмеченная в $F(\alpha)$ и не входящая в нижнюю секвенцию, а φ — связанная переменная второго порядка, имеющая такое же число аргументных мест, что и α [разумеется, не входящая в $F(\alpha)$]. Переменная α называется *собственной переменной* этого правила.

Правила второго порядка для ∃:

∃-слева

$$\frac{F(\alpha),\ \Gamma\to\Delta}{\exists\varphi F(\varphi),\ \Gamma\to\Delta},$$

где α и φ — то же, что и в предыдущем случае.

∃-справа

$$\frac{\Gamma\to\Delta,\ F(R)}{\Gamma\to\Delta,\ \exists\varphi F(\varphi)},$$

где R — произвольная предикатная константа или свободная переменная второго порядка, а φ имеет такое число аргументных мест, что и R.

Боковые и главные формулы этих правил определяются так же, как и в остальных случаях. Правила 5 и 6 в § 6.3 в отличие от правил данного параграфа называются правилами первого порядка для ∀ и ∃ соответственно.

Определение. Выводы в системе BC и относящиеся к ним понятия определяются так же, как в § 6.3, таким образом, можно определить понятия:

«вывод, оканчивающейся секвенцией S, или вывод секвенции S»;

«секвенция S выводима»;

«нить секвенций»;

«секвенция, расположенная в выводе ниже или выше другой секвенции» и т. д.

Можно показать, что эта система непротиворечива (см. § 6.4). Аналогично гл. 6 можно доказать следующее:

Предложение. 1. $P(R)$ — некоторый BC-вывод секвенции $S(R)$, где R — произвольная предикатная константа или свободная переменная второго порядка. Пусть R' — произвольная предикатная константа или свободная переменная второго порядка, не входящая в $P(R)$. Предположим, что R и R' имеют одинаковое значение. Тогда $P(R')$ — вывод секвенции $S(R')$.

2. Вывод называется регулярным, если, во-первых, все собственные переменные в нем отличны друг от друга и, во-вторых, всякая переменная второго порядка, входящая в некоторую секвенцию S этого вывода в качестве собственной, входит, кроме того, только в секвенции, расположенные выше S. Тогда, если какая-то секвенция S выводима в BC, то она имеет в BC некоторый регулярный вывод.

Определение. Понятие системы аксиом определяется так же, как в § 6.3, т. е. система аксиом A (в языке L_2) — это произвольное множество предложений (языка L_2). Пункты 2—6 определения системы аксиом в § 6.3 можно приспособить для систем второго порядка. Предложение третьего раздела системы аксиом принимает здесь следующий вид.

Предложение. Пусть A — система аксиом и пусть BC A-система, получающаяся из BC добавлением в качестве начальных секвенций вида $\to A$ для всех A из A. Тогда секвенция $\Gamma\to\Delta$ выводима в BC, если только для некоторых предложений $A_1, \ldots, A_m$ из A секвенция $A_1, \ldots, A_m, \Gamma\to\Delta$ выводима в BC.

При рассмотрении систем второго порядка удобно работать с полутермами и полуформулами.

Определение. 1. Индивидные константы и переменные первого порядка (свободные и связанные) являются *полутермами:* если $t_1, \ldots, t_n$ — полутермы и f — некоторая n-местная функциональная константа, то $f(t_1, \ldots, t_n)$ — тоже полутерм.

2. *Полуформулы и свободные вхождения в них связанных переменных* определяются следующим образом.

Пусть R — некоторая i-местная предикатная константа или переменная второго порядка (свободная или связанная) и пусть $t_1, \ldots, t_i$ — полутермы. Тогда $R(t_1, \ldots, t_i)$ — атомарная полуформула, все связанные переменные полутермов $t_1, \ldots, t_i$ входят свободно в $R(t_1, \ldots, t_i)$, и, если R — связанная переменная, то R входит свободно в $R(t_1, \ldots, t_i)$. Если B и C — полуформулы, то $B\wedge C$ — тоже полуформула, и все связанные

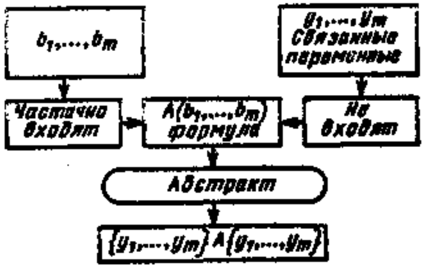


Рис. 7.5. Пояснение к определению абстракта формулы

переменные, входящие свободно в B или C, входят свободно в $B\wedge C$. Для остальных пропозиционных связок определение проводится аналогично. Если $F(x)$ — полуформула, в которой все вхождения связанной переменной первого порядка x являются свободными и отмеченными, то $\forall xF(x)$ — тоже полуформула, причем свободные вхождения связанных переменных в $\forall xF(x)$ те же, что и в $F(x)$, кроме вхождений переменной x. Если $F(\varphi)$ — полуформула, в которой все вхождения связанной переменной второго порядка φ являются свободными и отмеченными, то $\forall\varphi F(\varphi)$ — тоже полуформула и свободные вхождения в $\forall\varphi F(\varphi)$ — те же, что и в $F(\varphi)$, кроме вхождений переменной φ. Для $\exists$ определяется аналогично.

Очевидно, термы — это полутермы без связанных переменных, а формулы — это полуформулы, не имеющие свободных вхождений связанных переменных. Дадим определения абстракта и подстановки.

Определение. Пусть $A(b_1, \dots, b_m)$ — формула, в которой отмечены некоторые вхождения переменных $b_1, \dots, b_m$ (возможно, что какие-то из переменных $b_1, \dots, b_m$ не входят в эту формулу). Пусть $y_1, \dots, y_m$ — связанные переменные первого порядка, не входящие в $A(b_1, \dots, b_m)$. Тогда метавыражение $\{y_1, \dots, y_m\}A(y_1, \dots, y_m)$ называется *абстрактом* формулы $A(b_1, \dots, b_m)$ (рис. 7.5). Абстракты являются метавыражениями, а не формальными выражениями языка L_2 и будут использоваться только во вспомогательных целях.

Абстракт вида $\{y_1, \dots, y_m\}A(y_1, \dots, y_m)$ называется m-местным. Абстракты будут обозначаться буквами $V, U\dots$ Абстракт вида $\{y_1, \dots, y_m\}\ \alpha(y_1, \dots, y_m)$ будет часто отождествляться с α. Если V обозначает абстракт $\{y_1, \dots, y_m\}A(y_1, \dots, y_m)$ и $t_1, \dots, t_m$ — полутермы, то $V(t_1, \dots, t_m)$ обозначает полуформулу $A(t_1, \dots, t_m)$.

Определение. Подстановка абстракта вместо свободной переменной второго порядка в полуформулу определяется следующим образом. Пусть $F(\alpha)$ полуформула, где некоторые вхождения переменной α отмечены, и пусть V — абстракт, имеющий такое же число аргументных мест, что и α. (В дальнейшем не будем оговаривать последнее условие, поскольку подстановка определяется только для таких α и V, которые имеют одинаковое число аргументных мест). Определим *подстановку* V вместо α в $F(\alpha)$, результат которой обозначим через $F\binom{\alpha}{V}$ или $F(V)$. Для упрощения обозначений предполагаем, что α и V являются одноместными. Итак, пусть V имеет вид $\{y\}A(y)$. Тогда $F\binom{\alpha}{V}$ определяется индукцией по логической сложности формулы $F(\alpha)$ с помощью следующих положений:

1) (i) Если $F(\alpha)$ есть $\alpha(s)$ и это α отмечено в $F(\alpha)$, то (F^{α}_{V}) есть $A(s)$; (ii) если $F(\alpha)$ есть $\alpha(s)$ и это α не помечено или же $F(\alpha)$ есть $\beta(s)$ для некоторого β, отличного от α, то $F\binom{\alpha}{V}$ совпадает с $F(\alpha)$.

В остальных случаях вначале заменим все связанные переменные в F, входящие в V, на связанные переменные, не входящие в V, таким образом, чтобы каждая переменная заменялась на другую переменную того же порядка, различные переменные на различные, а i-местные связанные переменные второго порядка на другие i-местные переменные. Таким образом, можно предположить, что F не содержит связанных переменных, входящих в V.

2) $F(\alpha)$ — одна из формул $\neg B(\alpha)$, $B(\alpha)\wedge C(\alpha)$, $B(\alpha)\vee C(\alpha)$ или $B(\alpha)\supset C(\alpha)$. Тогда $F\begin{pmatrix}\alpha\\ V\end{pmatrix}$ есть $\neg B\begin{pmatrix}\alpha\\ V\end{pmatrix}$, $B\begin{pmatrix}\alpha\\ V\end{pmatrix}\wedge C\begin{pmatrix}\alpha\\ V\end{pmatrix}$, $B\begin{pmatrix}\alpha\\ V\end{pmatrix}\vee C\begin{pmatrix}\alpha\\ V\end{pmatrix}$ или $B\begin{pmatrix}\alpha\\ V\end{pmatrix}\supset C\begin{pmatrix}\alpha\\ V\end{pmatrix}$ соответственно.

3) $E(\alpha)$ имеет вид $\forall xG(x)(\alpha)$, $\exists xG(x)(\alpha)$, $\forall\varphi G(\varphi)(\alpha)$ или $\exists\varphi G(\varphi)(\alpha)$. Тогда $F\begin{pmatrix}\alpha\\ V\end{pmatrix}$ есть $\forall x\left(G(x)\begin{pmatrix}\alpha\\ V\end{pmatrix}\right)$, $\exists x\left(G(x)\begin{pmatrix}\alpha\\ V\end{pmatrix}\right)$, $\forall\varphi\left(G(\varphi)\begin{pmatrix}\alpha\\ V\end{pmatrix}\right)$ или $\exists\varphi\left(G(\varphi)\begin{pmatrix}\alpha\\ V\end{pmatrix}\right)$ соответственно. Очевидно, что $F\begin{pmatrix}\alpha\\ V\end{pmatrix}$ есть полуформула. Очевидно также, что если $F(\alpha)$ — формула, то $F\begin{pmatrix}\alpha\\ V\end{pmatrix}$ — тоже формула.

Имеющуюся в п. 2 и 3 определения подстановки неопределенность в выборе новых связанных переменных можно устранить, если попробовать, чтобы эти переменные были первыми в списке связанных переменных первого или второго порядка, которые удовлетворяют нужным условиям. Это условие не так существенно, так как справедливо следующее утверждение (предложение).

Предложение. Пусть A и B — любые две формулы, являющиеся алфавитными вариантами друг друга. Тогда в BC выводима формула $A\equiv B$.

Поэтому в дальнейшем может иметь дело с любым алфавитным вариантом данной формулы.

Пример 7.1. Пример имеет три части. 1. Пусть $F(\alpha)$ — формула $\forall x \forall y(x=y \supset (\alpha(x) \equiv \alpha(y)))$, где оба вхождения α отменены, и пусть V — абстракт $\{u\} \exists x(x+u=5)$, где подразумевается, что 5 — это индивидная константа, + — функциональная константа и = — предикатная константа языка L_2. Так как переменная x из $F(\alpha)$ входит в V, сначала заменим ее на другую связанную переменную, например $z: \forall z \forall y(z=y \supset (\alpha(z) \equiv \equiv \alpha(y)))$. Обозначим полученную формулу через $F'(\alpha)$.

Проведем шаг за шагом подстановку абстракта V в $F'(\alpha)$ вместо α:

$$\alpha(z)\binom{\alpha}{V}: \exists x(x+z=5);$$

$$\alpha(y)\binom{\alpha}{V}: \exists x(x+y=5);$$

$$(\alpha(y) \equiv \alpha(y))\binom{\alpha}{V}: \exists x(x+z=5) \equiv \exists x(x+y=5);\quad F'\binom{\alpha}{V},$$

т. е.

$$\forall z \forall z \forall y(z=y \supset (\alpha(z) \equiv \alpha(y)));$$

$$\binom{\alpha}{V}: \forall z \forall y(z=y \supset (\exists x(x+z=5) \equiv \exists y(y+z=5))).$$

Такая формула нам знакома — это аксиома равенства. Если бы мы предварительно не заменили x на z, то в результате получили бы выражение

$$\forall x \forall y(x=y \supset (\exists x(x+x=5) \equiv \exists x(x+y)=5)),$$

которое даже не является формулой. Этот пример можно обобщить для произвольного абстракта $\{u\}B(u)$ (при условии, что нет «коллизии» связанных переменных) и получить таким образом формулу $\forall x \forall y(x=y \supset B(x) \equiv \equiv B(y))$, представляющую собой аксиому равенства. Таким образом, из простой схемы

$$\forall x \forall y(x=y \supset (\alpha(x) \equiv \alpha(y)))$$

с помощью подстановки получаются все аксиомы равенства.

2. Пусть $F(\alpha)$ обозначает формулу $\alpha(0) \wedge \forall x(\alpha(x) \supset \alpha(x')) \supset \forall x\alpha(x)$, где все вхождения переменной α в $F(\alpha)$ отмечены, и пусть V есть абстракт $\{u\}B(u)$. Предположим, что x не входит в V. Тогда $F\binom{\alpha}{V}$ есть формула $B(0) \wedge \forall x(B(x) \supset B(x')) \supset \forall xB(x)$, представляющая собой аксиому индукции для арифметики (в соответствующем языке).

3. Пусть $F(\alpha)$ обозначает формулу

$$\forall x \forall y \forall z(\alpha(x, y) \wedge \alpha(x, z) \supset y=z) \supset \exists v \forall y(y \in v \equiv \exists x(x \in u \wedge \alpha(x, y))),$$

где все вхождения α в $F(\alpha)$ отмечены, и пусть V обозначает абстракт $\{x^1, y^1\}B(x^1, y^1)$, где $B(x^1, y^1)$ — полуформулы языка теории множеств. Тогда $F\binom{\alpha}{V}$ есть формула $\forall x \forall y \forall z(B(x, y) \wedge B(x, z) \supset y=z) \supset \exists v \forall y(y \in v \equiv \equiv \exists x(x \in u \wedge B(x, y)))$, представляющая собой аксиому подстановки в теории множеств. Заметим, что формула $B(x, y)$ может содержать кроме x и y и другие переменные, в том числе u, но не должна содержать v [поскольку v — связанная переменная в $F(\alpha)$]

Индукцией по числу логических символов в формуле $F(\alpha)$ легко доказывается следующее.

Предложение. Для любой формулы $F(\alpha)$ и произвольных абстрактов U и V в системе BC выводима секвенция

$$\forall x(U(x) \equiv V(x)), \qquad F(U) \to F(V).$$

при этом предполагается, что связанные переменные соответствующим образом переименованы.

Определение. В этом определении три раздела. 1. Пусть $A(b_1, \ldots, b_m, c_1, \ldots, c_n, \beta_1, \ldots, \beta_k)$ — формула, все свободные переменные которой содержатся среди переменных $b_1, \ldots, b_m, c_1, \ldots, c_n, \beta_1, \ldots, \beta_k$ (хотя необязательно все они входят в A), причем все вхождения этих переменных отмечены. Тогда всякое предложение вида

$$(*) \forall z_1, \ldots, \forall z_n \forall \psi_1, \ldots, \forall \psi_k \exists \varphi \forall y_1, \ldots, \forall \\ \forall y_m(\varphi(y_1, \ldots, y_m)) \equiv A(y_1, \ldots, y_m, z_1, \ldots, z_n, \\ \psi_1, \ldots, \psi_k)$$

называется *аксиомой выделения.*

Пусть V обозначает абстракт

$$\{y_1, \ldots, y_m\} A(y_1, \ldots, y_m, c_1, \ldots, c_n, \beta_1, \ldots, \beta_k).$$

Тогда приведенную выше аксиому выделения можно записать в другом виде:

$$(**) \forall z_1 \ldots, \forall z_n \forall \psi_1, \ldots, \forall \psi_k \exists \varphi \forall y_1, \ldots, \forall y_m \times \\ \times (\varphi(y_1, \ldots, y_m) \equiv U(y_1, \ldots, y_m)),$$

где абстракт U получается из V замещением переменных c_i и β_j переменными z_i и ψ_j соответственно $(i=1, \ldots, n;\ j=1, \ldots, k)$.

2. Пусть K — произвольное множество формул. Формулу, принадлежащую множеству K, назовем *K-формулой,* и если $A(b_1, \ldots, b_m)$ есть K-формула, то абстракт $\{y_1, \ldots, y_m\} A(y_1, \ldots, y_m)$ назовем *K-абстрактом.* Если формула A в аксиоме выделения (см. $*$) является K-формулой, то предложение $(*)$ называется *аксиомой K-выделения.*

3. Множество формул K называется *замкнутым относительно подстановок,* если для всякого K-абстракта V формула $A(V)$ тоже принадлежит K.

Определение. Пусть K — некоторое множество формул. 1. K-система получается из системы BC добавлением к ней в качестве начальных секвенций всех аксиом K-выделения (т. е. всех секвенций вида $\to A$, где A — аксиома выделения).

2. Система KC получается из системы BC добавлением следующих правил вывода, где $F(\alpha)$ — произвольная формула,

V — произвольный K-абстракт, а $F(V)$ и $F(\varphi)$ получаются из $F(\alpha)$ подстановкой V и φ соответственно вместо отмеченных вхождений α:

правило $\forall$-слева второго порядка

$$\frac{F(V),\ \Gamma\to\Delta}{\forall\varphi F(\varphi),\ \Gamma\to\Delta};$$

правило $\exists$-справа второго порядка

$$\frac{\Gamma\to\Delta,\ F(V)}{\Gamma\to\Delta,\ \exists\varphi F(\varphi)}.$$

Главные и боковые формулы этих правил вывода определяются как обычно.

Так как система KC представляет интерес, только когда множество K замкнуто относительно подстановок, будем предполагать, что K удовлетворяет этому условию.

Предложение. Для произвольного множества формул K (замкнутого относительно подстановок) система KC эквивалентна K-системе.

Доказательство. Можно показать, что аксиомы K-выделения выводимы в KC, а с другой стороны, нижние секвенции правил второго порядка $\forall$-слева и $\exists$-справа выводимы из соответствующих верхних секвенций в K-системе.

Определение. Если K — множество всех формул языка второго порядка, то система KC называется исчислением предикатов второго порядка с полным выделением и обозначается через G^1LC.

Предложение. Если для системы G^1LC верна теорема об устранении сечений, то эта система непротиворечива.

7.3. ПРОСТАЯ ТЕОРИЯ ТИПОВ

Этот вариант теории типов ориентирован на построение исчисления предикатов высшего (конечного) порядка, которое будет строиться в виде исчисления секвенций. Кроме того, ограничимся рассмотрением только предикатных букв. Введем термин «тип» (переменной). Иногда применяется в системах второго порядка термин «порядок» (переменной). Отсчет типов начнем с 0, а не с 1. Таким образом, индивидные объекты имеют тип 0 [62].

Определение. 1. Типы определяем индуктивно: 0 есть тип, если $\tau_1, \ldots, \tau_n$ — типы ($k\geqslant 1$), то $[\tau_1, \ldots, \tau_k]$ — тоже тип; типами являются только те объекты, которые получаются таким образом.

2. Символы нашего языка классифицируются следующим образом.

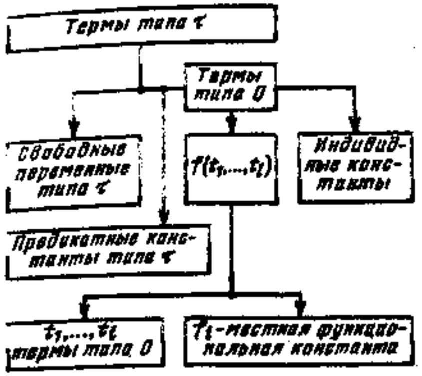


Рис 7.6. Пояснение к определению термов типа τ

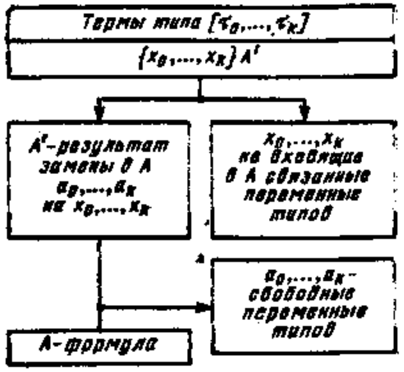


Рис. 7.7. Пояснение к определению типа $[\tau_0, \ldots, \tau_k]$

Константы:
индивидные константы $c_0, c_1\ldots$;
функциональные константы (i-местные)

$$b_0^i,\ b_1^i\ .\ .\ .\qquad (i=1,\ \ 2\ .\ .\ .);$$

предикатные константы типа τ

$$(\tau\neq 0): R_0^\tau;\ R_1^\tau\ .\ .\ .$$

Переменные:
свободные переменные a_0^τ, a_1^τ ... для каждого типа τ;
связанные переменные x_0^τ, x_1^τ ... для каждого типа τ.

Логические символы $\neg$ (не), $\wedge$ (и), $\vee$ (или). $\supset$ (влечет), $\forall$ для всех, $\exists$ существует;

Вспомогательные символы (,), {,}, [,].

Язык высшего порядка (язык простой теории типов) задан, когда заданы все его константы. Под предикатной переменной понимается любая переменная (свободная или связанная) типа $\tau\neq 0$. Символы языка будут использоваться в качестве метапеременных. Верхние индексы, обозначающие типы, иногда будут опускаться. Кроме того, остаются справедливыми все соглашения и принятые в § 6.3 обозначения.

Интуитивно переменные типа 0 обозначают индивидные объекты, а переменные типа $[\tau_1, \ldots, \tau_k]$ — предикаты, которые сопоставляются с подмножествами множества $T_1\times\ldots\times T_k$, где T_i — множество объектов типа τ_i.

Определение. *Термы* (данных типов), *формулы* и их *внешние логические символы* определяются (рис. 7.6) одной и той же следующей индукцией:

1. Индивидные константы — термы типа 0.

Рис. 7.8. Пояснение к определению атомарной формулы

2. Свободные переменные типа τ — термы типа τ.

3. Если f — некоторая i-местная функциональная константа и $t_1, \ldots, t_i$ — термы типа 0, то $f(t_1, \ldots, t_i)$ — термы типа 0.

4. Предикатные константы типа τ — термы типа τ.

5. Если A — формула, $a_0^{\tau_0}, \ldots, a_k^{\tau_k}$ — различные свободные переменные указанных типов и A' — результат одновременной замены в A переменных $a_0, \ldots, a_k$ на $x_0, \ldots, x_k$ соответственно, то $\{x_0, \ldots, x_k\}\ A'$ — терм (называемый также *абстрактом*) типа $[\tau_0, \ldots, \tau_k]$ (рис. 7.7).

6. Если α — предикатная константа или свободная переменная типа $[\tau_1, \ldots, \tau_k]$ и $t_1, \ldots, t_k$ — термы типов соответственно, то $\alpha[t_1, \ldots, t_k]$ — формула, называемая *атомарной*. Она не имеет внешнего логического символа (рис. 7.8).

7. Если A и B — формулы, то $(\neg A)$, $(A \wedge B)$, $(A \vee B)$, $(A \supset B)$ — формулы, и их внешними логическими символами являются соответственно $\neg$, $\wedge$, $\vee$, $\supset$.

8. Если A — формула, a^τ — свободная переменная, x^τ — не входящая в A связанная переменная того же типа и A' получается из A заменой всех вхождений a^τ на x^τ, то $\forall x^\tau A'$ и $\exists x^\tau A'$ — формулы, и их внешними логическими символами являются $\forall$ и $\exists$ соответственно.

Эти правила образования формул и термов могут привести к чрезмерно большому количеству скобок. Поэтому некоторые из скобок будут опускаться, если это не приводит к двусмысленности.

Заметим, что в отличие от § 7.2 абстракты здесь понимаются как формальные объекты.

Понятие *алфавитного варианта* определяется так же, как и раньше: выражение A называется алфавитным вариантом выражения B (и наоборот), если A и B отличаются только наименованием своих связанных переменных.

Определение. *Высота* типа определяется индуктивно следующим образом:

$$h(0) = 0; \qquad h([\tau_1, \ldots, \tau_k]) = \max(h(\tau_1), \ldots, h(\tau_k)) + 1.$$

Под высотой $h(t)$ терма t понимается высота его типа.

Логическая сложность формулы или абстракта A определяется как общее число логических символов и пар символов абстракции $\{,\}$, входящих в A.

Подстановка в формулу (в абстракт) A терма t типа τ вместо свободной переменной a типа τ определяется двойной индукцией: по высоте типа τ и по сложности формулы (абстракта) A. Результаты такой подстановки обозначим через $A\binom{a}{t}$.

Далее следуют два раздела определения высоты типа.

1. Базис: высота типа τ равна 0, т. е. тип τ есть 0. Тогда $A\binom{a}{t}$ определяется просто как $A\binom{a}{t}$ в соответствии с определением п. 6.3.4 или как некоторый алфавитный вариант этого выражения.

Пусть a и b — некоторые свободные переменные типа 0 и пусть t, s — некоторые термы типа 0.

Можно доказать следующее:

1) если A — формула (терм типа τ), то $A\binom{a}{t}$ — формула (терм типа τ);

2) если A — алфавитный вариант B, то $A\binom{a}{t}$ — алфавитный вариант $B\binom{a}{t}$;

3) $A\binom{a}{t}$ содержит только те свободные переменные, которые входят в A и в t;

4) если s не содержит a, то $A\binom{a}{t}\binom{b}{s}$ совпадает с

$$A\begin{pmatrix} b \\ \\ s \end{pmatrix}\begin{pmatrix} a \\ \\ t\begin{pmatrix} b \\ s \end{pmatrix} \end{pmatrix}.$$

2. Шаг индукции: допустим, что $h(\tau) = n \neq 0$ и для всех $m \leqslant n$ уже определена подстановка терма t типа τ (такого, что $h(\sigma) = m$) вместо свободной переменной a типа σ так, что выполняются условия:

1) если A — формула (терм типа σ), то $A\binom{a}{t}$ — формула (терм типа σ);

2) если A — алфавитный вариант B, а s — алфавитный вариант t, то $A\binom{a}{s}$ — алфавитный вариант $B\binom{a}{t}$;

3) $A\binom{a}{t}$ содержит только те свободные переменные, которые входят в A или в t;

4) $A\,(\{x_1, \ldots, x_k\}\, a[x_1, \ldots, x_k])$ есть алфавитный вариант A;

5) $\{x_1, \ldots, x_k\}\, a\,[x_1, \ldots, x_k]$ есть алфавитный вариант t.

6) если s не содержит a и высота b меньше h, то

$$A\binom{a}{t}\quad\binom{b}{s}$$

есть алфавитный вариант

$$A\binom{b}{s}\begin{pmatrix} a \\ t\binom{b}{s}\end{pmatrix};$$

7) если A не содержит a, то $A\binom{a}{s}$ есть алфавитный вариант A.

Пусть a и t — соответственно свободная переменная и терм типа τ, такие, что $h\,(\tau) = n$.

Если t — свободная переменная или предикатная константа, то $A\binom{a}{t}$ снова определяется как некоторый алфавитный вариант $A\binom{a}{t}$. Допустим, что t — некоторый абстракт. Определим $A\binom{a}{t}$ индукцией по сложности A. Пусть t есть $\{x_1, \ldots, x_k\}\, U\,(x_1, \ldots, x_k)$, где x_i имеет тип τ_i, a имеет тип $\tau\{\tau_1, \ldots, \tau_k\}$ и $\max\,(h\,(\tau_1), \ldots, h\,(\tau_k)) + 1 = n$. Отметим сначала, что для всякого терма s типа 0 $s\binom{a}{t}$ определяется как s.

Для доказательства справедливости условий 1—7 заметим, что:

2.1) если A имеет вид $b[t_1, \ldots, t_k]$, где b — некоторая предикатная константа или переменная, отличная от a, то

$$A\binom{a}{t}\text{— это в }\left[t_1\binom{a}{t}, \ldots, t_k\binom{a}{t}\right];$$

2.2) если A имеет вид $a[t_1, \ldots, t_k]$, то

$$A\binom{a}{t}\text{— это } U\,(b_1, \ \ldots, \ b_k)\,\times$$

$$\times\begin{pmatrix} b_1 \\ t_1\binom{a}{t}\end{pmatrix}, \ldots, \begin{pmatrix} b_k \\ t_k\binom{a}{k}\end{pmatrix},$$

где $b_1, \ldots, b_k$ отличны от всех свободных переменных в A, и

$$U\,(b_1, \ \ldots, \ b_k)$$

есть

$$\left(U\,(x_1, \ \ldots, \ x_k)\,\frac{x_1, \ \ldots, \ x_k}{b_1, \ \ldots, \ b_k}\right).$$

Здесь $h\,(b_i) < n$ и каждый t_i «проще», чем A, поэтому

$$B\begin{pmatrix} b_i \\ t_i\binom{a}{t}\end{pmatrix}$$

уже определено для произвольного B;

2.3) если A имеет вид $\neg B$, $B \wedge C$, $B \vee C$ или $B \supset C$, то $A\binom{a}{t}$ — это

$$\neg B\binom{a}{t}, \qquad B\binom{a}{t} \wedge C\binom{a}{t}, \qquad B\binom{a}{t} \vee C\binom{a}{t}$$

или $B\binom{a}{t} \supset C\binom{a}{t}$ соответственно;

2.4) если A имеет вид $\forall x F\,(x)$ или $\exists x F\,(x)$, то $A\binom{a}{t}$ — это соответственно $\forall y G\,(y)$, где $G\,(b)$ есть $F\,(b)\binom{a}{t}$, b отлично от a и не входит в $G\,(b)$;

2.5) если A имеет вид $\{x_1, \ldots, x_k\}$ В $(x_1, \ldots, x_k)$, то $A\binom{a}{t}$ — это $\{y_1, \ldots, y_k\}\, C\,(y_1, \ldots, y_k)$, где $C\,(b_1, \ldots, b_k)$ есть $B(b_1, \ldots, b_k)\binom{a}{t}$, $b_1, \ldots, b_k$ отличны от a и не входят в A и ни одно из y_i не входит в $C\,(b_1, \ldots, b_k)$. Тогда для $h\,(a) = n$ можно доказать справедливость свойств 1—7 (доказательство этих свойств опускается).

Часто вместо $A\binom{a}{t}$ будет использоваться обозначение $A\,(t)$. Буквы U, V с верхними индексами используются для типов, а без таких индексов — в качестве метапеременных для абстрактов. Кроме того, для обозначения свободных переменных часто будем использовать буквы α, β... вместо букв a, b..., а для обозначения связанных переменных — буквы φ, ψ... вместо x, y... (как правило, в тех случаях, когда имеем в виду переменные типа $\neq 0$).

Пример 7.2. Пусть A есть формула $\forall\varphi \exists x(\alpha[a] \equiv \varphi[x])$, где φ и α имеют тип [0], а x и a — тип 0. Пусть V — абстракт $\{z\} \forall\varphi \exists x(\varphi[x] \wedge \beta[z])$, где φ и β имеют тип [0], а x и z — тип 0. Тогда V — абстракт типа [0].

Найдем формулу $A\binom{\alpha}{V}$. Подстановка выполняется поэтапно в соответствии с определением высоты типа. Так как α имеет тип [0], то начинаем с п. 2 этого определения и неоднократно возвращаемся к п. 1. В силу п. 2.2 и 1 $\alpha[a]$ есть

$$\forall\varphi\exists x(\varphi(x)\land\beta[a])\equiv\gamma(b).$$

Отсюда в силу п. 2.4 получаем, что $A\binom{\alpha}{V}$ есть

$$\forall\psi\exists y(\forall\varphi\exists x(\varphi[x]\land\beta[a])\equiv\psi[y]).$$

Упражнение. Пусть A — следующий абстракт:

$$\{y\}\alpha^2[\{z\}(\alpha^1[z]\equiv\beta^1[y])],$$

и пусть V^2 есть

$$\{\psi^1\}\forall\varphi^1\exists x(\psi^1[a]=\varphi^1[x]),$$

где $1=[0]$ и $2=[1]$. Найти $A\binom{\alpha^2}{V^2}$.

Определение. Пусть V — некоторый абстракт вида

$$\left\{x_1^{\tau_1},\ .\ .\ .\ ,\ x_n^{\tau_n}\right\}A\left\{x_1^{\tau_1},\ .\ .\ .\ x_n^{\tau_n}\right\}$$

и пусть $v_1,\ \dots,\ v_n$ — термы типов $\tau_1,\ \dots,\ \tau_n$ соответственно.

Тогда $V[v_1,\ \dots,\ v_n]$ определяется как

$$A\left(a_1^{\tau_1},\ .\ .\ .\ ,\ a_n^{\tau_n}\right)\binom{a_1^{\tau_1}}{V_1}\ .\ .\ .\ \binom{a_n^{\tau_n}}{V_n},$$

где $a_1^{\tau_1},\ .\ .\ .,\ a_n^{\tau_n}$ —свободные переменные указанных типов, не входящие в термы $v,\ v_1,\ \dots,\ v_n$, а $A\left(a_1^{\tau_1},\ .\ .\ .,\ a_n^{\tau_n}\right)$ есть

$$\left(A\left(x_1^{\tau_1},\ .\ .\ .\ ,\ x_n^{\tau_n}\right)\frac{x_1^{\tau_1},\ .\ .\ .\ ,\ x_n^{\tau_n}}{a_1^{\tau_1},\ .\ .\ .\ ,\ a_n^{\tau_n}}\right).$$

Определение. Формальная система *простой теории типов* определяется аналогично системе G^1LC, рассмотренной в §7.2 (исчисление предикатов второго порядка). Секвенции, как обычно, имеют вид $\Gamma\to\theta$, где Γ и θ состоят из конечного числа формул. Правила вывода в ней такие же, как и в системе G^1LC, со следующим обобщением. Здесь повторим с комментариями из предыдущих разделов все нужные определения и обозначения.

Правило $\forall$-слева:

$$\frac{F(V),\ \Gamma\to\theta}{\forall\varphi F(\varphi),\ \Gamma\to\theta},$$

где V — произвольный терм (любого типа) и φ — связанная переменная того же типа, что и V [если $F(V)$ есть $F\binom{\alpha}{V}$, то $F(\varphi)$ есть $F\binom{\alpha}{\varphi}$].

Правило $\forall$-справа:

$$\frac{\Gamma\to\theta,\ F(\alpha)}{\Gamma\to\theta,\ \forall\varphi F(\varphi)},$$

где α не входит в нижнюю секвенцию и φ имеет тот же тип, что и α. Правила $\exists$-слева и $\exists$-справа определяются симметрично.

Всякое предложение вида

$$\forall y_1,\ .\ .\ .\ ,\ \forall y_m\exists\varphi\forall x_1,\ .\ .\ .\ ,\ \forall x_n(\varphi(x_1,\ .\ .\ .\ ,\ x_n)\equiv$$
$$\equiv A(x_1,\ .\ .\ .\ ,\ x_n,\ y_1,\ .\ .\ .\ ,\ y_m)),$$

где $A(x_1,\ \dots,\ x_n;\ y_1,\ \dots,\ y_m)$ — произвольная формула (и переменные $a_1,\ \dots,\ a_n,\ b_1,\ \dots,\ b_m$ имеют произвольный тип), называется *аксиомой выделения*.

Предложение. Все аксиомы выделения (для произвольных A) выводимы в простой теории типов. Рассмотрим подсистему простой теории типов типов, в которой правила $\forall$-слева и $\exists$-справа разрешается применять, только тогда V — свободная переменная или предикатная константа. В этой подсистеме, пополненной аксиомами выделения для всех A, допустимы общие правила $\forall$-слева и $\exists$-справа.

Определение. 1. *Полуформулами* называются выражения, построенные так же, как и формулы, и отличающиеся от них лишь тем, что полуформулы могут содержать свободные вхождения связанных переменных. *Полутермы* определяются аналогично.

2. *Логическая сложность* полуформул и полутермов определяется так же, как и для формул и абстрактов (см. определение высоты типа).

7.4. ОБЩАЯ ТЕОРИЯ ТИПОВ

В общем случае применительно к λ-исчислению с типами [38, 82], комбинаторной логике с типами, теории категорий (точнее, ДЗК) понятие типа вводится следующим образом.

Определение. Множество *типов*, обозначаемое Тур, индуктивно [38] определяется следующим образом:

$$0\in\mathrm{Typ};\quad (1)$$

$$\sigma,\tau\in\mathrm{Typ}\Rightarrow(\sigma\to\tau)\in\mathrm{Typ}.\quad (2)$$

Это определение интерпретируется (поясняется) следующим образом.

Определение. Пусть X — любое множество. Определим множества X_σ индукцией по $\sigma \in$ Тур:

$$X_0 = X;$$

$X_{\sigma\to\tau} = X_\tau^{X_\sigma}$ — совокупность всех теоретико-множительных функций из X_σ в X_τ. Это определение очень напоминает одно из свойств ДЗК, называемое «допускает экспоненсирование», не выходя из ДЗК (см. гл. 4). Это позволяет доказать, что ДЗК является некоторой системой типов.

В целом система типов является вместе с теорией категорий одним из способов описания, моделирования множеств.

Замечание. (*i*). В литературе применяются обозначения, отличные от $(\sigma\to\tau)$, например $(\sigma\tau)$, $(\sigma)\tau$, $\sigma(\tau)$ или $F\sigma\tau$.

(*ii*). 0 иногда называют *основным типом* (или *типом индивидов*). Иногда допускаются несколько основных типов.

(*iii*). $\sigma_1\to\sigma_2\to\ldots\to\sigma_n$ — это сокращение для $(\sigma_1\to(\sigma_2\to\to(\ldots\to\sigma_n)))$. Заметим, что каждый тип имеет вид $\sigma=\sigma_1\to\ldots\to\sigma_n\to 0$, где $n\geqslant 0$.

(*iiii*). Иногда правило 2 из определения типов заменяется правилом

$$\sigma_2, \ldots, \sigma_n, \qquad \tau\in \text{Тур} \Rightarrow (\sigma_1, \ldots, \sigma_n; \tau)\in \text{Тур}.$$

Интерпретация задается равенством

$$X_{(\sigma_1, \ldots, \sigma_n, \tau)} = X_\tau^{X_{\sigma_1}\times\ldots\times X_{\sigma_n}},$$

где × обозначает декартово произведение.

С помощью метода сведения функций нескольких переменных к одноместным [38] тип $(\sigma_1, \ldots, \sigma_n; \tau)$ можно заменить на $\sigma_1\to\sigma_2\to\ldots\to\sigma_n\to\tau$, и поэтому типы (;) не всегда постулируются.

Часть 4

АЛГЕБРЫ И ИСЧИСЛЕНИЯ

Глава 8

РЕЛЯЦИОННАЯ АЛГЕБРА И ИСЧИСЛЕНИЕ

8.1. ОБЩИЕ СВЕДЕНИЯ И ОСНОВНЫЕ ПОНЯТИЯ

Реляционная алгебра и исчисление применяются к построению реляционной базы данных, реляционного банка данных [30, 36, 79]. Однако логические особенности этого аппарата позволили в дальнейшем применять его в моделях баз знаний и баз целей и в естественно-языковых системах общения.

В основе моделей реляционной алгебры (реляционной модели данных) лежит понятие отношений между данными, понятие, которое, как правило, задается в виде таблицы.

Пример 8.1. Рассмотрим отношение

$R_1Г_1$	C_2	A_3
А1—01	503820	СТЕПАНОВ И. В.
А1—01	503821	ЮРГЕНЕВ А. Б.
К1—01	505703	АБРОСИМОВ Ю. И.
К1—01	505706	НОВИКОВ Н. Н.

Такая таблица называется *форматным представлением данных*, или форматом, а набор связанных между собой таблиц — *форматной базой данных*.

Реляционная модель данных представляет собой набор отношений, которые преобразуются, изменяются во времени, при этом образуются новые отношения в зависимости от запроса, содержащего требование сформировать новую таблицу (отношение) (рис. 8.1). На этом рисунке R_1, $R_2\ldots$ обозначают отношения, а $П_1$, $П_2\ldots$ — операции (процедуры) под отношениями. Отношения (таблицы) могут носить семантический и прагматический характер.

В реляционных алгебре и исчислении имеется набор понятий и операций, которые позволяют, во-первых, записывать любое отношение в виде некоторой формулы или формального выражения (α-выражения) и, во-вторых, производить преобразования отношений (α-выражений). Таблица отношений представляет собой одну из разновидностей форматного представления данных, а формальный язык реляционной алгебры называется *форматным языком*. В теоретических исследованиях рассматривают реляционную алгебру и языки алгебры (форматные языки) и реляционное исчисление и языки исчисления (α-языки) (рис. 8.2).

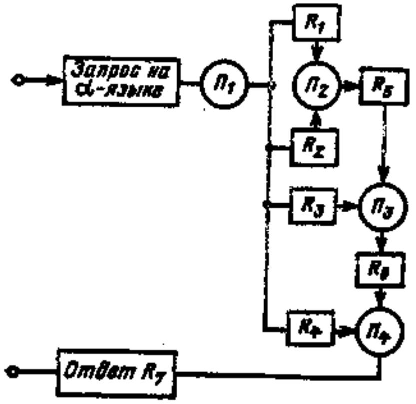


Рис. 8.1. Схема получения ответа на α-запрос в реляционном банке

Рис. 8.2. Алгебры и исчисления реляционной базы данных

Связь реляционного исчисления с реляционной алгеброй осуществляется с помощью процедуры (алгоритма) редукции, которая сводит любое α-выражение к набору стандартных операций реляционной алгебры, и, наоборот, может любое отношение и операции над отношениями записать в виде α-выражения.

Отношение или нормализованное отношение определяется как некоторая таблица (матрица) отношений, подчиняющаяся определенным условиям.

Столбцы таблицы называются *доменами* или *простыми доменами*. Каждый домен должен иметь свое имя или наименование. Например, в примере 8.1 первый домен — это группа (Г), второй домен — студент (С), номер студенческой зачетной книжки, третий домен — фамилия и инициалы студента (А). Домены могут быть пронумерованы. В этом случае говорят, что отношение упорядочено по доменам. Но очень часто на практике отношения имеют настолько большие степени, что в процессе их преобразований удобнее использовать имена доменов (идентифицировать домены по именам) при поиске и хранении информации. Преобразование отношений основано на операциях над доменами: некоторые домены исключаются, некоторые добавляются, меняется их расположение в отношении и т. д.

Пример 8.2. С помощью доменов отношение для примера 8.1 может быть записано в виде $R_1(D_1, D_2, D_3)$, где D_1 — имя Г, D_2 — имя С, D_3 — имя А:

$R_1(D_1$	D_2	$D_3)$
А1—01	503820	СТЕПАНОВ И. В.
А1—01	503821	ЮРГЕНЕВ А. Б.
К1—01	505704	АБРОСИМОВ Ю. И.
К1—01	505706	НОВИКОВ Н. Н.

Очень важное требование к отношениям заключается в том, что строки в отношении должны быть разные. Если в результате операций над отношениями появляются одинаковые строки, то в результирующем отношении автоматически оставляется любая из одинаковых строк. Таким образом, после каждого преобразования отношений полученное новое отношение нормализуется и, в частности, исключаются одинаковые строки.

Как говорилось ранее, часто используется интерпретация отношений как файла. Однако следует иметь в виду, что n-арное отношение (т. е. отношение с n столбцами) обладает следующими свойствами:

каждая строка представляет собой n-выборку из R;

порядок строк не играет роли;

все строки различны (в отличие от файлов);

порядок столбцов важен — он соответствует упорядочению S_1, S_2, ..., S_n доменов, на которых определено отношение R (но иногда рассматривают *доменно-упорядоченные* и *доменно-неупорядоченные отношения*);

каждый столбец отношения помечается именем соответствующего домена.

Пример 8.3. Рассмотрим отношение степени 4, названное *календарным планом* (КП) *студента*:

КП	(Дата	Группа	Дисциплина	Вид отчетности)
	23.10.	К3—01	МАТ. АНАЛИЗ	Контр. работа
	16.12.	К3—01	МАТ. АНАЛИЗ	ЗАЧЕТ
	17.12.	К3—01	ФИЗИКА	ЗАЧЕТ
	08.01.	К3—01	МАТ. АНАЛИЗ	ЭКЗАМЕН
	14.01.	К3—01	ФИЗИКА	ЭКЗАМЕН

Общая *база данных* на логическом уровне может рассматриваться как набор отношений, изменяющихся во времени. Эти отношения имеют самые разные степени. С течением времени в каждое n-арное отношение вставляются дополнительно n-выборки, удаляются имеющиеся выборки, изменяются любые существующие компоненты выборки.

Обычно пользователь, работающий с базой данных, не должен помнить порядок доменов внутри отношения, т. е. он имеет дело не с доменно-упорядоченными структурами, а со связями, в которых участвуют доменно-неупорядоченные компоненты. Поэтому домены должны иметь уникальные имена, по крайней мере внутри одного отношения, что позволяет манипулировать ими без использования позиции. Если встречаются несколько идентичных доменов, то для них должны быть даны различные *ролевые имена*, указывающие на роль, которую играют домены в данном отношении. Обычно один или несколько доменов, рас-

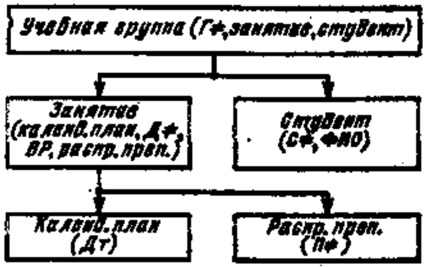


Рис. 8.3. Организация данных об учебном процессе

смотренных в совокупности, имеют значения, однозначно определяющие n-выборки в отношении. Такой домен называется *вероятным* ключом (его можно сравнить с ключом записи). Вероятный ключ называется *неизбыточным*, если никакая его часть не обладает свойством однозначно определять выборку. Когда в отношении имеются два и более неизбыточных вероятных ключа, один из них выбирается произвольно в качестве *основного ключа отношения*. Смысл выделения основных ключей состоит в том, чтобы обеспечить возможность ссылок на выборки того же самого или другого отношения. Назовем домены в отношении R *внешним ключом*, если он не является основным ключом R, но его элементы есть значения основного ключа некоторого отношения S (случай идентичности R и S не исключается).

До сих пор рассматривались примеры отношений, определенных на *простых доменах*, т. е. таких доменах, которые стоят из неделимых элементов. Однако требование неделимости необязательно, и некоторые домены в качестве своих элементов могут иметь отношения. Эти отношения могут, в свою очередь, определяться на непростых доменах и т. д. Когда при дальнейшем рассмотрении баз данных будет определяться термин *атрибут*, то это есть налог простого домена; термин *повторяющаяся группа* является аналогом непростого домена.

Как говорилось ранее, отношение с простыми доменами может быть представлено как файл с однородными записями. Более сложная структура данных необходима для представления отношения с непростыми доменами. Поэтому непростые домены ликвидируются с помощью простой процедуры, называемой *нормализацией*. В качестве примера рассмотрим организацию данных об учебном процессе в вузе (рис. 8.3). Из рисунка видно, что ЗАНЯТИЕ и СТУДЕНТ — непростые домены отношения УЧЕБНАЯ ГРУППА, КАЛЕНДАРНЫЙ ПЛАН и РАСПРЕДЕЛЕНИЕ ПРЕПОДАВАТЕЛЕЙ — непростые домены отношения ЗАНЯТИЕ.

Нормализация проводится следующим образом. Начиная с отношения в корне дерева, возьмем основной ключ Г и включим его во все подчиненные отношения, в результате чего получим их расширения. Основной ключ расширенного отношения содержит свой собственный основной ключ, бывший до расширения, а также основной ключ из вышестоящего отношения. Исключив теперь из старшего отношения все непростые домены, удалим вершину дерева и применим всю процедуру для каждого из оставшихся поддеревьев.

Получим нормализованный набор отношений следующего вида:

СТУДЕНТ (Г≠, С≠, ФИО);
КАЛЕНД. ПЛАН (Г≠, Дт, Д≠, ВР≠);
РАСПР. ПРЕП. (Г≠, Д≠, ВР≠, П≠),

где

Г≠ — номер учебной группы;
С≠ — номер зачетной книжки студента;
ФИО — фамилия, имя, отчество студента;
Дт — номер учебной недели;
Д≠ — наименование изучаемой дисциплины;
ВР≠ — вид работы (отчета);
П≠ — фамилия преподавателя.

Процедура нормализации применима, если выполняются следующие условия:

1) граф связей непростых доменов имеет вид дерева;

2) основной ключ не содержит в качестве своего компонента непростой домен.

Дальнейший анализ процедуры нормализации дает возможность указать три нормальные формы отношений [30, 36, 80]. В самой простой, третьей нормальной форме отношения свободны от транзитивных зависимостей доменов внутри данного отношения. Нормализованные отношения имеют простую интерпретацию в виде файлов: отношения с непростыми доменами (см. рис. 8.3) можно рассматривать как логическую иерархическую базу, а приведенный алгоритм нормализации есть не что иное, как перечисление всех путей, идущих из корня дерева в его висячие вершины.

Если пользователь в своей реляционной модели применяет нормальную форму, то это упрощает указание имен данных. Самое общее имя имеет вид

$$R(g).r.d,$$

где R — имя отношения; g — идентификатор поколения; r — ролевое имя; d — имя домена. Наиболе часто используется запись $R \cdot d$ (g становится необходимым, если существуют несколько поколений данного отношения; r необходимо, если в отношении R два домена или более с именем d).

8.2. РЕЛЯЦИОННАЯ АЛГЕБРА

Так же как в любой алгебре, в реляционной алгебре используется понятия множества. *Простой домен* определяется как множество, все элементы которого являются цепочками символов. Отношение, определенное на простых доменах, называется *простым нормальным отношением* (для краткости, если не делается специальных оговорок, только такие отношения и будем рассматривать, применяя к ним просто термин *отношения*). В реляционной алгебре вводится понятие *расширенного декартового произведения*, которое в дальнейшем будет называться просто *произведением*.

Декартово произведение двух множеств C и D формально определяется как

$$C \times D = \{(c,\ d): c \in C \wedge d \in D\}.$$

Расширенное декартово произведение χ n-множеств D_1, $D_2, \ldots, D_n$ формально определяется выражением

$$\chi(D_1,\ D_2,\ .\ .\ .\ ,\ D_n) = \{(d_1,\ d_2,\ .\ .\ .\ ,\ d_n);\ d_j \in D_j,\ j=1,\ .\ .\ .\ ,\ n\}.$$

Элементы этого множества называются n-выборками или просто выборками; формально отношение на множествах D_1, D_2, ..., D_n определяется как подмножество множества $\chi(D_1$, D_2, ..., $D_n)$ и является особым видом множества. Все его элементы являются выборками-строками, содержащими n элементов, где n — степень отношения (первый элемент отношения d_1 берется из множества D_1, второй — из множества D_2 и т. д.). Таким образом, отношения — это множества, элементы которых представляют собой *кортежи* (упорядоченные цепочки символов). Множества D_1, D_2, ..., D_n называются *доменами отношения*. Каждое вхождение домена в декартово произведение называется *атрибутом*. Заметим, что декартово произведение может иметь несколько вхождений одного и того же домена. Так, для создания реляционной модели СЛУЖАЩИЙ может дважды понадобиться домен ФАМИЛИЯ. Первый раз — для фамилии служащего, второй — для фамилии начальника отдела, где работает служащий.

Тип отношения задается строкой $R(A_1,\ A_2,\ \ldots,\ A_n)$ или *схемой отношения*, где R — имя отношения, n — степень отношения, A_i — имя i-го атрибута. Каждая строка — элемент отношения и множества, которым определяется это отношение. Порядок строк и столбцов в табличном представлении отношения произвольный. Число атрибутов таблицы-отношения называется *степенью отношения*. Отдельный атрибут может рассматриваться как отношение, имеющее один столбец или одну ступень. Если атрибутам разрешается быть неупорядоченными в отношении (т. е. если они обладают некоторой свободой), то строки в отношении считаются упорядоченными.

Если $d = (d_1,\ d_2,\ \ldots,\ d_m)$ и $e = (e_1,\ e_2,\ \ldots,\ e_n)$, то *конкатенацией* d с e является выборка

$$\widehat{de} = (d_1,\ d_2,\ .\ .\ .\ ,\ d_m,\ e_1,\ e_2,\ .\ .\ .\ ,\ e_n).$$

Операции реляционной алгебры. Экземпляр отношения представляет собой множество, элементами которого являются кортежи или строки. Поэтому к ним применимы теоретико-множественные операции: объединение, пересечение и разность. Необходимо только, чтобы отношения были совместимы по объединению, т. е. имели одинаковую степень. Рассмотрим эти операции на примерах.

Пример 8.4. Пусть имеются два экземпляра отношений:

$R\ (A,\ B)$;	$S\ (C,\ D,)$
a_1b_1	c_1d_1
a_2b_2	a_1b_1

Тогда соответствующие теоретико-множественные операции будут иметь следующий вид:

объединение: $R \cup S(M,\ N)$

a_1b_1

a_2b_2

c_1d_1

пересечение: $R \cap S(M,\ N)$

$a_1,\ b_1$

разность: $R/S = R - S(M,\ N)$

$a_2,\ b_2$

В соответствии с операциями над множествами разность двух множеств определяется как дополнение на множестве, состоящем из объединения уменьшаемого и вычитаемого, подмножества K, состоящего из общих элементов уменьшаемого R и вычитаемого S (их пересечения), и элементов множества вычитаемого S, не содержащихся в уменьшаемом. Поэтому элементы вычитаемого, отличные от элементов уменьшаемого, в разность не входят.

Кроме того, в определении этих трех операций проявляется непоследовательность изложения реляционной алгебры Кодда, так как за элементы множества-отношения принимаются домены, а не строки.

Операция произведения в реляционной алгебре вводится оригинальным способом, отличным от общепринятого. Если обычное декартово произведение $R \times S$ отношений R (степени m) и S (степени n) имеет степень 2, то расширенное декартово произведение $R \otimes S$ имеет степень $m+n$ и определяется формулой

$$R \oplus S = \left\{\left(\widehat{rs}\right): r \in R \wedge s \in S\right\}.$$

Рассмотрим пример.

Пример 8.5. Пусть заданы отношения R степени $m=2$ и S степени $n=1$.

$R\ (A,\ B)$;		$S\ (c)$.
a_1b_1	$(=y_1)$	c_1.
a_2b_2	$(=y_2)$	c_2.

Тогда расширенное декартово произведение будет равно:

$$\begin{array}{lll} R\otimes S\,(A, & B, & C). \\ a_1 & b_1 & c_1 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_1 \\ a_2 & b_2 & c_2. \end{array}$$

Как видно из записи, это отношение имеет степень $m+n$ и совпадает с обычным декартовым произведением, если за элементы множеств брать строки. Так, если в отношении R строки обозначать через y_1 и y_2, то обычное декартово произведение будет иметь вид

$$\begin{array}{ll} y_1 & c_1 \\ y_1 & c_2 \\ y_2 & c_1 \\ y_2 & c_2. \end{array}$$

Если в это отношение подставить выражения для y_1 и y_2, то получится расширенное декартово произведение $R\otimes S$ как отношение степени $m+n=3$. Таким образом, декартово произведение существенно зависит от того, что считать элементом множества. Так, если каждый элемент отношения считать элементом множества, то получится следующее декартово произведение:

$$\begin{array}{lll} R\times S & (M, & N), \\ & a_1 & c_1 \\ & a_1 & c_2 \\ & a_2 & c_1 \\ & a_2 & c_2 \\ & b_1 & c_1 \\ & b_1 & c_2 \\ & b_2 & c_1 \\ & b_2 & c_2, \end{array}$$

которое имеет степень 2.

В определении расширенного произведения по Кодду как бы сохраняется структура каждого из сомножителей-отношений.

Операция проекции предназначена для изменения числа столбцов в отношении. Если требуется исключить из строк-кортежей какие-нибудь атрибуты, например номера зачетных книжек С в примере 8.1, то применяется операция проектирования на атрибуты (Г, А).

Допустим, что r — выборка (строка-кортеж) из m-арного отношения R. Обозначим через $r(j)$ j-й компонент строки, где $j=1, 2, \ldots, m$. Для остальных значений $jr(j)$ не определено.

Введем в рассмотрение список $A=(j_1, j_2, \ldots, j_k)$ целых чисел из $1, 2, \ldots, m$ и набор элементов строки

$$r[A]=(r(j_1],\quad r[j_2],\ .\ .\ .\ ,\quad r[j_k]).$$

Если A — пустой список, то $r[A]=r$.

Пусть R — отношение степени m, а A — список целых чисел (но необязательно различных) из множества $\{1, 2, \ldots, m\}$. Тогда проекция R на A определяется формулой

$$R[A]=\{r[A]:r\in R\}.$$

Заметим, что так как в списке A индексы могут повторяться, то атрибуты можно исключать, добавлять, менять местами. Если A есть перестановка из списка $(1, 2, \ldots, m)$, то $R[A]$ является отношением, которое состоит из тех же доменов (атрибутов), но в другом порядке.

Пример 8.6. Пусть дано отношение третьей степени ($m=3$):

$$\begin{array}{lll} R\,(D_1, & D_2, & D_3). \\ a & 2 & f \\ b & 1 & g \\ c & 3 & f \\ d & 3 & g \\ e & 2 & f \end{array}$$

Пять его проекций запишутся в следующем виде:

$$\begin{array}{ccc} R[1]\,M_1; & R[2]\,M_1; & R[3]\,M_1; \\ a & 2 & f \\ b & 1 & g \\ c & 3 & \\ d & & \\ e & & \end{array}$$

$$\begin{array}{ll} R[3,\ 2]\,(M_1 & M_2); \\ f & 2 \\ g & 1 \\ f & 3 \\ g & 3 \end{array} \qquad \begin{array}{lll} R[3,\ 2,\ 2]\,(M_1 & M_2 & M_3). \\ f & 2 & 2 \\ g & 1 & 1 \\ f & 3 & 3 \\ y & 3 & 3 \end{array}$$

В этих отношениях-проекциях используется условие нормализации отношения, согласно которому в отношении запрещается иметь одинаковые строки.

Операции соединения. По определению θ-соединение отношения R по домену A с отношением S по домену B задается формулой

$$R[A\theta B]S=\left\{\left(\widehat{rs}\right):r\in R\wedge s\in S\wedge(r[A]\,\theta s[B])\right\}$$

при условии, что каждый элемент $R[A]$ является θ-соединением, сравнимым с каждым элементом $S[B]$. Символ обозначает любое из соотношений $=, \neq, <, \leqslant, >, \geqslant$.

Считается, что $x\ \theta$ сравним с y, если $x\theta y$ имеет значение «истина» или «ложь» (но не является неопределенным).

Пример 8.7. Пусть даны два отношения

$R\,(A,\ B,\ C_1);\quad S\,(D,\ E).$

a	2	1
b	1	2
c	2	5
c	5	3

2	*f*
3	*g*
4	*f*

Четыре возможных θ-соединения этих двух отношений будут иметь вид

$R\,[C = D]\ S\,(A,\ B,\ C,\ D,\ E);$

b	1	2	2	*f*
c	5	3	3	*g*

$R\,[C < D]\ S\,(A,\ B,\ C,\ D,\ E);$

a	2	1	2	*f*
a	2	1	3	*y*
a	2	1	4	*f*
b	1	2	3	*y*
b	1	2	4	*f*
c	5	3	4	*f*

$R\,[B > C]\ (A,\ B,\ C);$

a	2	1
c	5	3

$R\,[B \neq C]\ (A,\ B,\ C).$

a	2	1
b	1	2
c	2	5
c	5	3

Очевидно, что операция соединения позволяет объединять строки различных отношений по критерию сравнения элементов каких-нибудь двух доменов (например, в списке кадров выделить список всех сотрудников, родившихся после 1946 г.).

Операция деления. Если проекции $R[A]$ и $S[B]$ объединимы (т. е. имеется одинаковое число доменов), то деление R по A на S по B определяется соотношением

$$R\,[\div B]\,S = \{r\,[\overline{A}] : r \in R \wedge S\,(B) \subseteq g_R\,(r\,[\overline{A}])\}, \qquad (8.1)$$

где T — бинарное отношение, которое определяет элементам x множество образов y в соответствии с соотношением

$$g_T\,(x) = \{y : (x,\ y) \in T\};$$

A — список, определяющий домены (без дублирования) для отношения R; $\overline{A}$ — дополнительный к A список, т. е. если степень отношения $Rm = 5$ и $A = (2, 5)$, то $\overline{A} = (1, 3, 4)$.

В соотношении (8.1) делимое рассматривается как бинарное отношение доменов A и A. При заданной выборке $r \in R$ для прообраза $r[\overline{A}]$ можно определить бинарные образы $g_R(r[\overline{A}])$, являющиеся подмножеством $R[A]$. Соотношение (8.1) означает, что операция деления — это поиск такого прообраза $r[\overline{A}]$, определяемого доменами дополнительного списка, для которого множество образов делимого из подмножества $R[A]$ показывает множество делителя подмножества $S[B]$.

Заметим, что, как будет видно из дальнейшего, операция деления представляет собой алгебраическую интерпретацию квантора общности. Кроме того, *операцию деления* можно определить с помощью уже ранее введенных операций следующим образом:

$$R\,[C \div D]\,S = R\,[\overline{C}] - ((R\,[\overline{C}] \otimes S\,[D]) - R)\,[\overline{C}].$$

Положение о том, что ряд операций реляционной алгебры можно выразить через другие операции, вносит некоторую неопределенность в количество операций реляционной алгебры.

Операция ограничения. Пусть даны отношение R и два списка B и A, которые определяют домены для отношения R. Символ θ определяет одно из следующих отношений: $=, \neq, <, \leqslant, >, \geqslant$.

Тогда θ-ограничение по спискам A и B определяется как

$$R\,[A \theta B] = \{r : r \in R \wedge (r\,[A]\,\theta_r\,[B])\}$$

при условии, что каждый элемент из отношения $R[A]$ является θ-сравнимым с каждым элементом из отношения $R[B]$.

Пример 8.8. Пусть отношение R имеет вид

$R\,(A,\ B,\ C);\quad R\,[B = C]\,(A,\ B,\ C);$

p	2	1
q	2	3
q	5	4
r	3	3

r	3	3

$R\,[B > C]\,(A,\ B,\ C)$

p	2	1
q	5	4

С помощью ранее рассмотренных операций операцию ограничения можно определить следующим образом:

$$Y[A\theta B] = \left(R\left[\widehat{AB} = \widehat{AB}\right](R[A]\ [A\theta B]\ R[B])\right)_{[L]}, \qquad (8.2)$$

где L — список, определяющий все домены из отношения R в порядке их возрастания; $\widehat{AB}$ — конкатенация списка A со списком B.

Аналогично можно определить θ-соединение R с S через произведение и θ-ограничение:

$$R[A\theta B]\,S = (R \otimes S)[A\theta B]. \qquad (8.3)$$

Кроме того, очевидно, что

$$(R \otimes S) \div S = R.$$

8.3. РЕЛЯЦИОННОЕ ИСЧИСЛЕНИЕ

Как уже указывалось ранее, в реляционном исчислении путем использования ограниченного исчисления предикатов, кванторов и переменных выборки удается описать отношения и операции над ними в виде аналитического выражения или формулы.

В реляционном исчислении используются алфавиты (или нотации, т. е. совокупность обозначений), термы (или элементарные конструкции) и формулы, представляющие собой некоторые аксиомы, записанные с помощью предикатов, термов и обозначений в виде аналитических выражений.

В реляционном исчислении вводится следующий алфавит A:
отдельные константы: $a_1, a_2, a_3\ldots$;
индексы: 1, 2, 3...;
переменные выборки: $r_1, r_2, r_3\ldots$;
предикатные константы
  монадические: $P_1, P_2, P_3\ldots$;
  диадические: $=, <, >, \leqslant, \geqslant, \neq$;
логические символы: $\exists, \forall, \wedge, \vee, \neg$;
разделители: [ ] ( ).

В реляционном исчислении вводится взаимное однозначное соответствие между монадическими предикатами и отношениями в конкретной базе данных. Так, если даны отношения $R_1, R_2, \ldots, R_N$, то предикат P_j показывает принадлежность выборок к отношению R_j ($j = 1, 2, \ldots, N$).

В реляционном исчислении вводятся следующие термы.

Терм диапазона (ТД). Монадический предикат, за которым следует переменная выборки, называется термом диапазона. Терм диапазона $P_j r$ понимается как некоторое утверждение, заключающееся в том, что переменная выборки r имеет в качестве своего диапазона отношение. Для определения N-го компонента выборки r вводится понятие операции *индексной выборки*, которая имеет форму $r[N]$, где r — переменная выборки, а N — индексная константа.

Терм объединения (ТО). Пусть λ и μ — индексные выборки и α — отдельная константа. Тогда λθμ и λθα называются термами объединения, где θ — операция объединения. Терм объединения может состоять из следующих монадических переменных:

$$\begin{array}{ll}
v[a] = \alpha; & v[a] \neq \alpha; \\
v[a] < \alpha; & v[a] \leqslant \alpha; \\
v[a] > \alpha; & v[a] \geqslant \alpha, \\
v_1[a] = v_2[b]; & v_1[a] \neq v_2[b]; \\
v_1[a] < v_2[b]; & v_1[a] \leqslant v_2[b]; \\
v_1[a] > v_2[b]; & v_1[a] \geqslant v_2[b].
\end{array}$$

Эти два терма являются единственными в реляционном исчислении.

Из термов с помощью логических символов и разделителей строятся формулы реляционного исчисления.

Правильно построенная формула (*ППФ*) реляционного исчисления рекурсивно определяется следующим образом:

1) любой ТД или ТО есть ППФ;

2) если Г — ППФ, то и $\neg$Г — тоже ППФ, где $\neg$ — символ логического отрицания;

3) если Γ_1 — ППФ и Γ_2 — ППФ, то и $(\Gamma_1 \wedge \Gamma_2)$, и $(\Gamma_1 \vee \Gamma_2)$ — ППФ;

4) если Г — ППФ, в которой r используется как свободная переменная, то $\exists r(\Gamma)$ и $\forall r(\Gamma)$ — ППФ;

5) никакие другие формулы не являются ППФ.

Пример 8.9. Приведем ППФ без ТД и укажем в ней свободные переменные, т. е. такие переменные, которые не связаны кванторами.

ППФ без ТД	Свободные переменные
$r_1[3] > a_1$	r_1
$\exists r_1 [r_1[3] > a_1]$	Нет
$\exists r_1 (r_1[3] = r_2[5])$	r_2
$\forall r_1 \exists r_2((r_1[3] = r_2[5]) \vee r_1[1] = a_1)$	Нет
$(r_1[3] = r_2[5]) \wedge \exists r_3 (r_3[1] = a_2)$	r_1, r_2

Пример 8.10. Приведем ППФ только с ТД

ППФ с ТД	Свободные переменные
$P_6 r_3$	r_3
$P_6 r_3 \wedge P_7 r_2$	$r_2 r_3$
$P_6 r_3 \vee P_8 r_3$	r_3

Диапазонная сепарабельность. Для определения диапазона переменной выборки как ППФ вводится ряд определений.

Диапазонной ППФ является свободная от кванторов ППФ, все термы которой являются ТД. *Диапазонная ППФ* по r — это диапазонная ППФ, единственная свободная в которой r. *Соответственная диапазонная ППФ* по r — это диапазонная ППФ по r, которая удовлетворяет двум ограничениям: синтаксическому, заключающемуся в том, что символ $\neg$ или совсем не встречается, или за ним непосредственно следует символ $\wedge$, и семантическому, которое заключается в том, что везде, где r встречается в двух и более ТД, предикаты диапазона должны быть связаны с отношениями, которые объединены. Оба ограничения защищают переменные выборки от появления диапазонов, которые недопустимы для данных отношений, а также защищают отношения, которые могут быть сгенерированы из них с помощью объединения, пересечения и вычитания.

Как связанные, так и свободные переменные должны иметь четко определенные диапазоны. Предположим, что Δ — ППФ, имеющая свободную переменную r, но не содержащая никаких ТД в r.

Пусть Γ — соответствующая диапазонная ППФ по r. Чтобы ввести Γ в выражение $\exists r(\Delta)$ или $\forall r(\Delta)$, $\exists r$ и $\forall r$ заменяются на $\exists \Gamma r$ и $\forall \Gamma r$ соответственно, при этом получим выражения

$$\exists \Gamma(\Delta) = \exists r(\Gamma \wedge \Delta);$$

$$\forall \Gamma(\Delta) = \forall r(\neg \Gamma \vee \Delta),$$

которые называются *диапазонно-связанными кванторами*.

Правильно построенная формула, имеющая четко определенные диапазоны для всех своих переменных, называется *диапазонно-сепарабельной* (ДС) и определяется как конъюнкция из форм

$$u_1 \wedge u_2 \wedge \ . \ . \ . \ \wedge u_n \wedge v,$$

где $n \geqslant 1$; $u_1, u_2, \ldots, u_n$ — соответствующие диапазонные ППФ по n различным переменным выборки; v или пусто, или оно ППФ с тремя свойствами; каждый квантор в v диапазонно связан; каждая свободная переменная в v принадлежит множеству, диапазоны которого определены в $u_1, u_2, \ldots, u_n$; v не содержит ТД.

Из этих требований следует, что ДС ППФ имеет, по крайней мере, одну свободную переменную.

Пример 8.11. Приведем различные ППФ.

ДС ППФ	Свободные переменные
$P_6 r_1 \wedge (r_1[3] = a_1)$	r_1
$P_7 r_2 \wedge \exists P_8 r_1 (r_1[3] = r_2[5])$	r_2

Не ДС ППФ	Свободные переменные
$P_8 r_1 \vee (r_1[3] = a_1)$	r_1
$P_7 r_2 \wedge \exists r_1 (r_1[3] = r_2[5] \vee P_8 r_1)$	r_2
$\neg P_8 r_1 \vee (r_1[3] = a_1)$	r_3

Альфа-выражение. Запрос в реляционном исчислении осуществляется на α-языке или в виде α-выражения. Общая структура α-выражения z имеет вид

z-целевой список/квалифицирующее выражение, или $z = t/w$.

Наклонная черта иногда заменяется двоеточием или словом *где*. В целевом списке указываются домены, образующие результирующее отношение (которое хотим получить). Квалифицирующее выражение определяет логические условия подбора данных из объявленных отношений, причем разрешается употреблять квантор существования $\exists$ и квантор общности $\forall$. Записанное α-выражение подвергается редукции или разложению (декомпозиции) в последовательность алгебраических операций над отношениями базы данных. Как правило, редукция осуществляется путем формирования расширенного декартова произведения из объявленных отношений, перевода на язык алгебры квалифицирующего выражения и извлечения окончательной проекции, определенной целевым списком.

Пример 8.12. Приведем α-запрос в случае, когда даны отношения[1]

R_1 (Г§, С§, А);
R_2 (Г§, Д§, Вр§, Фр§, П);
R_3 (Дг§, Г§, Д§, Вр);
R_4 (С§, Д§, Вр§, Дт§О)

---

[1] В данном примере для обозначения ключа использован символ § вместо ≠.

с именами *студент* (СТ), *преподаватель* (ПР), *план* (Пл) и *оценка* (Оц) соответственно. Зададим следующий α-запрос: «Найти список» (Г§, С§, Д§, Вр§, А), где Дт§ 760116ЛПЛ. Г§-ПР.Г§ЛПЛ.Д§-Пр.Д§ЛПЛ. Вр-Пр. Вр§ЛПл Г§-СТ.Г§). Этот запрос означает «составить список или отношение, содержащее домены, указанные в целом списке с элементами, удовлетворяющими условиям квалифицирующего выражения, т. е. необходимо составить в домене ДТ§ (в отношении Пл) только те элементы, для которых удовлетворяется условие «<760116» (например, зачетной книжки меньше 760116), и т. д.

Нетрудно видеть, что квалифицирующее выражение имеет вид конъюнкций диапазонных ППФ, состоящих из термов диапазона и объединения, точнее, из диадических термов.

Более строго простое α-выражение определяется как выражение вида

$$z=(t_1,\ t_2,\ \ldots,\ t_k):w\equiv t:w,$$

где t_1, t_2, ..., t_k — различные термы, каждый из которых состоит из переменных выборки или индексных переменных; w — диапазонно-сепарабельная ППФ.

Множеством переменных выборки, встречающихся в t_1, t_2, ..., t_k, является в точности множество свободных переменных в w. Пусть ρ_1, ρ_2, ..., ρ_n — различные переменные выборки в порядке их первого вхождения в целевой список α-выражения z, отношения S_1, S_2, ..., S_n (не обязательно различные) имеют диапазоны из ρ_1, ρ_2, .., ρ_n соответственно. Тогда z означает определенную проекцию из такого подмножества декартова произведения $S_1 \otimes S_2 \otimes \ldots \otimes S_n$, элементы которого удовлетворяют квалифицирующему выражению. Проекция определяется через индексы, связанные с переменными выборки в целевом списке.

8.4. СВЯЗЬ РЕЛЯЦИОННОЙ АЛГЕБРЫ И РЕЛЯЦИОННОГО ИСЧИСЛЕНИЯ. АЛГОРИТМ РЕДУКЦИИ

Под редукцией понимается некоторая универсальная процедура или алгоритм для перевода любого простого α-выражения z в семантически эквивалентное алгебраическое выражение. Эта процедура представляет интерес, по крайней мере, по двум соображениям. Во-первых, редукция необходима для реализации всей идеологии реляционного исчисления и алгебры, т. е. реляционных банков на ЭВМ. Общение (запросы) с реляционным банком идет на α-языке (рис. 8.4). Форма запроса может быть самая разнообразная. По-видимому, единственный разумный путь — перевести α-выражение в соответствующее выражение реляционной алгебры. Если имеются реализация всех операций алгебры на ЭВМ и соответствующие отношения в памяти, то с помощью алгоритма редукции можно получить нужное отношение как ответ на запрос (рис. 8.4). Следует обратить внимание на то обстоятельство, что в данном случае система, изображенная на рис. 8.4, представляет собой вопросно-ответную систему (ВОС), на входе которой вопрос задается на α-языке, а ответ получается в виде отношения (которое также можно перевести в α-выражение, но этого делать, как правило, не следует). Проблема интеллектуальности такой ВОС (т. е. является ли она ИВОС) требует специального рассмотрения. Во-вторых, с теоретической точки зрения универсальный алгоритм редукции является некоторым доказательством реляционной полноты пары: исчисления и алгебры, которая формулируется следующим образом. Алгебра и исчисление считаются *реляционно полными*, если для любого набора отношений R_1, R_2, ..., R_N (в простой нормальной форме, т. е. состоящих из простых доменов) выражения алгебры или исчисления позволяют определить любое отношение, определяемое из R_1, R_2, ..., R_N с помощью α-выражений.

Рис. 8.4. Общая схема прохождения запроса в реляционной базе данных

Совокупность α-выражений как подмножество слов над алфавитом A реляционного исчисления называется языком запросов к базе данных (α-язык), схема которого описывается схемами связей отношений R_1, R_2, ..., R_N. Говорят еще, что ответом на запрос служит его интерпретация в алгебру, порожденную состояниями связей в некоторый момент времени. Алгоритм редукции означает интерпретацию произвольного α-выражения при заданной интерпретации входящих в это выражение предикатов.

Прежде всего, алгоритм редукции основан на соответствии между отношением R_i из алгебры и диапазонным предикатом P_i, операциями над отношениями (множествами) и логическими операциями исчисления и термами из исчисления, которое представлено на рис. 8.5.

Разработан итеративный алгоритм редукции [36], который вместо непосредственного получения нескольких алгебраических выражений, скомбинированных (собранных) в одно составное выражение T и представляющих собой алгебраическую интерпретацию исходного α-запроса, предлагает вычислять последовательно одно за другим составные выражения (T_n, T_{n-1}, ..., T_1, T). В соответствии с этим алгоритмом (рис. 8.6):

1) сначала получаются диапазоны существующих в α-выражении переменных выборки путем просмотра указанных отношений из базы данных и путем объединений, пересечений и теоретико-множественных вычитаний по необходимости (в соответствии с рис. 8.6);

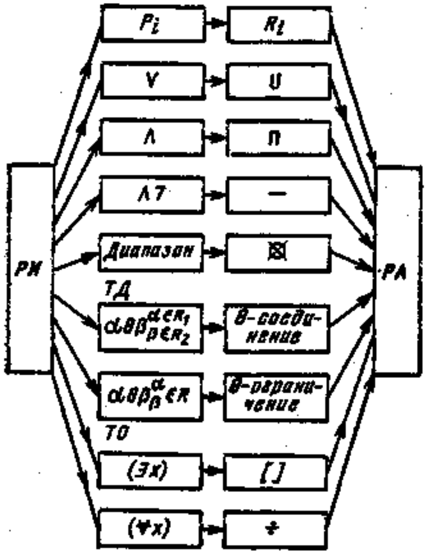


Рис. 8.5. Соответствие операций реляционного исчисления и реляционной алгебры

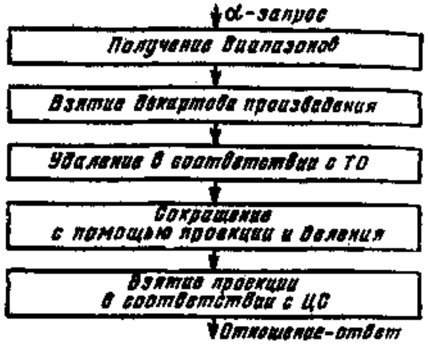


Рис. 8.6. Этапы выполнения алгоритма редукции

2) берется декартово произведение этих диапазонов (из этого произведения будет извлекаться окончательное отношение);

3) выборки, которые не удовлетворяют комбинации термов объединения, удаляются из произведения;

4) оставшиеся произведения сокращаются с помощью операции проекции и деления с тем, чтобы удовлетворить требованиям, содержащимся в величине V в α-выражении;

5) выполняется проекция в соответствии с целевым списком и тем самым определяется требуемое отношение.

Прежде чем формализовать эти этапы, целесообразно преобразовать α-выражение к некоторому каноническому виду и ввести некоторые обозначения.

Допустим, имеется простое α-выражение вида

$$(t_1, \ldots, t_k): U_1 \wedge U_2 \wedge \ldots \wedge U_p,$$

где U_i, $i=1, 2, \ldots$; p — ППФ, определенная под индексной переменной выборки r_i (причем $r_i \neq r_j$ при $i \neq j$) или строковой переменной.

В частности, для α-выражения, не содержащего кванторы, т. е. без формулы V, достаточно реализации 1-го, 2-го и 5-го этапов.

Пример 8.13. Рассмотрим α-выражение вида.

$$(t_1, \ldots, t_k): U_1 \wedge U_2 \wedge \ldots \wedge U_p \vee U.$$

Предположим, что выполнены следующие условия:

1) все строковые переменные занумерованы в порядке их первого появления после двоеточия;

2) в формуле все кванторы вынесены влево, а бескванторная часть преобразована в дизъюнктивную нормальную форму;

3) символы отрицания ┐ исключены путем замены отрицаний ТО

$$\neg\lambda=\mu,\quad \neg\lambda\neq\mu,\quad \neg\lambda<\mu,\quad \neg\lambda\leqslant\mu,\quad \neg\lambda>\mu,\quad \neg\lambda\geqslant\mu$$

на соответствующие ТО без отрицания

$$\lambda\neq\mu,\ \lambda=\mu,\ \lambda\geqslant\mu,\ \lambda\leqslant\mu,\ \lambda<\mu,$$

где $\lambda=V(a)$ — индексная переменная (V — переменная выборки, a — индексная константа), а μ может быть или константой, или индексной переменной $v_t=v_1$ $[b]$.

Очевидно, что эти ограничения не уменьшают общности, так как им всегда можно удовлетворить, они облегчают формализацию алгоритма редукции, который приводится далее.

Алгоритм продукции. Шаг 1. Заменить U_i связью C_i по правилам, приведенным на рис 8.5, где $i=1, \ldots, p+g$, т. е.

вместо	подставить
а) $P_i\ r_i$	R_i
б) ∪	∨
в) ∩	∧
г) —	∧ ┐ (при условии, когда п. в неприменим).

Шаг 2. Введем обозначение для объединения $A(R_i)$ данного отношения R_i со всеми отношениями R_i базы данных, которые с ним объединены. Обозначим через S_i любое подмножество из $A(R_i)$, причем будем считать по определению, что

$$A(S_i) = A(R_i).$$

Введем переменные x_i следующим образом:

$$x_i = \begin{cases} S_i, & \text{если } n \text{ свободна или связана квантором существования;} \\ A(S_j), & \text{в других случаях (т. е. если переменная связана квантором общности).} \end{cases}$$

Итогом данного шага является формирование произведения

$$S = x_1 \otimes x_2 \otimes \ldots \otimes x_{p+q}.$$

Пусть степень отношения x_j есть n_j $j=1, 2, \ldots, p+q$. Положим

$$\mu_l = \sum_{i=1}^{j-1} n_i.$$

Тогда домен с позиций j в отношении x_j будет иметь позицию $j+\mu_j$ в декартовом произведении S.

Шаг 3. Удалить из формулы V все кванторы и связанные ими переменные. Если оставшаяся формула V' пуста, то положить $T_{p+q}=S$. Если оставшаяся бескванторная формула V' непуста, то для получения уравнения, определяющего T_{p+q}, следует выполнить в ней следующие замены:

1) $\vee \rightarrow \cup$;
2) $\wedge \rightarrow \cap$;
3) $r_j\,[j]\,\theta r_k\,[K] \rightarrow S\,[(j+\mu_j)\,\theta\,(K+\mu_k)]\,S$;
4) $(r_i\,[j]\,\theta a] \rightarrow S\,[(j+\mu_i)\,\theta 1]\,(a)$.

Правая часть уравнения $T_{p+q}=S$ получается путем соответствующих замен в формуле V' согласно п. 1—4.

Шаг 4. Положим $j=p+q, \ldots, p+1$ и будем последовательно вычислять T_p на основании следующей формулы:

$$T_{j-1}\begin{cases} T_j\,(1,\ 2,\ \ldots,\ \mu_l) \text{ — если переменная выборки связана квантором существования;} \\ T_j\,[E_{j+1} \div D_j]\,S_j \text{ — если переменная выборки связана квантором общности,} \end{cases}$$

где

$$E_j=(\mu_1+1,\ \mu_j+2,\ \ldots,\ \mu_j+n_j);\quad D_j=(1,\ 2,\ \ldots,\ n_j).$$

Полученное уравнение для T_{j-1} отражает действие кванторов (существования или общности) из формул V, начиная с внутреннего Q_{p+q} и кончая самым внешним Q_{p+1}, где

$$Q_i=\begin{cases} \text{или квантор существования } \exists; \\ \text{или квантор общности } \forall. \end{cases}$$

Шаг 5. Составить итоговое уравнение, которое учитывает проекцию, содержащуюся в целом списке:

$$T=T_p\,[C_1{}^\frown C_2{}^\frown \ldots {}^\frown C_k],$$

где

$$C_k=\begin{cases} (\mu_j+1,\ \mu_j+2,\ \ldots,\ \mu_j+n_j=\mu_j+1, \text{ если } t_k=r_j; \\ (\mu_j+j), \text{ если } t_k=r_j\,[j]. \end{cases}$$

Далее путем простой подстановки определяем T непосредственно в терминах $R_1, R_2, \ldots, R_N$ с их степенями

$$n_1,\ n_2,\ \ldots,\ n_N.$$

Пример 8.14. Пусть БД задается следующими отношениями и доменами:

Символ	Имя отношения	Степень	Диапазон предекатов
R_1	Поставщики	3	P_1
R_2	Проекты	2	P_2
R_3	Поставка	3	P_3

Домен 1	Домен 2	Домен 3
Поставщик	Имя	Местоположение
Проект	Имя	
Поставщик	Часть	Проект

Положим $A(R_i)=R_i$ ($i=1,\ 2,\ 3$), т. е. предполагаем, что отношения несовместимы по объединению.

Сформулируем запрос: найти имя и местоположение всех поставщиков, каждый из которых снабжает все проекты. Соответствующий α-запрос будет иметь вид

$$(r_1\,[2]\,r_2\,[3] : P_1 r_1 \wedge \forall P_2 r_2 \exists P_3 r_3\,(r_1\,[1]=r_3\,[1]) \wedge (r_3\,[3]=r_2\,[1]).$$

Применяя алгоритм редукции, получаем

$$S_i=R_i\,(i=1,\ 2,\ 3);$$

$$S=S_1 \otimes S_2 \otimes S_3;$$

$$T_3=S\,[1=6] \cap S\,[b=4] \quad \text{(соответствует ППФ);}$$

$$T_2=T_3\,[1,\ 2,\ 3,\ 4,\ 5] \quad \text{(соответствует } \exists\text{);}$$

$$T_1=T_2\,[(4,\ 5)\div(1,\ 2)]\,S_2 \quad \text{(соответствует } \forall\text{);}$$

$$T=T_1\,[2,\ 3] \quad \text{(получается с помощью проекции в соответствии с целевым списком).}$$

8.5. НОРМАЛИЗАЦИЯ РЕЛЯЦИОННОЙ МОДЕЛИ ДАННЫХ

Одним из центральных вопросов любой организации данных, претендующей на некоторую универсальность, является вопрос сохранения жизнеспособности прикладных программ при изменениях в базе данных. Обсудим этот вопрос применительно к реляционной модели данных, когда некоторые отношения базы данных изменяются [30, 36, 79]. Поставим перед собой следующие цели:

1) освободить набор отношений от нежелательных включений при модернизации и вычеркивании зависимостей;

2) снизить необходимость перестройки набора отношений, когда вводятся новые типы данных, что позволит увеличить период жизни прикладных программ;

3) сделать реляционную модель более информативной для использователя;

4) сделать набор отношений нейтральным для запросных статистик, когда эти статистики изменяются во времени.

Функциональная зависимость. Атрибут B отношения R функционально зависит от атрибута A отношения R, если в каждый момент времени каждое значение в A не имеет более одного значения в B, соотнесенного с ним посредством R. Другими словами, проекция $\Pi_{A,B}(R)$ определяет в каждый момент времени функцию $\Pi_B(R)$ от $\Pi_A(R)$ (эта функция может изменяться со временем). По определению $\Pi_A(R) = \{r[A] \mid r \in R\}$. Такие определения легко расширяются до набора атрибутов.

Будем писать $RA \rightarrow R.B$, если B функционально зависит от A в отношении R и $R.A \not\rightarrow R.B$, если такой зависимости нет. Если одновременно имеют место зависимости $R.A \rightarrow R.B$ и $R.B \rightarrow R.A$, то будем писать $R.A \rightarrow R.B$, устанавливая факт взаимно однозначного соответствия. Функциональная зависимость вида $R.D \rightarrow RE$, где E — подмножество D, а D — некоторое подмножество атрибутов R, называется *простой зависимостью*, в противном случае — *сложной зависимостью*.

Пример 8.15. Для иллюстрации простой и сложной функциональных зависимостей рассмотрим отношение

$$\text{СТ}\,(\text{Г}\neq,\ \text{С}\neq,\ \text{ФИО}).$$

Предположим, что каждый студент имеет единственный номер зачетной книжки и что он числится только в одной группе. Тогда

$$\text{СТ.}\ \text{С}\neq\ \rightarrow \text{СТ.}\ \text{ФИО};$$

$$\text{СТ.}\ \text{С}\neq\ \rightarrow \text{СТ.}\ \text{Г}\neq.$$

Можно также предположить, что в одной группе обучаются однофамильцы, т. е.

$$\text{СТ.}\ \text{ФИО} \not\rightarrow \text{СТ.}\ \text{С}\neq;$$

$$\text{СТ.}\ \text{Г}\neq\ \not\rightarrow \text{СТ.}\ \text{ФИО}.$$

Примером простой зависимости является $\text{СТ.}\,(\text{Г}\neq,\ \text{С}\neq) \rightarrow \text{СТ.}\,\text{С}\neq$, так как $\text{С}\neq$ включено в $(\text{Г}\neq,\ \text{С}\neq)$, а сложной — $\text{СТ.}\,(\text{С}\neq,\ \text{Г}\neq) \rightarrow \text{СТ.}\,\text{ФИО}$.

Вероятные ключи. Вероятный ключ K отношения R является комбинацией атрибутов (возможно, единственным атрибутом), обладающей следующими свойствами.

1. *Единственная идентификация.* В каждой выборке отношения R величина K единственным образом определяет эту выборку, т. е.

$$R.K \rightarrow R.\Omega,$$

где Ω обозначает набор всех атрибутов специфицированного отношения.

2. *Неизбыточность.* Не существует атрибута в вероятном ключе K, который мог бы быть удален без разрушения свойства 1. Всегда существует, по крайней мере, один вероятный ключ, так как комбинация всех атрибутов R удовлетворяет свойству 1. Следовательно, он удовлетворяет и свойству 2.

Некоторые другие свойства вероятных ключей можно вывести из свойств 1 и 2.

3. Каждый атрибут из отношения R функционально зависит от каждого вероятного ключа из R.

4. Набор атрибутов из отношения R в вероятном ключе K есть максимально независимое множество (т. е. каждое внутреннее подмножество атрибутов из K функционально независимо от каждого другого внутреннего подмножества атрибутов из K, и никакие другие атрибуты из R не могут быть добавлены без нарушения этой функциональной независимости).

Таким образом, можно убедиться в следующем:

свойство 1 логически эквивалентно свойству 3;

из свойств 1 и 2 следует свойство 4;

максимальное функционально независимое множество необязательно есть вероятный ключ.

Если в отношении имеются несколько вероятных ключей, то один из них выбирается в качестве основного. Для простоты в дальнейшем будем рассматривать случай, когда в отношении имеется единственный вероятный ключ, который, очевидно, является и основным.

Вторая нормальная форма. Напомним, что нормализованное отношение — это такое отношение, все домены которого являются простыми. Рассмотрим отношение РАСПРЕДЕЛЕНИЕ ПРЕПОДАВАТЕЛЕЙ (РП):

$$\text{РП}\,(\text{Г, Д, ВР, П, ПФИО}),$$

где Г — учебная группа; Д — дисциплина; ВР — вид работы; П — шифр преподавателя; ПФИО — фамилия, имя и отчество преподавателя. Будем считать, что одна и та же работа в группе по конкретной дисциплине может производиться несколькими преподавателями. Например, экзамен по физике в группе КЗ-01 могут принимать несколько преподавателей: Павлов В. В., Смирнов А. Г. и др. Приведем несколько выборок из отношения РП:

РП	Г	Д	ВР	П	ПФИО
	К2-01	ФИЗИКА	ЗАЧ.	062008	СМИРНОВ А. Г.
	К3-01	МАТ. АНАЛИЗ	ЭКЗ.	062201	ПАВЛОВ В. В.
	К3-01	МАТ. АНАЛИЗ	ЭКЗ.	175213	ЩЕНКОВ Б. В.
	К4-01	ФИЗИКА	ЭКЗ.	175213	ЩЕНКОВ Б. В.

Из сказанного следует, что атрибуты Г, Д, ВР, П образуют вероятный и основной ключи отношения РП. Будем называть атрибуты, образующие основной ключ, основными атрибутами отношения, остальные — неосновными. Тогда Г, Д, ВР, П — основные атрибуты, а ПФИО — неосновной атрибут отношения РП. По определению ключа атрибут ПФИО функционально зависит от атрибутов Г, Д, ВР, П:

$$\text{Р.П.}\ (\text{Г, Д, ВР, П}) \rightarrow \text{РП.ПФИО}.$$

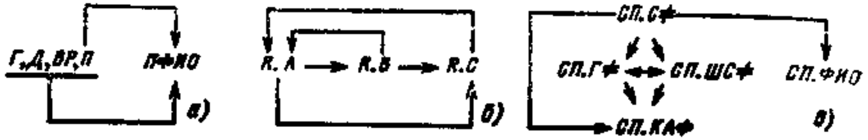


Рис. 8.7. Примеры зависимостей в отношениях:
а — функциональная; б — транзитивная; в — нетривиальная

Однако этот атрибут, кроме того, функционально зависит еще и от атрибута П, входящего в основной ключ (рис. 8.7, *а*): РП.П → РП.ПФИО.

Эта зависимость нежелательна по следующим причинам: во-первых, происходит неоправданное дублирование данных; во-вторых, шифр преподавателя связан с его фамилией в отношении, предназначенном для отражения текущих кратковременных зависимостей.

Если распределение работы преподавателей часто изменяется, то это приводит к необходимости соответствующего изменения выборок отношения РП. Это значит, что удаление некоторой выборки может привести вообще к удалению связи фамилии преподавателя и ее шифра из базы данных, что в принципе недопустимо.

Устранение зависимости неосновных атрибутов отношения от части атрибутов, образующих ключ, связано с переходом ко второй нормальной форме отношения (2НФ) и в общем случае связано с заменой отношения его проекциями. Первоначальное отношение восстанавливается из проекций путем обычного эквисоединения. Для рассматриваемого примера необходимо образовать две проекции РП: первую из атрибутов Г, Д, ВР, П; вторую из атрибутов П, ПФИО.

Чтобы дать определение 2НФ, введем понятие *полной зависимости*. Пусть F и E — два различных подмножества атрибутов некоторого отношения R и $RF \to R.E$. Если E функционально не зависит от любого подмножества F (не совпадающего с F), то E называется полностью зависимым от F в R. Таким образом, для отношения РП = {Г, Д, ВР, П} E = {ПФИО}, и нельзя сказать, что E полностью зависит от F, так как ПФИО функционально зависит от П.

Говорят, что отношение R находится на *второй нормальной форме* (2НФ), если, во-первых, оно нормализовано, т. е. находится в первой нормальной форме (1НФ), и, во-вторых, каждый неосновной атрибут полностью зависит от ключа отношения.

Отношение РП, состоящее из тех же самых атрибутов, может находиться в 2НФ, если предположить, что каждая работа выполняется только одним преподавателем. Например, экзамен у данной группы по данной дисциплине может принимать только один преподаватель. В этом случае ключ отношения состоит из атрибутов Г, Д, ВР, а атрибуты П и ПФИО находятся в функциональной зависимости от них, причем полной. Следовательно, РП находится в 2НФ.

Однако перечисленные нежелательные свойства этого отношения остались, следовательно, они бывают присущи и отношениям, находящимся в 2НФ. Причина заключается в транзитивной зависимости неосновных атрибутов от ключа отношения.

Транзитивная зависимость. Предположим, что для трех различных наборов атрибутов A, B, C отношения R выполняются следующие условия:

$$R.A \to R,B; \qquad R.B \to R.C; \qquad R.B \not\to R.A.$$

Отсюда следует, что $R.A \to R.C$ и $R.C \not\to R.A$ (рис. 8.7, *б*), при этом функциональная зависимость $R.C \to R.B$ не запрещается и не требуется. Тогда C транзитивно зависит от A в отношении R. Если $R.C \to R.B$, то B и C транзитивно зависят от A в отношении R. Рассмотрим отношение

СП (С≠, ФИО, Г≠, ШС≠, КАФ),

где С≠ — номер зачетной книжки студента; ФИО — фамилия, имя и отчество студента; Г≠ — номер группы; ШС — шифр специальности; КАФ — профилирующая кафедра.

Ясно, что каждый студент обучается в одной группе, а каждая группа имеет один единственный шифр специальности. Известно также, что каждая группа относится к какой-то профилирующей кафедре. Схематически все нетривиальные зависимости в СП показаны на рис. 8.7, *в*.

Если бы шифр специальности отсутствовал, то существовала бы только одна транзитивная зависимость атрибута КАФ от атрибута С≠. При наличии ШС≠ существуют две транзитивные зависимости: Г≠ и ШС≠ от С≠. Однако атрибут КАФ транзитивно не зависит ни от Г≠, ни от ШС≠.

Рассмотрим несколько выборок из отношения СП и проанализируем нежелательные свойства этого отношения:

СП (С≠,	ФИО,	Г≠	ШС≠	КАФ)
005	АЛЕКСАНДРОВ Б.А.	К1—01	0025	22
017	ПЕТРОВ А. Г.	К1—01	0025	22
301	ИВАНОВ А.В.	К2—01	1748	22
302	КУБРИН Е. Д.	К2—01	1748	22
623	ИВАНОВ А. В.	К3—01	2713	29
625	СИДОРОВ Б. Б.	К3—01	2713	29

Если изменяется специальность группы К1—01, то требуется обработать все выборки (их число изменяется со временем), чтобы провести соответствующее изменение. Аналогичная ситуация возникает при переводе группы на другую кафедру. Удаление выборки, связанной с некоторым студентом, приводит к удалению соответствующей информации о группе, если подобная строка является единственной.

Третья нормальная форма. Чтобы избавиться от ряда недостатков, присущих рассмотренной схеме и ей подобным, требуется перейти к третьей нормальной форме отношений (3НФ). Преобразование СП в 3НФ состоит в замене этого отношения двумя его проекциями:

$$СП_1 = П_{С\neq,\ ФИО,\ Г\neq}(СП); \qquad СП_2 = П_{Г\neq,\ ШС\neq,\ КАФ}(СП);$$

Представленные выборки теперь будут выглядеть следующим образом:

СП₁ (С≠	ФИО	Г≠);	СП₂ (Г≠	ШС≠	КАФ)
005	АЛЕКСАНДРОВ Б. А.	К1—01	К1—01	0025	22
017	ПЕТРОВ А. Г.	К1—01	К2—01	1748	22
301	ИВАНОВ А. В.	К2—01	К3—01	2713	29
623	ИВАНОВ А. В.	К3—01			
625	СИДОРОВ Б. Б.				

Система из указанных двух отношений, каждое из которых находится в 3НФ, обладает лучшими свойствами по сравнению с предыдущей. В любой момент можно восстановить первоначальное отношение путем обычного объединения.

Говорят, что отношение *R* находится в 3НФ, если оно представлено в 2НФ и каждый неосновной атрибут в отношении нетранзитивно зависит от каждого вероятного ключа в этом отношении. Любое отношение *R* в 3НФ обладает следующим дополнительным свойством. Каждый неосновной атрибут в отношении *R* полностью зависит, причем нетранзитивно, от каждого вероятного ключа в этом же отношении. Заметим, что сформулированное свойство не запрещает транзитивной зависимости основных атрибутов от вероятного ключа в отношении *R*.

Пусть C_2 — некоторый набор отношений в оптимальной 2НФ и требуется преобразовать его в 3НФ. Результирующий набор C_3 находится в оптимальной 3НФ относительно C_2 при выполнении следующих условий:

1. Набор отношений C_3 содержит минимально возможное число отношений (как и в случае оптимальной 2НФ), находящихся в 3НФ.
2. Каждое отношение в C_3 не содержит такой пары атрибутов, в которой один член пары транзитивно зависит от другого в некотором отношении из набора C_2 (т. е. атрибуты, состоящие «в дальнем родстве», отделяются друг от друга).

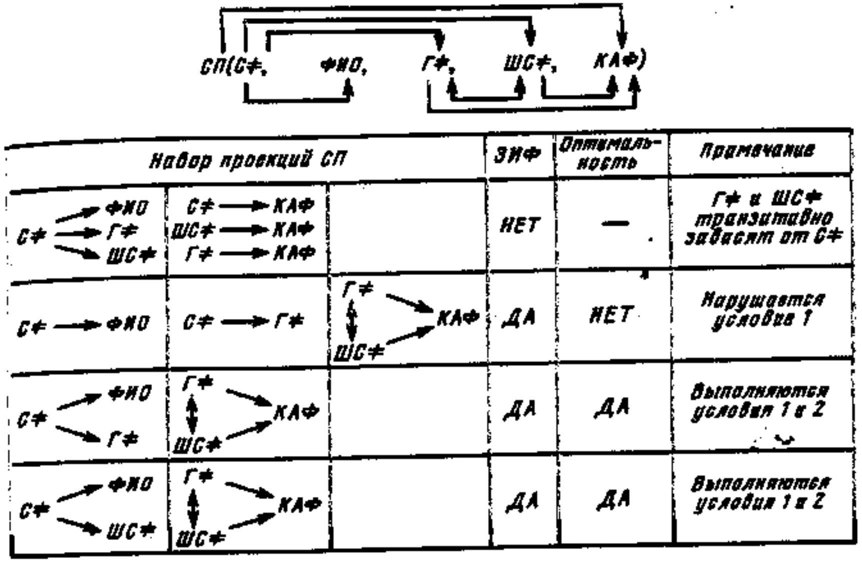


Рис. 8.8. Приведение отношений к третьей нормальной форме

Атрибут *С строго транзитивно* зависит от атрибута *А* в отношении *R*, если существует такой атрибут *В*, что

$$R.A \to R.B, \qquad R.B \nrightarrow R.A,$$
$$R.B \to R.C, \qquad R.C \nrightarrow R.B.$$

Преобразование отношения СП в оптимальную 3НФ показано на рис. 8.8. Заметим, что оптимальная 3НФ не единственна.

Рассмотрим снова отношение СП (С≠, ФИО, Г≠, ШС≠, КАФ) и отношения СП₁ (С≠, ФИО, Г≠) и СП₂ (Г≠, ШС≠, КАФ). Легко показать, что данные, представленные в виде двух отношений СП₁ и СП₂, несут больше информации, чем отношение СП.

Пусть СП₂ (Г≠, ШС≠, КАФ) представлено следующими выборками:

К1—01	0025	22
К2—01	1748	22
К3—01	2713	29
К4—01	0543	28

Если взять объединение СП₁ и СП₂ по Г≠, то получим состояние СП, которое рассматривалось ранее, но существенная

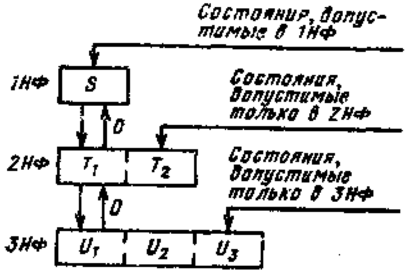


Рис. 8.9. Допустимые состояния базы данных

информация, заключенная в последней выборке К4-01 0543 28, потеряна.

Таким образом, схема, состоящая из двух отношений, позволяет включать и удалять выборки, не разрешенные схемой с одним отношением. Следовательно, эти схемы неэквиваленты по включению и удалению.

Запросная эквивалентность (Q-эквивалентность). Если θ — реляционная алгебра, B — выбор отношений и R — отношение, выводимое из B только с помощью операций θ, то говорят, что $R\theta$-выводимо из B. Пусть имеются две базы данных A и B, которые в момент времени t находятся в состояниях A_t и B_t. Говорят, что A_t и B_t Q-эквиваленты, если они θ-выводимы одно из другого и θ —полная реляционная алгебра.

Из рис. 8.9 видно, что множество S всех допустимых состояний базы данных в 1НФ Q-эквивалентны подмножеству T_1 всех допустимых состояний базы данных в 2НФ. Аналогично объединение подмножеств $T_1 \cup T_2$ для всех допустимых состояний базы данных в 2НФ Q-эквивалентно подмножеству T_2 все допустимых состояний 3НФ базы данных.

Итак, определение понятия функциональной зависимости позволило получить две новые формальные формы. При этом известны четыре формы, в которых может быть представлена база данных:

1) *ненормализованная форма.* Допускает домены, которые имеют отношение в качестве своих элементов (фактически иерархические базы);

2) *первая нормальная форма (1НФ).* Уничтожает сложные домены, т. е. домены, имеющие отношения в качестве элементов;

3) *вторая нормальная форма (2НФ).* Уничтожает неполные зависимости нескольких основных атрибутов от вероятных ключей;

4) *третья нормальная форма (3НФ).* Уничтожает транзитивные зависимости неосновных атрибутов от вероятных ключей.

Хотя три нормальные формы Q-эквивалентны, существует разница в их информационном содержании. Вторая нормальная форма более информативна, чем первая; третья более информативна, чем вторая. Ввиду большой информативности вводим дополнительные отношения. Следует также отметить, что 2НФ и особенно 3НФ более понятны случайному пользователю и более приспособлены к реализации.

8.6. ПРИКЛАДНЫЕ ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ РЕЛЯЦИОННОГО ИСЧИСЛЕНИЯ

Рассмотрим последовательно усложняющуюся серию примеров применения изложенных концепций реляционного исчисления. Все запросы на поиск данных будут связаны с конкретной базой данных, используемой в предыдущих примерах. Однако для того чтобы более четко представлять себе логическую структуру базы данных, сведем в таблицу все используемые отношения, а также нарисуем граф связей этих отношений (табл. 8.1).

Напоминаем, что символом ≠ отмечены ключевые домены отношений; Г — номер студенческой группы; С — номер зачетной книжки студента; ФИО — фамилия, имя и отчество студента; Д — дисциплина; ВР — вид работы (отчетности); ДТ — дата сдачи (текущая); П — шифр преподавателя; О — поставленная оценка. Приведем примеры запросов на языке.

Запрос 1. Найти тех преподавателей, которые ведут занятия в группе КЗ-01:

$$\{v_1[\text{П}]/(v_1\in\text{РП})\wedge(v_1\{\text{Г}\neq]=\{\text{КЗ-01}\})\}.$$

Запрос относительно прост. Здесь v_1[П] — целевой список; $(v_1\in\text{РП})$ — терм диапазона квалифицирующего выражения; $(v_1[\text{Г}\neq]=\{\text{КЗ-01}\})$ — терм объединения. Используя операции реляционной алгебры, выполняем сначала ограничение отношения РП, получая промежуточное отношение R_1:

$$R_1(\text{Г}\neq,\ \text{Д},\ \text{ВР},\ \text{П})=\text{РП}(\text{Г}=\{\text{КЗ-01}\}).$$

Наконец, получим окончательный список преподавателей в виде отношения R_r(П), проектируя отношение R_1 на домен П:

$$R_2(\text{П})=R_1(\text{П}).$$

Таблица 8.1

Отношения	Обозначения отношений	Домены
СТУДЕНТ	СТ	Г≠С≠ФИО
ОЦЕНКА	ОЦ	Г≠С≠Д.ВР.ДТ.П.О
КАЛЕНДАРНЫЙ ПЛАН	КП	ДТ.Т≠Д.ВР
РАСПРЕДЕЛЕНИЕ ПРЕПОДАВАТЕЛЕЙ	РП	Г≠Д.ВР.П

Запрос 2. Найти фамилии студентов тех групп, занятия в которых проводит преподаватель с шифром 062201. На α-языке запрос выглядит следующим образом:

$$U_1[\text{ФИО}]/(U_1 \in \text{СТ}) \wedge \exists (U_2 \in \text{РП}) \wedge U_1[\text{Г}] = U_2[\text{Г}] \wedge (U_2[\text{П}] = \\ = \{062201\}).$$

При формулировке данного запроса использованы две переменные выборки: U_1 и U_2. Переменная выборки U_2 связана квантором существования. На языке реляционной алгебры это означает:

1) ограничение отношения РП

$$R_1 = \text{РП}[\text{П} = \{062201\}].$$

2) выполнение проекции R_1 на домен Г для получения списка групп, в которых преподаватель 062201 проводит занятия

$$R_2(\text{Г}) = R_1(\text{Г});$$

3) соединение отношений СТ и R_2 по домену Г

$$R_3 = \text{СТ}\ (\text{Т} = \text{Г})\ R_2;$$

4) получение искомого списка студентов

$$R_4 = R_3[\text{ФИО}].$$

Запрос 3. Найти шифр преподавателей, которые не ведут занятия в группе КЗ-01. Запрос тесно связан с запросом 1. Сформулировать его на языке алгебры предельно просто. Действительно, необходимо составить список всех работающих преподавателей и исключить из него список, найденный при ответе на запрос 1. Это делается с помощью теоретико-множественного вычитания. Сформулировать же данный запрос на языке реляционного исчисления относительно сложно. В частности, запрос, сформулированный как

$$U[\text{П}]/(U \in \text{РП}) \wedge (U)[\text{Г}] \neq \{\text{КЗ-01}\},$$

неправильно отражает суть дела.

Правильно сформулированный запрос на языке реляционного исчисления для данного случая выглядит следующим образом:

Необходимо помнить, что надо очень внимательно относиться к формулировке запросов, в которых присутствует отрицание.

Запрос 4. Найти фамилии студентов группы КЗ-01, которые сдали все экзамены, предусмотренные календарным планом:

$$v_1[\text{ФИО}]/(v_1 \in \text{СТ}) \wedge \forall (U_2 \in \text{КП}) \wedge \exists (U_3 \in \text{ОЦ}) \wedge U_2[\text{Г}] = v_3[\text{Г}] \wedge \\ \wedge (v_3[\text{Д}] = U_3[\text{Д}] \wedge [U_2[\text{ВР}] = v_3[\text{ВР}]) \wedge (U_2[\text{ВР}] = \{\text{ЭКЗ}\}) \wedge \\ \wedge (U_3[\text{О}] > \{2\}) \wedge (v_3[C] = v_1[C]) \wedge (v_2[\text{Г}] = \{\text{КЗ-01}\}).$$

Заметим, что порядок расположения кванторов имеет существенное значение. Можно дать следующую алгебраическую интерпретацию этого выражения:

1) произведем ограничение отношения ОЦ

$$R_1(\text{Г},\ \text{С},\ \text{Д},\ \text{ВР},\ \text{ДТ},\ \text{П},\ \text{О}) = \text{ОЦ}[\text{О} > \{2\}; \\ \text{ВР} = \{\text{ЭКЗ}\}\quad \text{Г} = \{\text{КЗ-01}\}];$$

2) произведем ограничение отношения КП

$$R_2(\text{ДТ},\ \text{Г},\ \text{Д},\ \text{ВР}) = \text{КП}[\text{Г} = \{\text{КЗ-01}\};\quad \text{ВР} = \{\text{ЭКЗ}\}];$$

3) выполним операцию деления отношения R_1 на R_2 по домену Д

$$R_3(\text{Г},\ \text{С},\ \text{ВР},\ \text{ДТ},\ \text{П},\ \text{О}) = R_1[\text{Д} \div \text{Д}]\,R_2.$$

В полученном отношении R_3 содержатся номера зачетных книжек студентов, сдавших все экзамены. Чтобы получить их фамилии, необходимо соединить полученное отношение с отношением СТ и произвести необходимые проекции.

Для расширения избирательной способности реляционного исчисления можно использовать открытую библиотеку функций. Рассмотрим некоторые наиболее важные функции этой библиотеки:

COUNT — подсчитывает число элементов в конечном множестве;

MAX — выбирает максимальное значение в конечном множестве чисел;

MIN — выбирает минимальное значение в конечном множестве чисел.

Все указанные функции в качестве своего аргумента используют либо имя множества, либо выражения множества.

Запрос 5. Найти количество студентов, получивших пятерки по физике:

$$\{\text{COUNT}\ U_1[C \neq 1]/(U_1 \in \text{ОЦ}) \wedge ((U_1[0] = \{5\}) \wedge (U_1[\text{Д}]) = \\ = \{\text{ФИЗИКА}\})\}.$$

Формулировка запроса достаточно проста, особенностью его является использование функции множества. При реализации реляционного исчисления кроме средств, обеспечивающих поиск требуемых данных, необходимо обеспечить средства коррекции отношений. Обычно вводятся дополнительные процедуры коррекции типа УДАЛИТЬ, ВСТАВИТЬ и т. д.

8.7. ОПИСАНИЕ РЕЛЯЦИОННОЙ МОДЕЛИ БАЗЫ ДАННЫХ С ПОМОЩЬЮ КАТЕГОРИЙ

При разработке категориального метода представления базы данных [91] исходим из тех же посылок, что и при проектировании базы знаний, т. е. оперируем объектами данных, опирающимися на два понятия: объекты и отношения.

Представление бинарных отношений. Рассмотрим бинарный предикатный символ R и пусть $r: A\times B\to\{1,\ 0\}$ есть характеристическая функция множества $R\subseteq A\times B$.

Тогда диаграмма на рис. 8.10 коммутативна. Приняты следующие обозначения:

$$\rho_t^m(\langle x_1,\ .\ .\ .\ ,\ x_m\rangle)=x_t;$$

$$\rho_u^m(\langle y_2,\ .\ .\ .\ ,\ y_m\rangle)=y_u;$$

$$\widehat{ev}\,(tR_u)^m(\langle x_1,\ .\ .\ .\ ,\ x_m\rangle,\ \langle y_1,\ .\ .\ .\ ,\ y_m\rangle)=$$

$$=\begin{cases}1, & \text{если true;}\\ 0 & \text{в противном случае.}\end{cases}$$

Рассмотрим пару атрибутов $\ulcorner A\urcorner, \ulcorner B\urcorner$, которые являются именами доменов (областей) A, B соответственно. Будем считать, что элементами домена A служат A_i, $i=1,\ \ldots,\ k$, а элементами домена B являются B_j, $j=1,\ \ldots,\ l$.

1. Выполним коразложение A для отношения $R: A\to B$ согласно диаграмме на рис. 8.11, что дает вектор-строку морфизмов $(R_i: A_i\to B/i=1,\ \ldots,\ k)$, которую запишем в виде $(R_1,\ R_2,\ \ldots,\ R_k)$.

2. Выполним разложение A_i для морфизмов $R_1: A_l\to B$ согласно диаграмме на рис. 8.12, что дает вектор-столбец морфизмов $\{(R_i)^j: A_i\to B_j \mid j=1,\ \ldots,\ l\}$ для каждого i, т. е.

$$\begin{pmatrix}(R_i)^1\\ .\ .\ .\\ (R_i)^l\end{pmatrix},\qquad i=1,\ 2,\ .\ .\ .\ ,\ k.$$

$$\begin{tikzcd}
(A\times B)^m \arrow[r, "\langle\rho_t^m{,}\,\rho_u^m\rangle"] \arrow[dr, swap, "{[|tR_u|]}^m"] & A\times B \arrow[d, "r"] \\
 & \{1,0\}
\end{tikzcd}$$

Рис. 8.10. Пояснение к представлению бинарных отношений

$$\begin{tikzcd}
A_i \arrow[r, "in_i"] \arrow[dr, swap, "R_i"] & A \arrow[d, "R"] \\
 & B
\end{tikzcd}$$

Рис. 8.11. Пояснение к коразложению A

$$\begin{tikzcd}
A_i \arrow[r, "in_i"] \arrow[d, swap, "(R_i)^j"] \arrow[dr, "R_i"] & A \arrow[d, "R"] \\
B_j & B \arrow[l, "\pi_j"]
\end{tikzcd}$$

Рис. 8.12. Пояснение к разложению A_i

3. Опуская скобки, получаем матрицу

$B \diagdown A$	A_1	A_2	. . .	A_k
B_1	R_1^1	R_2^1	. . .	R_k^1
B_2	R_1^2	R_2^2	. . .	R_k^2
. . .	. . .	. . .	. . .	. . .
B_l	R_1^l	R_2^l	. . .	R_k^l

4. Рассмотрим коммутативную диаграмму на рис. 8.13. Нетрудно заметить, что эта диаграмма соответствует характеристической функции $r_i^j: A_i\times B_j\to\{1,\ 0\}$ для $R_i^j\in A_i\times B_j$, т. е. для $A_iR_i^jB_j$, которая указывает на принадлежность пары $(A_i,\ B_j)$ матрице **R** [значение $r_i^j(A_i,\ B_j)=1$] или на непринадлежность [значение $r_i^j(A_i,\ B_j)=0$].

Перейдем теперь к характеристической функции матрицы **R**, которую назовем характеристической матрицей для **R**. Она содержит единицы в тех позициях, которые соответствуют паре значений, принадлежащих отношению **R**. В остальных случаях в позициях содержатся нули. Матрица **r** имеет вид

$$\mathbf{r}=\begin{array}{c|c|c|c}
B\diagdown A & A_1 & \ldots & A_k\\ \hline
B_1 & r_1^1 & \ldots & r_k^1\\ \hline
\ldots & \ldots & \ldots & \ldots\\ \hline
B_l & r_1^l & \ldots & r_k^l
\end{array}$$

Анализ диаграмм позволяет записать в терминах матриц **R** и **r** следующее равенство:

$$\mathbf{r}=\widehat{ev}\,(\mathbf{R}),$$

которое назовем характеристическим.

$$\begin{tikzcd}
(A_i \times B_j)^m \arrow[r, "\langle \rho_t^m, \rho_u^m \rangle"] \arrow[dr, "{[\![t R_i^j u]\!]}"'] & A_l \times B_j \arrow[d, "r_i^j"] \\
& \{1, 0\}
\end{tikzcd}$$

Рис. 8.13. Пояснение к категориальному представлению матриц

Пример 8.16. Возьмем знак $\approx$ в качестве бинарного предикатного символа. Тогда характеристическая матрица определяется элементами

$$S_{A_i}^{B_j}(A_i, B_j) = \begin{cases} 1, & \text{если } A_i = B_j; \\ 0, & \text{в противном случае.} \end{cases}$$

Такой подход к представлению отношений справедлив и для случая n-арных отношений, однако требует рассмотрения n-мерных матриц. Это обобщение носит традиционный характер.

Реляционная алгебра. Остановимся на анализе матричной алгебры n-арных отношений, для которой поясним особенности представления операций реляционной алгебры. Пусть r — характеристическая функция отношения R и A — атрибут (домен) из R. Тогда $r.A$ обозначает A-компонент R. Если A — список $(A1, A2, \ldots, Ak)$ атрибутов из R, то $r.A = (r.A1, \ldots, r.Ak)$.

Пусть $C = (C_1, \ldots, C_n)$ — список всех атрибутов из R. Пусть теперь A — подсписок длины k из C и r — характеристическая функция отношения $R \subseteq C_1 \times C_2 \times \ldots \times C_n$. В этом случае через $r.\neg A$ обозначим $(n-k)$-характеристическую функцию r, где B — дополнительный к C список атрибутов, которые упомянуты в A. В терминах реляционной алгебры (РА) проекция отношения R на список атрибутов A определяется посредством РА $(R) = \{rA \mid r \in R\}$. Для перехода к характеристической матрице r для R приведем пример.

Пример 8.17. Пусть отношению $R(A, B)$ соответствует характеристическая матрица

$$\mathbf{r} = \begin{array}{c|ccc} {}_{B}\backslash^{A} & A_1 & \dots & A_k \\ \hline B_1 & r_1^1 & \dots & r_1^l \\ \dots & \dots & \dots & \dots \\ B^l & r_l^1 & \dots & r_l^k \end{array}$$

Тогда проекция R на домен A вычисляется следующим образом:

$$PA\ (R) = \left(\bigvee_{j=1}^{l} r_j^1 (A_1, B_j), \bigvee_{j=1}^{l} r_j^2 (A_2, B_j), \ \ldots\ , \bigvee_{j=1}^{l} r_j^k (A_k, B_j) \right),$$

где $\bigvee_{j=1}^{l}$ — символ дизъюнкции, а проекция R на домен B есть

$$\mathrm{PB\ (R)} = \begin{pmatrix} \bigvee_{i=1}^{k} r_1^i (A_i, B_1) \\ \bigvee_{i=1}^{k} r_2^i (A_i, B_2) \\ \dots\dots\dots \\ \bigvee_{i=1}^{k} r_l^i (A_i, B_l) \end{pmatrix}.$$

Отметим, что $\mathrm{PA(R)}$ — вектор-строка, а $\mathrm{PB(R)}$ — вектор-столбец.

Конкатенация (или произведение) двух отношений R и S определяется посредством

$$R \otimes S = \{(r, s);\quad r \in R \;\&\; s \in S\}.$$

При определении конкатенации характеристических матриц

$$\mathbf{r} = \widehat{ev}\,(R) \quad \text{и} \quad \mathbf{s} = \widehat{ev}\,(S)$$

получаем матрицу размерности dim, равной сумме размерностей матриц r и s, т. е. $\dim(\mathbf{r} \otimes \mathbf{s}) = \dim(\mathbf{r}) + \dim(\mathbf{s})$.

Для перехода к характеристической матрице r для R приведем пример.

Пример 8.18. Пусть R и S являются 1-матрицами, т. е. $R(\alpha)$ и $R(\beta)$. Тогда

$$\widehat{ev}\,(R \otimes S) = \widehat{ev}\,(R) \otimes \widehat{ev}\,(S) = r \otimes S$$

вычисляются согласно рис. 8.14.

$$r = \begin{array}{ccc} \alpha_1 & \dots & \alpha_k \\ \hline \multicolumn{1}{|c|}{r^1} & \multicolumn{1}{c|}{\dots} & \multicolumn{1}{c|}{r^k} \\ \hline \end{array}$$

$$S = \begin{array}{|c|c} \hline \beta & \\ \hline s_1 & \beta_1 \\ \hline \dots & \dots \\ \hline s_e & \beta_e \\ \hline \end{array} \qquad \begin{array}{|c|c|c|c|} \hline {}^{\beta}\backslash^{\alpha} & \alpha_1 & \dots & \alpha_k \\ \hline \beta_1 & r^1 \cdot s_1 & \dots & r^k \cdot s_e \\ \hline \dots & \dots & \dots & \dots \\ \hline \beta_e & r^1 \cdot s_e & \dots & r^k s_e \\ \hline \end{array} = r \otimes S = [\![R \otimes S]\!]$$

Рис. 8.14. Декартово произведение 1-матриц $r_j{}^i = r^j s_i$, где $\cdot$ — знак конъюнкции $r_i{}^j(\alpha_j, \beta_i) = r^j(\alpha_j) s_i(\beta_i) = r^j s_i(\alpha_j, \beta_i)$

$$r = \begin{array}{|c|c|c|c|}\hline \beta \backslash \alpha & \alpha & \cdots & \alpha_k \\ \hline \beta_1 & r_1^1 & \cdots & r_1^k \\ \hline \cdots & \cdots & \cdots & \cdots \\ \hline \beta & r_e^1 & \cdots & r_e^k \\ \hline \end{array} \qquad S = \begin{array}{|c|c|c|c|}\hline \delta \backslash \gamma & \gamma_1 & \cdots & \gamma^m \\ \hline \delta_1 & S_1^1 & \cdots & S_1^m \\ \hline \cdots & \cdots & \cdots & \cdots \\ \hline \delta_n & S_n^1 & \cdots & S_n^m \\ \hline \end{array}$$

Атрибуты	Переменные индексы
α	$x=1, \ldots, k$
β	$y=1, \ldots, e$
γ	$z=1, \ldots, m$
δ	$w=1, \ldots, n$

Рис. 8.15. Вычисление декартова произведения 2-матриц $r_{yw}^{xz} = r_y^x s_w^z = r^x r_y s^z s_w$, где · — знак конъюнкции

Пример 8.19. Пусть R и S являются 2-матрицами, т. е. $\mathbf{R}(\alpha, \beta)$ и $\mathbf{S}(\delta, \gamma)$. В этом случае $\widehat{ev}(R \otimes S) = r \otimes S$ вычисляются согласно рис. 8.15.

Результат вычисления характеристической 4-матрицы $r(\alpha, \beta) \otimes s(\gamma, \delta)$ представлен на рис. 8.16.

Пусть R, S — отношения, а $A=(A_1, \ldots, A_k)$ и $B=(B_1, \ldots, B_k)$ — списки атрибутов равной длины из R и S соответственно. В терминах PA соединением R по A с S по B считается отношение $R \ast S(A \theta B) = \{(r, s) : r \in R \& S \in S \& (r.A \theta r.B)\}$, где (r, s) — конкатенация характеристических функций r, s;

$$\theta \in \{=, \neq, >, <, \leqslant, \geqslant\};$$

$$r.A \theta r.B = (r.A_1 \theta s.B_1) \ \& \ . \ . \ . \ \& \ (r.A_k \theta s.B_k).$$

Наиболее часто на практике используется эквисоединение с модифицированным определением:

$$R \ominus S\,(A=B) = \{(r, s \urcorner B) : r \in R \,\&\, S \in S \,\&\, (r.A = r.B)\}.$$

Представим шаги вычисления характеристической функции эквисоединения отношений R и S:

1) вычисляется декартово произведение $r(A_1, \ldots, A_k) \otimes \otimes s(B_1, \ldots, B_n)$ с характеристической матрицей, элементы которой

$$r^{i_1 \ldots i_m} s_{j_1 \ldots j_n} = r_{j_1 \ldots j_n}^{i_1 \ldots i_m};$$

2) составляются характеристические матрицы для $r.A = r.B$:

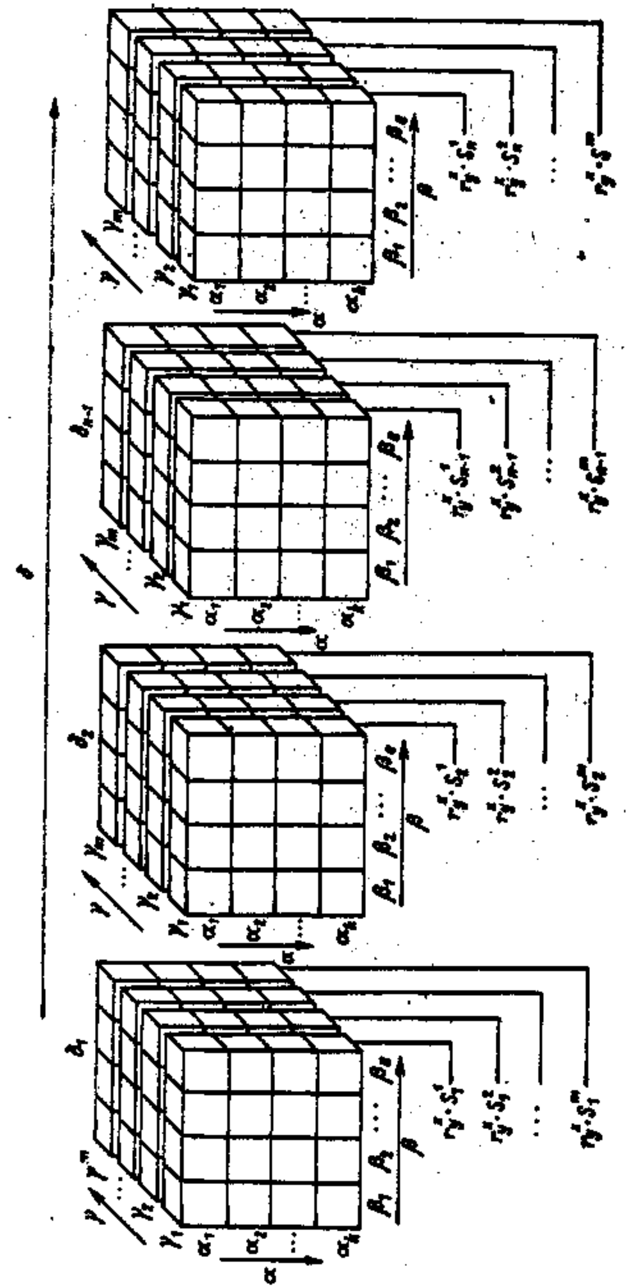


Рис. 8.16. Представление 4-матрицы декартова произведения 2-матриц $(r(\alpha, \beta) \otimes s(\gamma, \delta))$

$$\begin{array}{c|c|c|c} B_1 \backslash A_1 & A_1^1 & \ldots & A_l^1 \\ \hline B_1^1 & \delta_1^1 & \ldots & \delta_1^l \\ \hline \ldots & \ldots & \ldots & \ldots \\ \hline B_p^1 & \delta_p^1 & \ldots & \delta_p^l \end{array} \ldots \delta[k] = \begin{array}{c|c|c|c} B_{k_1} \backslash A_k & \delta_1^1 & \ldots & A_t^k \\ \hline B_{k_1} & \delta_1^1 & \ldots & \delta_t^l \\ \hline \ldots & \ldots & \ldots & \ldots \\ \hline B_{k_q} & \delta_q^1 & \ldots & \delta_q^1 \end{array},$$

где

$$\delta[\rho]_j^l(A_{\rho_l}, B_{\rho_j}) = \delta_j^l(A_{\rho_l}, B_{\rho_j}) = \begin{cases} 1, & \text{если } A_{1_l} = B_{1_j}; \\ 0 & \text{в противном случае;} \end{cases}$$

3) в произведении $r = r(A_1, \ldots, A_m) \otimes S(B_1, \ldots, B_n)$ выполняется диагонализация в подматрицах, участвующих в эквисоединении, т. е. элементы этих подматриц умножаются на соответствующие δ-элементы;

4) вычисляются проекции $P(A_1, \ldots, A_m, \neg B)(r)$, где $\neg B$ — список атрибутов, дополнительный к списку.

Пример 8.20. Представим эквисоединение отношений R и S по атрибуту $\beta : R(\alpha, \beta) \ominus S(\beta, \gamma)(\beta = \beta)$. Вычисление значения этого выражения и эквисоединение соответствующих характеристических матриц представлены на рис. 8.17. Из этого рисунка видна связь с декартовым произведением (см. рис. 8.16) эквисоединения характеристических матриц.

Другие, практически использующиеся операции PA представимы в терминах характеристических матриц точно таким же способом. В частности, относительным дополнением отношения R, определенного на доменах $D_1, \ldots, D_n$, считается отношение, определенное на тех же самых доменах, но посредством выражения

$$R'' = \{r : r \in (D_1 \times D_2 \times \cdots \times D_n) \& r \notin R\}.$$

В терминах характеристических матриц получаем условие, связывающее элементы r и r'' характеристических матриц отношений R и R'' соответственно, $\neg r_{i_1 i_2 \ldots i_n} = \overset{\bullet}{r}_{i_1 i_2 \ldots i_n}$.

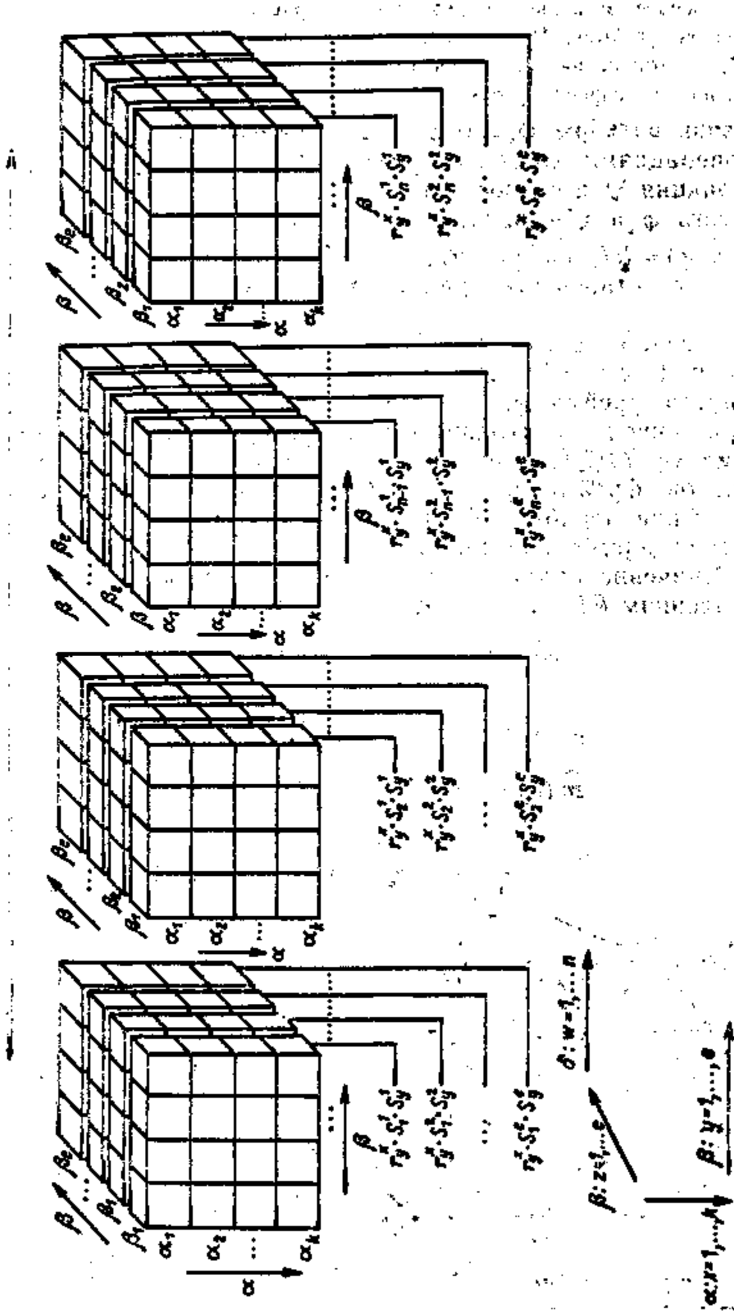


Рис. 8.17. Эквисоединения двух 2-матриц

На основе введенных определений операций *RA* установим их связь с алгеброй фреймов. Подчеркнем, что фреймы образуют содержание базы знаний, а отношения, являющиеся значениями фреймов, расширение базы знаний или расширение фреймов.

Связь алгебры фреймов и реляционной алгебры. Считаем, что операциями алгебры фреймов служат их конъюнкция &, дизъюнкция $\vee$ и разность & ¬[36, 91].

Пусть φ и ψ — фреймы; их значения обозначим посредством $\widehat{ev}(\varphi)=R1$, $\widehat{ev}(\psi)=R2$, где $R1$, $R2$ — расширения фреймов, т. е. отношения, рассматриваемые в реляционной алгебре.

Согласуем списки атрибутов во фреймах по следующему правилу. Пусть $X1$ — список атрибутов фрейма φ; $X2$ — список атрибутов фрейма ψ; Y — список общих атрибутов длиной n, т. е. n — число, подходящее как для фрейма φ, так и для фрейма ψ; $Y1 \subseteq X1$ — m — подсписок фрейма φ; $Y2 \subseteq X2$ — m — подсписок фрейма ψ; $Y1_i = Y2_i$ для $i=1, \ldots, m$, т. е. $Y1_i$ и $Y2_i$ — общие атрибуты фреймов φ и ψ, по которым будут выполняться алгебраические операции.

1. Значение m-конъюнкции фреймов φ и ψ вычисляется по их значениям $R1$ и $R2$ следующим образом:

$$\widehat{ev}(\varphi \& \psi)^m = \widehat{ev}(\varphi)^m \cap \widehat{ev}(\psi)^m;$$

где

$$\widehat{ev}(\varphi)^m = PY1\,(R1 \mathbin{⊜} R2\,(Y1=Y2));$$

$$\widehat{ev}(\psi)^m = PY2\,(R1 \mathbin{⊜} R2\,(Y1=Y2));$$

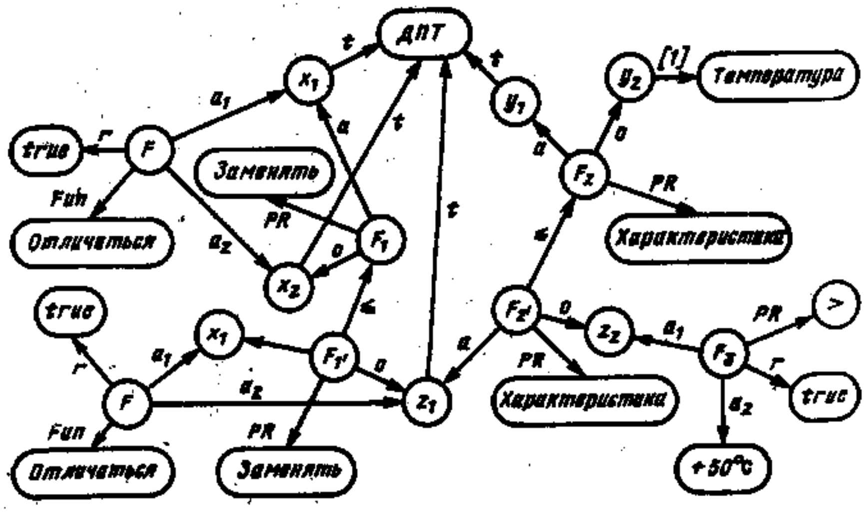


Рис. 8.18. Пример конъюнкции фреймов

$r_1 =$

ДПТ ДПТ		x_1		
		e2	e3	e4
x_2	e4	0	1	0
	e5	1	0	0
	e6	0	0	1
	e7	0	0	1
	e8	0	0	1

$r_2 =$

ДПТ Температ.		y_1						
		e2	e3	e4	e5	e6	e7	e8
y_2	30	0	1	0	0	1	0	0
	40	1	0	0	1	0	0	0
	55	0	0	1	0	0	0	0
	60	0	0	0	0	0	1	0
	70	0	0	0	0	0	0	1

Рис. 8.19. Характеристические матрицы расширения фреймов F_1 и F_2

$R1$ — значение фрейма φ; $R2$ — значение фрейма ψ; P — символ операции проекции; $⊜$ — символ операции эквисоединения.

2. Значение m-дизъюнкции фреймов φ и ψ вычисляется по их значениям $R1$ и $R2$ следующим образом:

$$\widehat{ev}(\varphi \vee \psi)^m = \widehat{ev}(\varphi)^m \cup \widehat{ev}(\psi)^m,$$

где

$$\widehat{ev}(\varphi)^m = PY1\,(R1 \otimes R2);$$

$$\widehat{ev}(\psi) = PY2\,(R1 \otimes R2).$$

3. Значение m-разности фреймов φ и ψ вычисляется по их значениям следующим образом:

$$\widehat{ev}(\varphi)^m = PY1\,(R1 * R2\,(Y1 \neq Y2));$$

$$\widehat{ev}(\psi)^m = R2''.$$

$r_1' =$

ДПТ ДПТ		x_1	
		e3	e4
z_1	e4	1	0
	e7	0	1
	e8	0	1

$r_2' =$

ДПТ Температ.		z_1		
		e4	e7	e8
z_2	55	1	0	0
	60	0	1	0
	70	0	0	1

Рис. 8.20. Характеристические матрицы расширения фреймов F_1' и F_2', участвующих в конъюнкции

$r'_1 =$

ДПТ \ ДПТ		x_1	
		e2	e4
z_1	e5	1	0
	e6	0	1

$r'_2 =$

ДПТ \ Температ.		z_1	
		e5	e6
z_2	30	0	1
	40	1	0

Рис. 8.21. Характеристические матрицы расширения фреймов F_1' и F_2' с учетом расширения F_3

Пример 8.21. Рассмотрим [91] такую ситуацию: «Двигатель постоянного тока (ДПТ) серии ДПМ заменяется на ДПТ серии ДПР, температура окружающей среды которых превышает +50 °С» (рис. 8.18). Для этого примера пусть $\widehat{ev}(F1)=R1$, $\widehat{ev}(F2)=R2$ является расширением фреймов $F1$ и $F2$ (рис. 8.19), тогда расширение $\widehat{ev}(F1\&F2)^1=\widehat{ev}(F1')^1\cap\widehat{ev}(F2')^1$, которое соответствует конъюнкции, показано на рис. 8.20. Расширение фрейма $F:\widehat{ev}(F)^1=\widehat{ev}(F1\&F2)\&\neg F3)^1$ приведено на рис. 8.21.

Глава 9

АЛГЕБРЫ И ИСЧИСЛЕНИЕ НЕЧЕТКИХ МНОЖЕСТВ

9.1. ОБЩИЕ СВЕДЕНИЯ

Модели логики предикатов, как и другие логико-алгебраические, не позволяют описать поведение систем, в которых фигурируют качественные, нечетко определенные понятия. Поэтому появились работы, посвященные нечетким кванторам и предикатам, и специалисты по СИИ обратили свое внимание на логики качественного характера, в частности на модальную логику [72].

Однако наибольший интерес у инженеров-кибернетиков вызвали работы по логике нечетких множеств [80, 81]. Методы принятия решений в детерминированных и случайных ситуациях относительно хорошо разработаны, но очень часто в реальных ситуациях человеку приходится принимать решения в условиях неопределенности, которые не носят ни случайный, ни игровой (антагонистический) характер. Эта неопределенность лежит в самом существе процесса принятия решения и проистекает, как правило, от неопределенностей природы, в которых приходится принимать решение, в том числе от неопределенностей противоположных сторон (партнеров, противников) в случае принятия решения в условиях коллектива, общества. Модели теории игр снимают такого рода неопределенность, сводя иногда ситуацию к модели случайной игры [30]. Однако совершенно ясно, что реальная неопределенность значительно шире и богаче игровых и других, ставших уже традиционными кибернетических моделей.

Логика нечетких множеств является попыткой создать более широкий математический аппарат для описания существующей неопределенности множеств принятия решений. При этом возникают две принципиальные особенности. Во-первых, само решение, получаемое с помощью моделей логики нечетких множеств, носит размытый, неопределенный характер. Во-вторых, возникают трудности в реализации систем принятия решений на ЭВМ с использованием логики нечетких множеств. С помощью алгебры нечетких множеств построены модели нечетких автоматов и грамматик, которые позволяют более точно описывать процессы принятия решений и естественные языки.

9.2. АЛГЕБРА НЕЧЕТКИХ МНОЖЕСТВ

В основе алгебры нечетких множеств лежат два основных понятия: нечеткого множества и нечетких операций (алгоритмов) над ними. При этом считается, что логика мышления при принятии решений отлична от двоичной и многозначной логики, так как она, как правило, имеет дело с нечеткими, размытыми, неопределенными классами объектов, понятий и размытыми, неопределенными преобразованиями над ними.

Определение. Нечеткое подмножество A множества элементов U области рассуждений определяется функцией принадлежности $\mu A: U\rightarrow[0, 1]$, которая связывает с каждым элементов y из U число $\mu A(y)$ в интервале [0, 1], которое представляет собой степень принадлежности y к A. Основой нечеткого множества A называется множество точек из U, для которых $\mu_A(y)$ положительно. *Размытым синглетоном* называется нечеткое множество, основой которого является единственная точка из множества U.

Если носитель множества A состоит из единственной точки, то такое нечеткое множество называется *одноточечным нечетким множеством* и обозначается $A=\mu/y$.

Определенное (четкое) *одноточечное* множество обозначается как $1/y$. Нечеткое множество A можно рассматривать как объединение составляющих его одноточечных множеств синглетонов и обозначается в общем случае

$$A=\int_U \mu_A(y)/y.$$

а при конечном числе элементов суммой

$$A=\mu_1/y_1+\mu_2/y_2+\ .\ .\ .\ +\mu_n/y_n=\sum_{i=1}^{n}\mu_i y_i.$$

Следует иметь в виду, что знак плюс в этой формуле означает не суммирование, а объединение (конъюнкцию). Понятие *нечеткого множества* поясним на следующих примерах.

Пример 9.1. Допустим, что

$$U=1+2+\ .\ .\ .\ +10.$$

Тогда нечеткое множество A (являющееся подмножеством, которое описывается понятием «несколько»), можно записать в следующем виде:

«несколько» $\triangleq$ 0,4/2+0,7/3+0,2/8+0,7/9, где знак $\triangleq$ означает равенство по определению.

Пример 9.2. Допустим, что

U = пинчер + гончая + овчарка + бульдог;

A — нечеткое множество, определяемое термином (признаком) *злой*. Тогда можем написать

Злой-сильно/пинчер + *слабо*/гончая + *средне*/овчарка + *средне*/бульдог

Нечеткие термины «сильно», «средне», «слабо» можно определить с помощью нечеткого множества $V=0+0{,}1+0{,}2+\ldots+1$, например, следующим образом:

сильно = 0,6/0,7 + 0,8/0,9 + 0,9/1;
средне = 0,6/0,3 + 0,8/0,5 + 0,5/0,9;
слабо = 0,8/0,1 + 0,1/0,7 + 0,5/0,6.

В этих примерах уже используются новые нечеткие или лингвистические переменные *несколько, злой, слабо* и т. д., характерные для логики нечетких множеств. Лингвистические переменные здесь используются, так же как и в повседневной практике, для обозначения наименования нечетких множеств типа цветов объекта *красный, желтый*, которые необходимо уточнить: *слабо, сильно* и т. д. Точной характеристикой цвета является длина волны, но всем известно, что в жизни она никогда не используется для характеристики цвета. Над нечеткими множествами выполняются те же операции, что и над обычными множествами.

Объединение нечетких множеств, обозначаемое как $A\bigcup B$, определяется соотношением

$$A\bigcup B\triangleq\int_V(\mu_A(y)\,\overline{\bigvee}\,\mu_B(y))/y,$$

где $\mu_A(y)\overline{\bigvee}\mu_B(y)=\max(\mu_A(y),\ \mu_B(y))$.

Эта операция соответствует союзу *или*, и поэтому, если имеются символы u и v нечетких множеств,

$$u\ \text{ или }\ v\triangleq u+v=u\bigcup v.$$

Пересечение нечетких множеств A и B, обозначаемое как $A\bigcap B$, определяется соотношением

$$A\bigcap B\triangleq\int_V(\mu_A(y)\,\underline{\bigwedge}\,\mu_B(y))/y,$$

где $\mu_A(y)\underline{\bigwedge}\mu_B(y)=\min(\mu_A(y),\ \mu_B(y))$.

Операция пересечения соответствует союзу u, и поэтому

$$u\ \text{ и }\ v\triangleq u\bigcap v.$$

Дополнение или *отрицание* определяется формулой нечеткого множества

$$\neg A\triangleq\int_U(1-\mu_A(y))/y.$$

Дизъюнктивная сумма определяется соотношением

$$A+B\triangleq(A\bigcup B)-(A\bigcap B).$$

Разность двух нечетких множеств определяется формулой

$$A-B\triangleq A\bigcap\neg B.$$

Проиллюстрируем введенные определения на примерах.

Пример 9.3. Пусть

$U=(u_1u_2u_3u_4u_5u_6u_7)$;
$A=\{(0{,}6/u_1)\ (0{,}4/u_2)\ (0{,}3/u_3)\ (0{,}8/u_4)\ (0{,}5/u_5)\ (1/u_6)\ (0{,}6/u_7)\}$;
$B=\{(0{,}9/u_1)\ (0{,}2/u_2)\ (1/u_3)\ (0{,}5/u_5)\ (0{,}8/u_6)\ (1/u_7)\}$;
$A\bigcup B\ \{(0{,}9/u_1)\ (0{,}4/u_2)\ (1/u_3)\ (0{,}8/u_4)\ (0{,}5/u_5)\ (0{,}8/u_6)\ (0{,}6/u_7)\}$;
$\neg A=\{(0{,}4/u_1)\ (0{,}6/u_2)\ (0{,}7/u_3)\ (0{,}2/u_4)\ (0{,}5/u_5)\ (0{,}4/u_7)\}$;
$A+B=\{(0{,}4/u_1)\ (0{,}4/u_2)\ (0{,}7/u_3)\ (0{,}8/u_4)\ (0{,}5/u_5)\ (0{,}2/u_6)\ (0{,}4/u_7)\}$;
$A-B=\{(0{,}1/u_1)\ (0{,}4/u_2)\ (0/u_3)\ (0{,}8/u_4)\ (0{,}5/u_5)\ (0{,}2/u_6)\}$.

Заметим, что если $\mu_A(y)$ принимает значение только 0 или 1, то множество A является обычным множеством. Запись $\mu_A(y)=1$ означает, что элемент $y\in U$ принадлежит множеству A, т. е. $y\in A$, а запись $\mu_A(y)=0$ означает, что $y\in U$ не принадлежит множеству A, т. е. $y\notin A$. Определенные ранее операции над нечеткими множествами дают правильные операции для обычных четких множеств. Продемонстрируем это на примерах.

Пример 9.4. Пусть U то же, что в примере 9.3.

Пусть $A=(u_1, u_2, u_3)$; $B=(u_3u_5u_7)$.

Тогда $A = \{(1/u_1)\,(1/u_2)\,(1/u_3)\,(0/u_4)\,(0/u_5)\,(0/u_6)\,(0/u_7)\}$;

$B = \{(0/u_1)\,(0/u_2)\,(1/u_3)\,(0/u_4)\,(1/u_5)\,(0/u_6)\,(1/u_7)\}$;

$A \cup B = \{(1/u_1)\,(1/u_2)\,(1/u_3)\,(0/u_4)\,(1/u_5)\,(0/u_6)\,(1/u_7)\} = (u_1u_2u_3u_5u_7)$;

$A \cap B = \{(0/u_1)\,(0/u_2)\,(1/u_3)\,(0/u_4)\,(0/u_5)\,(0/u_6)\,(0/u_7)\} = \{u_3\}$;

$A = \{(0/u_1)\,(0/u_2)\,(0/u_3)\,(1/u_4)\,(1/u_5)\,(1/u_6)\,(1/u_7)\} = (u_4u_5u_6u_7)$.

Произведение двух нечетких множеств A и B, обозначаемое как AB, определяется соотношением

$$AB \triangleq \int_U \mu_A(y)\,\mu_B(y)/y.$$

Аналогичным образом определяется операция *возведения в степень* неопределенного множества A, которая обозначается так:

$$A^\alpha \triangleq \int_U (\mu_A(y))^\alpha/y,$$

где $\alpha>0$.

В соответствии с этим вводится специально операция *концентрирования* неопределенного множества, т. е.

$$\mathrm{CON} \triangleq A^2.$$

Эта операция уменьшает степень принадлежности элементов множеству тем больше, чем меньше степень их принадлежности первоначальному множеству A.

Операция *растяжения* неопределенного множества является операцией, противоположной концентрации, и определяется соотношением $\mathrm{DIL}(A) \triangleq A^{0,5}$. Операция *контрастной интенсивности* определяется соотношением

$$\mathrm{INT}\,(A) \triangleq \begin{cases} 2A^2, & 0 \leqslant \mu_A(y) \leqslant 0{,}5; \\ \neg 2(\neg A)^2, & 0{,}5 \leqslant \mu_A(y) \leqslant 1, \end{cases}$$

где операция *отрицания* размытого множества определяется формулой

$$\neg A = \int_U (1-\mu_A(y))/y.$$

В соответствии с этой формулой имеем

A	INT (A)
0,4	0,32
0,7	0,82
0,9	0,98

Эта операция увеличивает значение $\mu_A(y)$, которое больше 0,5, и уменьшает те значения $\mu_A(y)$, которые меньше 0,5, уменьшая, таким образом, нечеткость A.

Операция *увеличения нечеткости*, которая обозначается как $F(A)$ или A, противоположна операции контрастной интенсивности и выполняет процедуру превращения четкого множества в нечеткое или увеличения степени нечеткости множества. Эта операция характеризуется ядром $K(y)$, представляющим собой нечеткое множество, которое является результатом применения операции F к одноэлементному множеству $1/y$, т. е.

$$K(y) \triangleq 1/y.$$

С помощью функции K операция F определяется соотношением

$$F(A;\ K) \triangleq \int_U \mu_A(y)\,K(y),$$

где $\mu_A(y)K(y)$ — произведение скаляра $\mu_A(y)$ на нечеткое множество $K(y)$, а $\int_U$ следует понимать как объединение нечетких множеств $\mu(y)K(y)$, $y \in U$. Определение операции увеличения нечеткости напоминает интегральные соотношения, использующие импульсную переходную функцию, роль которой в теории автоматического регулирования играет $K(y)$ [80, 81].

Пример 9.5. Рассмотрим соотношения

$U=1+2+3+4$; $A=0{,}8/1+0{,}6/2$; $K(1)=1/1+0{,}4/2$; $K(2)=1/2+0{,}4/1+0{,}4/3$. Тогда $F=0{,}8(1/1+0{,}4/2)+0{,}6(1/2+0{,}4/1+0{,}4/3)=0{,}8/1+0{,}32/2+0{,}6/2+0{,}24/1+$ $+0{,}24/3=0{,}8/1+0{,}6/2+0{,}24/3$.

9.3. АЛГЕБРА НЕЧЕТКИХ ОТНОШЕНИЙ

Наряду с нечеткими множествами и переменными в логике размытых множеств вводится в рассмотрение *нечеткое отношение* $R: X \to Y$ как нечеткое подмножество декартова произведения нечетких множеств $X \times Y$, которое определяется с помощью функции принадлежности двух переменных по формуле

$$R \triangleq \int_{X \times Y} \mu_R(x,\ y)/(x,\ y).$$

В общем случае n-арное отношение (n-отношение) определяется следующим образом. Пусть R — результирующее множество декартова произведения n множеств и μ — его функция принадлежности. Нечеткое n-отношение определяется как нечеткое подмножество R, принимающее какое-либо значение из μ в соответствии с формулой

$$R = \int_{X_1 \times \ldots \times X_n} \mu_R(x_1, \ldots, x_n)/(x_1, \ldots, x_n),$$

$$x_i \in X_i, \qquad i = 1, \ldots, n.$$

Приведем примеры.

Пример 9.6. Допустим, что

X = (Игорь, Ольга); Y = (Владимир, Людмила)

Дружба = 0,6 (Игорь, Владимир) + 0,9 (Игорь, Людмила) + 0,8 (Ольга, Владимир) + 0,2 (Ольга, Людмила).

Отношения удобно задавать с помощью матрицы отношений $\mu = \{\mu_{ij}\}$:

$$\mu = \begin{array}{c|cc} & \text{Владимир} & \text{Людмила} \\ \hline \text{Игорь} & 0{,}6 & 0{,}9 \\ \text{Ольга} & 0{,}8 & 0{,}2 \end{array}$$

Пример 9.7. Пусть

$$x = (x_1x_2x_3x_4x_5);$$

$$y = (y_1y_2y_3).$$

Нечеткое отношение R можно задать матрицей T_R, элементами которой будут $\mu(x, y)$:

$$\begin{array}{c|ccc} & y_1 & y_2 & y_3 \\ \hline x_1 & 1 & 0{,}8 & 0{,}6 \\ x_2 & 0{,}2 & 1 & 0{,}4 \\ x_3 & 0{,}0 & 1 & 0{,}7 \\ x_4 & 0{,}2 & 0{,}9 & 0{,}5 \\ x_5 & 0{,}3 & 0{,}4 & 1 \end{array}$$

В большинстве новых логико-алгебраических моделей, в том числе в реляционном исчислении и ситуационном управлении [33], в различных вариантах вводится исчисление отношений. Из того, что нечеткое отношение есть нечеткое множество, следует, что для него справедливы операции, определенные на нечетких множествах. Дадим определение нечеткого произведения нечетких отношений. Пусть заданы два нечетких отношения

$$R : X \to Y;$$

$$S : Y \to Z.$$

Тогда нечеткое произведение или композиция этих нечетких отношений определяется формулой

$$R \circ S \triangleq \int_{X \times Z} \overline{\bigvee_{y}} \left(\mu_R(x, y) \underline{\wedge} \mu_S(y, z)\right)/(x, y).$$

Напомним, что в этой формуле знаки $\overline{\vee}$ и $\underline{\wedge}$ обозначают операции взятия максимума и минимума соответственно. При конечных множествах значений x, y, z матрица $R \circ S$ равна максимальному произведению матриц отношений R и S, причем в этом произведении операции сложения соответствует знак $\overline{\vee}$, т. е. операция взятия максимума, операции умножения — $\underline{\wedge}$, т. е. операция взятия минимума, а $\circ$ — знак математической операции. Иначе композицию можно записать в виде

$$R \circ S = \int_{(x, z) \in X \times Z\, y \in Y} \max \min \{\mu_R(x, y), \mu_S(y, z)\}/(x, z).$$

Приведем пример произведения отношений.

Пример 9.8. Допустим, что

$$\mathbf{R} = \begin{pmatrix} 0{,}3 & 0{,}8 \\ 0{,}6 & 0{,}9 \end{pmatrix}; \qquad \mathbf{S} = \begin{pmatrix} 0{,}5 & 0{,}9 \\ 0{,}4 & 1{,}0 \end{pmatrix},$$

тогда

$$\mathbf{R} \circ \mathbf{S} \triangleq \begin{pmatrix} 0{,}3 & 0{,}8 \\ 0{,}6 & 0{,}9 \end{pmatrix} \circ \begin{pmatrix} 0{,}5 & 0{,}9 \\ 0{,}4 & 1{,}0 \end{pmatrix} = \begin{pmatrix} 0{,}4 & 0{,}8 \\ 0{,}5 & 0{,}9 \end{pmatrix}.$$

Так, например, первый элемент первого столбца получается следующим образом:

$$\underset{\min}{0{,}3 \cdot 0{,}5} + \underset{\min}{0{,}8 \cdot 0{,}4} = 0{,}3 \underset{\max}{+} 0{,}4 = 0{,}4.$$

Пример 9.9. Возьмем **R** из примера 9.5, а **S** определено следующим образом:

$$\mathbf{S} = \begin{array}{c|cccc} & z_1 & z_2 & z_3 & z_4 \\ \hline y_1 & 0{,}3 & 0{,}3 & 0{,}3 & 0{,}2 \\ y_2 & 0{,}3 & 0{,}2 & 0{,}7 & 0{,}8 \\ y_3 & 0{,}3 & 0{,}7 & 0{,}8 & 0{,}6 \end{array},$$

тогда

$$\mathbf{R} \circ \mathbf{S} = \begin{array}{c|cccc} & z_1 & z_2 & z_3 & z_4 \\ \hline x_1 & 0{,}3 & 0{,}6 & 0{,}7 & 0{,}8 \\ x_2 & 0{,}3 & 0{,}4 & 0{,}7 & 0{,}8 \\ x_3 & 0{,}3 & 0{,}7 & 0{,}7 & 0{,}8 \\ x_4 & 0{,}3 & 0{,}5 & 0{,}7 & 0{,}8 \\ x_5 & 0{,}3 & 0{,}7 & 0{,}8 & 0{,}6 \end{array}.$$

В общем случае в формуле композицию min можно заменить на какую-либо другую бинарную операцию $*$, тогда

$$R * S = \int_{X \times Z} \max_{y} \{\mu(x, y) * \mu(y, z)\}/(x, z).$$

Нечеткому отношению соответствует размытый граф, который определяется следующим образом.

Нечеткое подмножество такое, что

$$x^{(i)} \in x_i, \qquad i = 1,\ 2,\ .\ .\ .\ ,\ n;$$
$$\forall\ (x^{(1)},\ x^{(2)},\ .\ .\ .\ ,\ x^{(n)}) \in X_1 \times X_2 \times\ .\ .\ .\ \times X_n;$$
$$\mu\,(x^{(1)},\ x^{(2)},\ .\ .\ .\ ,\ x^{(n)}) \in M,$$

где M — множество принадлежности произведению $X_1 \times \times X_2 \times \ldots \times X_n$ называется *размытым графом.*

Введем для нечетких отношений операцию *проекции.*

Определение. **Проекцией** на произведение $x_{i1} \times x_{i2} \times \ldots \times \times x_{im}$ размытого отношения R, определенного на $X_1 \times X_2 \times \times \ldots X_n$, будет размытое множество с функцией принадлежности

$$\mu_R\,(x_{i1},\ x_{i2},\ .\ .\ .\ ,\ x_{im}) = \max_{x_{j1}} \max_{x_{j2}}\ .\ .\ .\ \max_{x_{jk}} \mu_R \times$$
$$\times\,(x_1,\ x_2,\ .\ .\ .\ ,\ x_n),$$

где $x_{j1},\ x_{j2},\ \ldots,\ x_{jk}$ — дополнение множества $(x_{i1},\ x_{i2},\ \ldots,\ x_{im})$ до $(x_1,\ x_2,\ \ldots,\ x_n)$.

Глобальной проекцией отношения называется число

$$h\,(R) = \max_{x_1} \max_{x_2}\ .\ .\ .\ \max_{x_n} \mu_R\,(x_1,\ x_2,\ .\ .\ .\ ,\ x_n).$$

Если $h(R) = 1$, отношение называется нормальным, если $h(R) < 1$, *субнормальным.*

Пример 9.10. Рассмотрим отношение R из примера 9.5. Для этого

$$\mu_R\,(x) = (1/x_1) + (1/x_2) + (1/x_3) + (0{,}9/x_4) + (1/x_5);$$
$$\mu_R\,(y) = (1/y_1) + (1/y_2) + (1/y_3);$$
$$b\,(R) = 1.$$

Отсюда заключаем, что отношение нормальное.

В исчислении нечетких отношений вводятся *нечеткие условные множества.*

Нечеткое подмножество $B(x) \subset Y$ будем называть условным по X, если его функция принадлежности зависит от $x \in X$ как параметр $\mu_B(y \| x)$.

Нечеткое множество $A \subset X$ будет порождать нечеткое множество $B \subset Y$, функция принадлежности которого

$$\mu_B\,(y) = \max\,(\min\,\{\mu_B\,(y \| x)\,\mu_A\,(x)\}),$$

т. е. $B\,(y) = B\,(y \| x)\,A\,(x)$.

Правило композиции условного нечеткого множества запишется в виде

$$B\,(y \| x) = (A \times B) + (A \times Y).$$

Пример 9.11. Пусть

$$A = \{(0{,}4/x_1)\,(0{,}7/x_2)\,(0{,}3/x_3)\};$$

$$T_R\,(y \| x) = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{cc} y_1 & y_2 \\ \left\{\begin{matrix} 0{,}3 & 0{,}7 \\ 0{,}2 & 0{,}6 \\ 1 & 0 \end{matrix}\right\} \end{array}.$$

Отсюда

$$B = \{(0{,}3/y_1)\,(0{,}5/y_2)\}.$$

Условные нечеткие отношения играют ту же роль, что и функции для формальных множеств. Для обычных множеств это можно выразить фразой «Если $x = a$, то $y = B$ под действием функции f», т. е.

$$x_i \sim\!\rightarrow y \quad \text{или} \quad y = f\,(x).$$

Для случая нечетких множеств пусть $x \in X$, а $y \in Y$ и существует R между X и Y. Тогда фраза «Если $X = A$, то $Y = B$ под действием отношения R» запишется как

$$A \sim\!\rightarrow B.$$

Замечание. Фраза «Если $X = A$, то $Y = B$ под действием отношения R» не означает справедливость утверждения «Если $Y = B$, то $X = A$ под действием отношения R'», где R' получается R транспозицией.

Для нечетких отношений можно определить такие свойства, как транзитивность, рефлективность, симметричность и т. д.

На базе теории размытых множеств можно построить нечеткую алгебру.

Дадим формальное определение нечеткой алгебры, под которой понимается совокупность нечетких операций над нечеткими множествами.

Нечеткая алгебра — это система вида $Z = \langle z + * - \rangle$, где Z имеет, по крайней мере, два разных элемента и $\forall x,\ y,\ z \in z$ и удовлетворяет следующему набору аксиом:

идемпотентность: $x + x = x,\ x * x = x$;

коммутативность: $x + y = y + x,\ x * y = y * x$;

ассоциативность: $(x + y) + z = x + (y + z), \quad (x * y) * z = x * * (y * z)$;

законы поглощения: $x + (y * x) = x,\ x * (y + x) = x$;

дистрибутивность: $x + (y * z) = (x + y) * (x + z)$;
$x * (y + z) = (x * y) + (x * z)$;

полнота: если $\bar{x} \in z$, то существует единственное дополнение $\bar{x} \vee x$, такое, что $\bar{x} \in z$ и $\bar{\bar{x}} = x$.

Законы де Моргана: $\overline{x+y}=\bar{x}*\bar{y}$, $\overline{x\times y}=\bar{x}+\bar{y}$. Будем рассматривать нечеткую алгебру определенной системы $z=<[0,\ 1]+,\ *,\ -,>$, где знаки $+$, $*$, $-$ понимаются как max, min и дополнение, определяемое формулой $\bar{x}=1-x$ для $\forall x$, $x\in[0,\ 1]$. Переменные с соответствующими индексами l_* и l_+ определяются как 0 и 1 соответственно.

Очевидно, что аксиоматике нечеткой алгебры подчиняются нечеткие множества и отношения.

В булевой алгебре существует $\bar{x}$, такое, что $x\bar{x}=0$ и $x+\bar{x}=1$, что несправедливо для нечеткой алгебры. Однако с помощью введенного ранее формализма нечеткой алгебры эти соотношения получаются как модификации соотношения идемпотентности следующим образом. Так, в булевой алгебре переменные могут принимать только значения 0 и 1. Эти соотношения идемпотентности для данного случая удовлетворяются. Действительно, $x*\bar{x}=0$, так как если $x=0$, то $\bar{x}=1-0=1$ и $x*\bar{x}=\min(x\bar{x})$; если $x=1$, то $\bar{x}=0$ и $x*\bar{x}=\min(x,\bar{x})=\min(1,0)=0$.

Аналогичным образом можно доказать, что

$$x+\bar{x}=\max\ (x,\ \bar{x})=1.$$

Поэтому можно утверждать, что булева алгебра — это размытая алгебра, но не наоборот. Можно ввести в рассмотрение нечеткие формы, которые определяются следующим образом:

1) числа 0 и 1 — нечеткие формы;
2) нечеткая переменная x_i — нечеткая форма;
3) если A — нечеткая форма, то $\bar{A}$ — нечеткая форма;
4) если A и B — нечеткие формы, то AB AB — нечеткие формы;
5) нечеткие формы — это те и только те формы, которые заданы п. 1—4.

Для нечеткой алгебры можно построить нормальные дизъюнктивные и конъюнктивные формы, используя следующие определения:

литерал — переменная x_i или $\bar{x}_i$;
статья — дизъюнкция одного или более литералов;
фраза — конъюнкция одного или более литералов;
форма — S-нормальная дизъюнктивная форма, если $S=P_1+$ $+P_2+\ldots+P_m$, $m\geqslant 1$ и P_{ikkm} являются статьями.

Дадим формальное определение лингвистической переменной, которая подробнее рассмотрена далее.

Лингвистическая переменная — это набор $(X,\ T(X)\cup G,\ M)$, где X — название переменной; $T(X)$ — терм, т. е. множество названий лингвистической переменной X, причем каждое из таких значений является размытой переменной со значениями из универсального множества; G — синтаксическое правило, порождающее названия для значений переменной X; M — семантическое правило, которое ставит в соответствие размытой переменной ее смысл.

Для нечеткой логики характерно то, что истинность принимает значения из интервала 0,1, является размытым подмножеством множества значений истинности и определяется функцией принадлежности, т. е. для высказывания A истинность есть нечеткое множество $T(X)$ с функцией принадлежности, определенной формулой

$$U(A)=\int_{x\in V}\mu(x)/X,$$

где $V\in[0,1]$ — интервал истинности. Тогда логические связи можно построить согласно принципу обобщения, а таблицы истинности — лишь приближенно.

В заключение введем понятие нечеткого конечного автомата и нечеткой грамматики.

Существуют несколько обобщений детерминированных конечных автоматов и преобразователей, одними из которых, например, являются вероятностные автоматы и преобразователи. *Нечеткий конечный автомат* представляет собой следующую пятерку:

$$A=(V,\ Q,\ \delta,\ q_0,\ F),$$

где V — входной алфавит (конечное множество символов); Q — конечное множество состояний автомата; δ — нечеткое отношение из множества $Q\times V$ в множестве Q; $q_0\in Q$ — начальное состояние автомата; $F\subseteq Q$ — множество заключительных состояний автомата.

Нечеткий конечный преобразователь в отличие от нечеткого конечного автомата помимо входного алфавита V_1 содержит выходной алфавит V_2, а δ представляет собой нечеткое отношение из $Q\times V_1$ в $Q\times V_2^*$.

Нечеткая грамматика, так же как и стохастическая [30], является обобщением формальной грамматики и представляет собой следующую четверку: $G(V_T,\ V_N,\ P,\ S)$, где V_T — словарь терминальных символов (непустое конечное множество); V_N — словарь нетерминальных символов (непустое конечное множество); P — конечное множество продукций вида $\alpha\to\beta$, причем $\alpha\in(V_T\cup V_N)^*$, $V_N(V_T\cup V_N)^*$, $\beta\in(V_T\cup V_N)^+$, i — степень, с которой цепочка α переводится в цепочку β.

Как и в случае обычных конечных автоматов, можно показать, что нечеткому автомату соответствует нечеткая автоматная грамматика, которая определяется аналогично стохастической автоматической грамматике.

9.4. ЛИНГВИСТИЧЕСКИЕ ПЕРЕМЕННЫЕ

Как было установлено, значениями лингвистических переменных являются элементы (символы) нечеткого множества в виде предложений на естественном или формализованном языке L.

Можно считать, что этот язык устанавливает соответствие между множеством терминов T и областью рассуждений v. Это соответствие можно описать с помощью нечеткого отношения N, которое каждому термину x из множества T ставит в соответствие суждение (объект) y из области v со степенью соответствия $\mu(x, y)$.

Пример 9.12. Допустим, что x=старый, y=60 лет. Тогда функция принадлежности μ_N (старый, 60) может быть равна 0,9.

Если зафиксировать x, то функция принадлежности $\mu_N(x, y)$ будет определять некоторое нечеткое подмножество $\mu(x)$ множества v с функцией принадлежности, определяемой соотношением

$$\mu_{M(x)}(y) = \mu_N(x, y), \quad x \in T, \quad y \in v. \qquad (9.1)$$

Нечеткое множество $M(x)$ является значением термина x, и хотя они из разных множеств [$x \in T$, а $M(x) \subset v$], очень часто вместо $M(x)$ будем использовать просто x.

В общем случае значение лингвистической переменной представляет собой составной термин $x = x_1, x_2, \ldots, x_n$, являющийся сочетанием элементарных терминов $x_1, x_2, \ldots, x_n$, например *не очень молодой*. Элементарные термины могут быть четырех типов:

1) первичные термины, являющиеся символами нечетких множеств области рассуждений (например, *молодой, старый* и т. д.);

2) отрицание *не* и союзы *и, или*;

3) неопределенности типа *очень, много, более или менее*;

4) *маркеры* типа вводных слов.

Основная проблема P_l, связанная с определением семантики лингвистических переменных, заключается в определении x по (9.1) [т. е. в определении $M(x)$] при заданных значениях каждого элементарного термина x_i $(i = 1, 2, \ldots, n)$ в составном термине $x = x_1, x_2, \ldots, x_n$.

Упростим задачу P_l до вычисления значения составного термина $x = hu$, где h — неопределенность, а u — термин с фиксированным значением.

Так, в терме x=*очень грубый человек* h=*очень*, u=*грубый человек*. Неопределенность можно рассматривать как оператор, который переводит нечеткое множество $M(u)$, представляющее собой значение u, в нечеткое множество $M(h, u)$. Очевидно, что неопределенности выполняют функции генерации большого множества значений для лингвистической переменной из небольшого набора первичных терминов. При составлении выражения для оператора неопределенности h удобно использовать над множествами типа концентрации, растяжения и увеличения нечеткости. На практике встречаются естественные неопределенности типа *очень, больше, меньше, много, слабо, вроде, вполне* и искусственные типа *минус, плюс*.

В обычном смысле неопределенность *очень* действует как усилитель и может быть интерпретирована с помощью операции концентрации, т. е.

$$\textit{очень} \quad x = x^2,$$

точнее

$$\textit{очень} \quad x = \int_v \mu^2 x(y)/y.$$

Пример 9.13. Допустим, что

$$U = 1 + 2 + 3 + 4 + 5;$$

слабый студент$=1/1+0{,}8/2+0{,}6/+0{,}3/4+0{,}5/5$; *очень слабый студент*$=1/1+$ $+0{,}64/2+0{,}36/3+0{,}16/7+0{,}04/5$.

Пример 9.14. Допустим, что x=*не очень слабый студент*. Первичный термин *слабый студент* определяется соотношением *слабый студент*$=1/1+$ $+0{,}8/2+0{,}6/3+0{,}4/4+0{,}2/5$.

Применяя операцию *очень* к множеству *слабый студент*, получаем *очень слабый студент*$=1/1+0{,}64/2+0{,}36/2+0{,}16/1+0{,}04/5$. Далее, учитывая формулу для операции отрицания $A \triangleq \int (1-\mu_A(y)/y$, получаем *не очень слабый студент*$=0{,}36/2+0{,}64/3+0{,}84/4+0{,}96/5+0{,}4/2+0{,}6/3+0{,}8/4+1/5$.

При составлении значения составного термина обычно используется *правило предшествования*, известное из булевой алгебры, которое в логике размытых множеств будет выглядеть следующим образом:

Предшествование	Операция
Первое	*h, не*
Второе	*и*
Третье	*или*

Для изменения порядка предшествования используются скобки и процедура разрешения неопределенностей путем объединений членов справа. Например, составной термин *плюс очень минус очень высокий* запишется в виде

плюс (очень (минус (очень (высокий)))).

При более общем подходе для определения значений лингвистической переменной может использоваться аппарат математической лингвистики Хомского при условии, что составные термины могут генерироваться бесконтекстной грамматикой.

Пример 9.15. Допустим, что значения лингвистической переменной представлены списком *маленький, не маленький, большой, не большой, очень маленький, не очень маленький, маленький или не очень большой, маленький и (большой или немаленький), не очень маленький и не очень большой* и т. д.

Пусть грамматика

$$G=(V_T, V_N, S, P),$$

где V_T — множество (словарь) терминальных символов (*маленький, большой, не* и т. д.); V_N — множество нетерминальных символов, обозначаемых через S; A, B, C, D, E, S — начальный символ; P — правила вывода. Результирующая система имеет вид

$$S\to A,\quad C\to D,$$

$$S\to S \ \textit{или}\ C\to E,$$

$$A\to B,\quad D\to \textit{очень}\ D,$$

$$A\to A\ \textit{и}\ B,\quad E\to \textit{очень}\ E,$$

$$B\to C,\quad D\to \textit{маленький},$$

$$B\to \textit{не}\ C,\quad E\to \textit{большой},$$

$$C\to (S).$$

Эти соотношения задают следующие правила вывода P:

$$\left.\begin{array}{l}
S\to S\ \ \textit{или}\ \ A\Rightarrow S_L=S_R+A_R;\\
A\to A\ \ \textit{и}\ \ B\Rightarrow A_L=A_R\cap B_R;\\
B\to\ \ \textit{не}\ C\Rightarrow B_L=\neg C_R;\\
D\to\ \ \textit{очень}\ D\Rightarrow D_L=D_R^2;\\
E\to\ \ \textit{очень}\ E\Rightarrow E_L=E_R^2;\\
D\to\ \ \textit{малый}\Rightarrow D_L=\textit{малый};\\
E\to\ \ \textit{большой}\Rightarrow E_L\ \ \textit{большой}.
\end{array}\right\} \qquad (9.2)$$

Нижние индексы L и R введены для того, чтобы различать символы, стоящие в левой и правой частях. Если задан составной термин X, то следует произвести его синтаксический анализ с помощью составленной ранее грамматики. Полученное синтаксическое дерево X служит основой для получения уравнений вида (9.2), которые гарантируют все нечетное множество значений X, в том числе и исходное. Например, если X=*не очень маленький и не очень большой*, то соответствующие уравнения для генерации всего нечеткого множества будут иметь вид

$$X=(\neg\ \text{малый}^2)\cap(\neg\ \text{большой}^2).$$

9.5. ИСЧИСЛЕНИЕ НЕЧЕТКИХ ВЫСКАЗЫВАНИЙ

Так же как в моделях логики исчисления высказываний, в логике нечетких высказываний (множеств) вводится операция импликации

$$A\Rightarrow B$$

или неопределенное высказывание «*если* A, тогда B», где A (антецедент) и B (консеквент) являются нечеткими высказываниями (множествами). Аналогичным образом можно записать

$$A\Rightarrow B\equiv\neg A\vee B.$$

Например, если x — *большой*, тогда y — маленький.

Очевидно, что неопределенное высказывание в виде связки-импликации можно определить как неопределенное отношение.

Как указывалось ранее, декартово произведение двух нечетких множеств A и B, определенных как подмножества областей рассуждений U и V соответственно, определяется соотношением

$$A\times B=\int_{U\times V}(\mu_A(u)\wedge\mu_B(v))/(u,\ v),$$

где $U\times V$ является декартовым произведением нечетких множеств, которое определяется как

$$U\times V=\{(u,\ v)/u\in U,\ v\in V\}.$$

Нечеткое произведение $A\times B$ является нечетким отношением U к V, так как оно представляет собой нечеткое множество упорядоченных пар $(u,\ v)$ со степенью принадлежности $(u,\ v)$ и $(A\times B)$, равной $\mu_A(u)>\mu_B(v)$.

Пример 9.16. Допустим, что

$$U=1+2;$$

$$V=1+2+3;$$

$$A=1/1+0{,}8/2;$$

$$B=0{,}6/1+0{,}9/2+1/3.$$

Тогда $A\times B=0{,}6/(1{,}1)+0{,}9/(1{,}2)+1/(1{,}3)+0{,}6/(2{,}1)+0{,}8/(2{,}2)+0{,}8/(2{,}3)$.

Напомним, что

$$a\triangle b=\begin{cases} a & \text{при}\ a\leqslant b;\\ b & \text{при}\ a>b.\end{cases}$$

Это произведение можно записать с помощью матрицы следующим образом:

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array}\\ \begin{array}{c} 1\\ 2 \end{array} & \begin{pmatrix} 0{,}6 & 0{,}9 & 1\\ 0{,}6 & 0{,}8 & 0{,}8 \end{pmatrix}. \end{array}$$

Произведение нечетких множеств удобно пояснить с помощью нечеткого условного высказывания «если A, тогда B, иначе C», которое более подробно можно записать как «если A, тогда B, если не A, тогда C». В этом случае нечеткое высказывание «если A, тогда B» можно понимать как «если A, тогда B, иначе C» с неопределенным C;

$$A \Rightarrow B = A \times B + (\neg A \times C), \tag{9.3}$$

где знак + обозначает объединение, а × — пересечение нечетких множеств.

Положив $C = V$, получим

$$A \Rightarrow B = A \times B + (\neg A \times V).$$

Наконец, полагая дополнительно, что $A = U$, получаем

$$A \Rightarrow B = U \times B + (\neg A \times V).$$

Основная проблема, которая возникает при синтезе нечетких правил вывода (или нечетких алгоритмов), заключается в том, что для заданного нечеткого отношения $R: U \leftrightarrow V$, определяемого нечеткими высказываниями, требуется для любого нечеткого подмножества $x \in U$ определить нечеткое подмножество $y \in V$, которое x индуцирует в V.

Пример 9.17. Допустим, что имеются два предложения, которые находятся в неопределенном отношении, определяемом соотношением (9.2):

1) x — *очень малый*;

2) *если* x — *очень мал, тогда* y — *большой, иначе* y *не очень большой.*

Требуется составить правило вывода для определения y, *если* x *очень мал*, которое будет обобщением известного правила.

Найдем неопределенное отношение $R: U \leftrightarrow V$, как неопределенное подмножество U Тогда y — как неопределенное подмножество V, которое индуцируется подмножеством x, определяется как произведение отношений:

$$y = x \circ R, \tag{9.4}$$

где x следует понимать как унитарное (один в один) отношение. Например:

$$\overset{x}{(0{,}2\ \ 1\ \ 0{,}3\ \ 0)} \circ \overset{R}{\begin{pmatrix} 0{,}8 & 0{,}9 & 0{,}2 \\ 0{,}6 & 1 & 0{,}4 \\ 0{,}5 & 0{,}8 & 1 \end{pmatrix}} = \overset{y}{(0{,}6\ \ 1\ \ 0{,}4)}.$$

Так, первый элемент произведения получается следующим образом:

$$\underset{\min}{0{,}2\cdot 0{,}8}+\underset{\min}{1{,}0\cdot 0{,}6}+\underset{\min}{0{,}3\cdot 0{,}5}=\underset{\max}{0{,}2+0{,}6+0{,}3}=0{,}6.$$

Допустим, что

$$U = 1+2+3+4+5;$$

малый $= 1/1+0{,}8/2+0{,}6/3+0{,}4/4+0{,}2/5$; *большой* $= 0{,}7/1+0{,}4/2+0{,}6/3+0{,}8/4+ +1/5$.

Подставляя в формулу для C *малый* вместо A, *большой* вместо B, *не очень большой* вместо C, получаем матрицу $\mathbf{R}$ для неопределенного высказывания *если малый, тогда большой, иначе не очень большой.* Произведение $x \circ R$, где x = *очень малый*, будет равно:

$$(1{,}0\ \ 0{,}64\ \ 0{,}36\ \ 0{,}16\ \ 0{,}04) \circ \left\{ \begin{matrix} 0{,}2 & 0{,}4 & 0{,}6 & 0{,}8 & 1 \\ 0{,}2 & 0{,}4 & 0{,}6 & 0{,}8 & 0{,}8 \\ 0{,}4 & 0{,}4 & 0{,}6 & 0{,}6 & 0{,}6 \\ 0{,}6 & 0{,}6 & 0{,}6 & 0{,}4 & 0{,}4 \\ 0{,}8 & 0{,}8 & 0{,}64 & 0{,}36 & 0{,}2 \end{matrix} \right\} =$$

$$= (0{,}36\ \ 0{,}4\ \ 0{,}6\ \ 0{,}81\ \ 1{,}0)$$

Если $R = A \Rightarrow B$ и $x = A$, то получаем выражение

$$y = A \circ (A \Rightarrow B) = B,$$

представляющее собой обобщенное правило modus ponens, которое строго выполняется для обычных множеств и приближенно для нечетких множеств A и B.

Глава 10

ОСНОВЫ λ-ИСЧИСЛЕНИЯ

10.1. ОБЩИЕ СВЕДЕНИЯ

Операция λ является удобным математическим средством связывания переменных, что имеет большое значение в общей теории программирования и обозначения [30, 38, 71, 82]. Заметим, что λ-операция впервые встретилась в исчислении предикатов, когда кванторы общности и существования связывали переменные и определялось множество истинности предиката $P(x_1, x_2, \ldots, x_n)$ с помощью так называемой операции абстракции λ-исчисления $\lambda(x_1, x_2, \ldots, x_n)$.

Пример 10.1. Так, областью предиката «$x>2$», определенного на множестве всех действительных чисел, является бесконечный интервал $(2, \infty)$, который обозначается с помощью λ-абстракции следующим образом: $(\lambda x) \times \times (x>2) = (2:\infty)$.

λ-функция (или лучше λ-абстракция) является более общим понятием, чем функция классического дифференциального и интегрального исчисления, так как дополнительно содержит в себе некоторый процесс вычисления функции (отсюда термин — «λ-исчисление»).

Пример 10.2. Функция двух аргументов $f(xy) = x^2 - y^2$ может быть определена двумя способами: $-\lambda xy(x^2-y^2)$ и $\lambda yx(x^2-y^2)$. Эти выражения отличаются тем, что в разной последовательности абстрагируются переменные x и y и в разной последовательности производится процесс присвоения переменным значений:

$$\left.\begin{array}{l}\lambda xy\,(x^2-y^2)\,(a,\ b)=\lambda y\,(a^2-y^2)=a^2-b^2;\\ \lambda xy\,(x^2-y^2)\,(a,\ b)=\lambda x\,(x^2-b^2)\,a=a^2-b^2,\end{array}\right\}$$

а конечные результаты совпадают.

Пример 10.3. Рассмотрим выражение $\int\limits_0^3 x^2dx=9$, левая часть которого является операцией с четырьмя аргументами: x, x^2, 0, 3. Обычно оно понимается как утверждение не о четырех объектах (x, x^2, 0, 3), а о трех объектах (функция, квадрат, 0, 3). Задать функцию — значит, указать закон соответствия, приписывающий значение функции каждому допустимому значению аргумента.

Допустим, имеется функция $f(x_1,\ x_2,\ldots,x_k)$, которая для каждого множества значений $x_1,\ x_2,\ldots,x_k$ ставит в соответствие значение функции $f(x_1,\ x_2,\ \ldots,\ x_k)$, причем x_i $(i=1,\ 2,\ \ldots,\ k)$ могут быть в свою очередь функциями. К примеру, выражение $x+y$ в общем виде можно считать либо функцией $f(x)$, либо функцией $g(y)$, либо $k(x,\ y)$. В λ-исчислении для различия этих трех функций вводится символ λ-абстракции $f=\lambda x.x+y$; $g=\lambda y.x+yk(x,\ y)=\lambda.x.y.x+y$. Говорят, что префикс λx абстрагирует функцию $\lambda x.x+y$ от выражения $x+y$, как бы образует новую функцию. Заметим, что здесь используется другой синтаксис λ-исчисления: «точка» вместо «скобки». Поэтому λ-оператор можно рассматривать как некоторый удобный механизм образования новых функций. Этот префикс λx дает способ построения значения функции f для каждого значения аргумента x. При этом благодаря операции абстрагирования λx переменная y и другие переменные остаются как бы в тени, в пассивном состоянии с помощью соответствующих операций абстрагирования. К примеру, в стандартном обозначении для функции f имеем $f(0)=0+y$, $f(1)=1+y$, а в обозначениях λ-исчисления эти выражения запишутся в виде

$$(\lambda x.x+y)\,(0)=0+y;$$
$$(\lambda x.x+y)\,(1)=1+y;$$

λ-оператор может быть распространен и на большее число переменных следующим образом. Например, выражению $x+y$ в теории λ-исчисления соответствуют две функции h, k от двух переменных $h(x,\ y)=x+y$; $k(x,\ y)=x+y$. Средствами λ-исчисления эти две функции запишутся в виде

$$h=\lambda x.y,\ x+y;\quad k=\lambda y.x,\ x+y.$$

Формальное (бестиповое) λ-исчисление — теория, обозначаемая через λ и изучающая функции и их аппликативное поведение

в отличие, например, от теории категорий, где функции изучаются в связи с их композицией. Поэтому *аппликация* (применение функции к аргументу) — исходная операция системы λ. Функция f в применении к аргументу a обозначается через fa. Операция, дополнительная к аппликации, называется *абстракцией*.

Пусть $t(\equiv t(x))$ — выражение, содержащее, быть может, переменную x. Тогда $\lambda x.t(x)$ — это такая функция J, которая сопоставляет аргумент a со значением $t(a)$. Иначе, имеет место

$$(*)\quad (\lambda x.t\,(x))\,a=t\,(a).$$

Поэтому можно заметить, что функции нескольких переменных можно свести к одноместным функциям. Например, если функция $f(x,\ y)$ — двухместная, то полагаем $f_x=\lambda y.f(x,\ y)$ и $a=\lambda x.f(x)$. Тогда $(ax)y=f_xy=f(x,\ y)$. Это очень важное свойство λ-исчисления, которое порождает равенство в теории категорий

$$X^{Y\times Z}=(X^Y)^Z.$$

Это равенство имеет место во всех декартово-замкнутых категориях, и почти все модели λ-исчисления являются в действительности объектами такой категории. Термы теории λ (λ-термы) составляют множество Λ и строятся из переменных с помощью аппликации и абстракции. Предложения теории λ — это равенства между λ-термами, а ее единственная математическая аксиома — это схема (∗).

Во избежание специальных обозначений для функции нескольких переменных используются такие функции, значениями которых являются не числа, а функции (функции от функции, рекурсивные). Например, вместо использования функции h двух аргументов представляют h с помощью h^* следующим образом:

$$h^*=\lambda x\,(\lambda y.x+y),$$

при этом

$$h^*\,(a)=\lambda y.a+y;$$
$$(h^*\,(a))\,(b)=(\lambda y.a+y)\,(b)=a+b.$$

Поэтому λ-исчисление может стать другой ветвью математики для исследования функции от функции, в частности теорией многократных интегралов от функции с переменным верхним пределом.

Продемонстрируем преимущества записи λ-исчисления как вычислительного предписания по сравнению со стандартным заданием функции [30, 38, 39, 73, 76, 82]. Что значит «функция $3x^2+1$»? Можно ввести функциональный символ f и утверждать: «функция $f:R\to R$, определенная отношением $f(x)=3x^2+$

$+1$. При этом, очевидно, переменную x можно здесь, не изменяя смысла, заменить на переменную y, так что x — связанная переменная. Лямбда-запись устраняет произвольность в выборе f в качестве функционального символа. Она предлагает вместо f выражение $\lambda x.3x^2+1$. Поэтому значение выражения $(\lambda x.3x^2+1)2$ равно 13. Говорят поэтому, что λ-запись преодолевает *нерегулярность обозначений* при обычно функциональном способе выражения.

Пусть P — оператор, т. е. сопоставление функций с функциями. Можно говорить, что «пусть $f(x)$ — функция и положим $g(x)=P(f(x))$». В общем случае это ведет к недоразумениям. Например, что такое $P(f)(x-1)$? То ли это $P(f)(x-1)$, то ли $P(g)(x)$, где $g(x)=f(x-1)$? А это не одно и то же, как это видно из следующего примера.

Пусть P определен соотношением

$$P(f)(x)=\begin{cases} 0 \text{ для } x\leqslant 0; \\ f(x) \text{ в противном случае.} \end{cases}$$

Тогда

$$P(f)(x-1)=\begin{cases} 0 \text{ для } x\leqslant 1; \\ f(x-1) \text{ в противном случае.} \end{cases}$$

$$P(g)(x)=\begin{cases} 0 \text{ для } x\leqslant 0; \\ f(x-1) \text{ в противном случае.} \end{cases}$$

Оказывается, на основании λ-абстракции можно построить различные варианты λ-исчисления. В λ-исчислении употребляются такие формальные системы, в которых используются общие способы применения одних функций к другим. λ-исчисление является одним из разделов математики, который называется *аппликативной теорией* (теорией аппликативных систем), где вместо операции взятия функции от аргументов применяется операция аппликации (приложения) к аргументам или переменным.

Частным случаем, одним из вариантов операции аппликации является λ-операция. В аппликации содержится, по крайней мере, последовательность вычисления функции (или выражения), процедура связывания переменных и т. д. Кроме того, λ-исчисление относится к внутренне противоречивым математическим теориям, что, несмотря на большие нарекания, оказалось привлекательным и полезным свойством для описания современных кибернетических систем и систем программирования, банков данных, систем искусственного интеллекта. λ-исчисление послужило основой для создания языка программирования LISP Маккарти [30], который является основным языком в искусственном интеллекте [38], а также языком-основой для всех конкретных языков программирования. После этих вводных замечаний перейдем к более строгому изложению λ-исчислений.

10.2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

Пусть M — некоторая формула, содержащая x в качестве переменной. Тогда $(\lambda x(M))$ есть функция, значение которой для любого значения аргумента α получается его подстановкой в M место x. Область интерпретации ω аргумента функции $(\lambda x(M))$ состоит из всех объектов $\alpha\in\omega$, таких, что формула M имеет смысл при подстановке в нее α вместо x. Множество всех значений формулы M, полученных для всех возможных значений аргумента $\alpha\in\omega$, назовем областью интерпретации формулы M.

Назовем операцию образования выражения $(\lambda x(M))$ из x и M λ-операцией. Введем λ-операцию для случая двух и более переменных.

Пусть M содержит в качестве переменных x_1 и x_2. Тогда $(\lambda x_1(\lambda x_2)M)$ есть функция, значение которой для любого значения α переменной x_1 получается его подстановкой в $(\lambda x_2(M))$ вместо x_1. Аналогично $(\lambda x_1(\lambda x_2(\lambda x_3)M))$ означает функцию трех переменных и т. д. Для n переменных $x_1, x_2, \ldots, x_n$ имеет место $(\lambda x_1(\lambda x_2(\ldots(\lambda x_n(M))\ldots)))$. Примем некоторые соглашения относительно обозначений. В дальнейшем будем иногда опускать скобки, когда смысл выражения ясен и без них, а для восстановления опущенных скобок будем следовать правилу «по ассоциации слева». Тогда предыдущие выражения будут выглядеть соответственно следующим образом:

$$\lambda x.M,\ \lambda x_1x_2M,\ \lambda x_1x_2x_3M \text{ и } \lambda x_1x_2\ .\ \ .\ \ .\ x_nM.$$

λ-исчисление как формальная система

Под формальной системой в современной математике понимается «алгебраическая» четверка: {Ал, Си, Ак, Се}, где Ал — алфавит; Си — синтаксис (синтаксические правила, правила построения правильных формул — ППФ); Ак — аксиомы и правила вывода; Се — семантика.

1. Алфавит. Под алфавитом a_λ λ-исчисления будем понимать:

три вспомогательных символа λ, (,), т. е. «лямбда», «левая скобка» и «правая скобка»;

счетное множество символов $a, b, c, \ldots, x, y, z, \ldots, a_1b_1, \ldots \ldots, x_ny_n\ldots$, которые назовем переменными.

Формулой или выражением λ-исчисления будем называть любую конечную (возможно, пустую) последовательность символов из алфавита a_λ.

В дальнейшем для обозначения формул λ-исчисления будем использовать заглавные буквы (при необходимости с индексами) $A, B, C, \ldots, M, N \ldots$

Пусть M есть формула, содержащая переменную x. Назовем вхождение переменной x в M связанным, т. е. x — связанной переменной тогда и только тогда, когда x содержится в части M, имеющей вид λx, N, где формула N содержит x. В противном случае x считается свободной переменной, а ее вхождение в M — свободным.

Назовем длиной формулы M общее количество вхождений в нее связанных переменных. В силу определения формулы ее длина всегда конечна.

2. Синтаксис. Синтаксис λ-исчисления определяется правильно построенными формулами (ППФ) λ-исчисления, которые определяются следующим образом:

а) свободная переменная x есть ППФ;

б) если F и G есть ППФ, то FG есть ППФ, и вхождение переменной x из F будет свободным или связанным в FG в зависимости от того, свободным или связанным является ее вхождение в F;

в) если M есть ППФ и содержит, по крайней мере, одно свободное вхождение x, тогда $\lambda x.M$ является *ППФ*; вхождение переменной y, отличной от x, в $\lambda x.M$ будет свободным или связанным, в зависимости от того, свободным или связанным является ее вхождение в M. Все вхождения x в $\lambda x.M$ являются связанными;

г) формула является правильно построенной, а вхождение переменной в нее свободным или связанным тогда и только тогда, когда она удовлетворяет условиям *а*—*в*.

3. Аксиомы. Аксиомами λ-исчисления являются:

а) $\lambda x.a$ есть a для всех x, отличных от a;

б) $(\lambda x.x)$ есть x.

4. Правила вывода I—III. Введем правила вывода, которые определяют условия, допускающие замену ППФ или ее частей на другие ППФ.

Предварительно введем синтаксическую конструкцию $S^x_N M$, означающую формулу, образуемую подстановкой N вместо x в M.

Другим обозначением $S^x_N M$ является $S^x_N M = [N/x]M$.

I. Любую часть вида $(\lambda x.M)N$ ППФ Q можно заменить выражением $S^x_N M$ в том и только в том случае, когда связанные переменные в M отличны от x и от свободных переменных из N.

II. Любую часть M правильно построенной формулы Q можно заменить выражением $S^x_y M$ в том и только в том случае, когда x является свободной переменной в M, а y не входит в M.

III. Любую часть вида $S^x_N M$ ППФ Q можно заменить выражением $(\lambda x.M)N$ в том и только в том случае, когда $(\lambda x.M)N$ есть ППФ, связанные переменные в M отличны от x и от свободных переменных из N. Это вводится, чтобы не спутать связанные и свободные переменные. Правило III в какой-то мере является «обратным» правилу II.

10.3. ОСНОВНЫЕ ПОНЯТИЯ ПРОЦЕДУРЫ И КОМПОНЕНТЫ λ-ИСЧИСЛЕНИЯ

Итак, мы ввели λ-исчисление как формальную систему. Перейдем теперь к рассмотрению необходимых нам для дальнейшего следстствий из введенных понятий.

Для произвольных ППФ M, M_1, M_2 и N свободной переменной x в M результат $S^x_N M$ подстановки N вместо каждого свободного вхождения x в M определяется следующим образом:

$$\text{1-е следствие: } S^x_N x \text{ есть } N; \tag{10.1}$$

$$\text{2-е следствие: } S^x_N a \text{ есть } a \text{ для всех } a\text{, отличных от } x; \tag{10.2}$$

$$\text{4-е следствие: } S^x_N (M,\ M_1)\text{, есть } \left(S^x_N M_1\right)\left(S^x_N M_2\right); \tag{10.3}$$

$$\text{5-е следствие: } S^x_N (\lambda x,\ M) \text{ есть } \lambda x.M; \tag{10.4}$$

$$\text{6-е следствие: } S^x_N (\lambda y.M) \text{ есть } \lambda y\left(S^x_N M\right) \text{ для всех } y\text{, отличных от } x \text{ и не входящих свободно в } N; \tag{10.5}$$

$$\text{7-е следствие: } S^x_N (\lambda y.M) \text{ есть } \lambda z\left(S^x_N\left(S^x_z M\right)\right) \text{ для всех } y\text{, отличных от } x \text{ и свободных в } N\text{, и всех } z\text{, не связанных ни в } M\text{, ни в } N. \tag{10.6}$$

Следствие 1 вытекает из аксиомы. Следствие 3 означает, что подстановка в произведение двух функций является произведением подстановок в каждую из функций. Следствия 4 и 5 означают запрет на связанные переменные (связанные переменные не трогать, они в «играх» не участвуют).

Конверсия-I (процедура imc). Если формула A может быть преобразована в формулу B с помощью одного из правил вывода I—III, то будем говорить, что A непосредственно конвертируется в B, или сокращенно A imc B.

Конверсия-II (процедура conv). Если существует конечная последовательность ППФ $A_1 A_2 \ldots A_m$, такая, что A_1 imc .... imc A_n, то будем говорить, что A_1 конвертируется в A_m, или сокращенно A_1 conv A_m. Процесс получения A_m из A_1 с помощью конечной последовательности применений правил вывода I—III назовем *конвертируемостью* A_1 в A_m, а отношение conv, которое связывает A_1 и A_m,— *отношением конвертируемости.*

Конверсия-III (процедура редукции). Преобразование, включающее только одно применение правила вывода II и не содержащее ни одного применения правила вывода III (количество применений правила вывода I при этом ограничивается), назовем *редукцией.*

Процедура imr. Если существует редукция ППФ A в ППФ B, то будем говорить, что A непосредственно редуктируется в B, или сокращенно A imr B.

Процедура red. Если существует преобразование ППФ A в ППФ B, состоящее из двух или более последовательных реакций, будем говорить, что A редуктируется в B, или сокращенно $A \operatorname{red} B$, причем применение правила вывода II к формуле будем называть *сжатием* или *операцией контрактирования части* $(\lambda x.M)N$ *формулы.*

Очевидно, имеют место следующие соотношения:

$$\lambda x.M \operatorname{imr} \lambda_y (S^x_y M), \tag{10.7}$$

если y не входит в M свободно и

$$(\lambda x.M)\, N \operatorname{imr} S^x_N M; \tag{10.8}$$

$$N \operatorname{red} M. \tag{10.9}$$

Нормальные формы. Будем говорить, что ППФ задана в нормальной форме (НФ), если она не содержит выражения вида $(\lambda x.M)N$.

Таким образом, нормальная форма λ-выражения будет иметь один из двух возможных видов:

а) нормальная форма 1 (НФ-1), обозначаемая

$$M, \tag{10.10}$$

где M — ППФ, не содержащая связанных переменных;

б) нормальная форма 2 (НФ-2)

$$\lambda x_1,\ x_2,\ .\ .\ .\ ,\ x_m M, \tag{10.11}$$

где M — ППФ, не содержащая связанных переменных.

Назовем B нормальной формой A, если B является нормальной формой и A conv B.

Очевидно, что A имеет нормальную форму, если существует формула B, являющаяся нормальной формой A.

Операции контрактирования и редукции служат для освобождения от связанных переменных. Говорят, что X контрактируется к Y тогда и только тогда, когда Y представляет собой результат замены части x вида $(\lambda x.M)N$ на $[N/x]M\{S_N{}^xM\}$. Говорят, что X редуктируется к Y (или $x \operatorname{red} Y$) тогда и только тогда, когда Y получается из X посредством конечной (возможно, пустой) серии контрактации и замен связанных переменных.

Выражение (терм) вида $(\lambda x.M)N$ называется *редексом*, а $[N/x]M$ *контрактом.* В этой терминологии можно считать, что x находится в нормальной форме 1, если не содержит редексов. Функция $\lambda x.M$, примененная к M, приобретает значение $[N/x]M$, и поэтому редукция в этом смысле может считаться процессом вычисления значений функций, которые входят в выражения.

Пример 10.4.

1. $(\lambda x.xy)\ F \operatorname{red} Fy$;
2. $(\lambda x.y)\ F \operatorname{red} y$;
3. $\big(\lambda x.\, (\lambda y.yx)z\big) v \operatorname{red} \big[S^x_v\big] \big(\lambda y.yx\big) z \equiv \big(\lambda y.yv\big) z \operatorname{red} \big[S^y_z\big] \big(yv\big) = zv.$

10.4. ФОРМАЛЬНОЕ ИЗЛОЖЕНИЕ λ-ИСЧИСЛЕНИЯ

Предыдущий материал был основан, главным образом, на оригинальных работах создателя λ-исчисления Черча [71]. На современном этапе λ-исчисления излагаются шаги с использованием других обозначений.

Предположим, что имеется бесконечная последовательность переменных и конечная или бесконечная последовательность констант. Атом определяется как переменная или константа.

Множество λ-термов определяется индуктивно: каждый атом есть λ-терм; если X и Y — термы, то (XY) есть λ-термы.

Примерами λ-термов могут служить

$$(\lambda x.\,(xy));\quad (\lambda y.y);\quad (\lambda x.\,(x.y));\quad (x\,(\lambda x.\,(\lambda x.x)));\quad (\lambda x.y).$$

В дальнейшем буквы x, y, z, u, v, ω будут обозначать переменные, причем различные буквы обозначают различные переменные. Заглавными буквами будем обозначать λ-термы. Скобки можно опускать, они восстанавливаются «по ассоциации влево», например W, X, Y, Z обозначает λ-терм $[((WX)Y)Z$, а $\lambda.xY$ обозначает $(\lambda x.(xY))]$. Тождество термов обозначается символом $\equiv$, так как символ $=$ будет использован для других целей. Наконец, вводится сокращение

$$\lambda x_1\ .\ .\ .\ x_n Y \equiv (\lambda x_1\,(\lambda x_2\,(\ .\ .\ .\ (\lambda x_n Y)\ .\ .\ .\ ))).$$

Предполагается, что все три класса термов (атомы $\lambda x.Y$, $x\,Y$) не пересекаются. Далее, если $xY \equiv UY$, то $x = U$ и $Y = V$, а если $\lambda x.Y \equiv \lambda$ и Y, то $x \equiv u$ и $Y \equiv U$.

Синтаксис и семантика λ-исчисления. После того как введены основные понятия и определены термы, приступим к построению λ-исчисления. Отношение X входит в Y (X является частью Y, или Y содержит X) и определяется индукцией по построению Y: x входит в X;

если X входит в U или в V, то X входит в (UV);

если X входит в U, то для любого Y/X X входит в $(\lambda y.U)$.

Предполагается интутивно ясным, что X входит в Y. Например, имеются два вхождения $(x.y)$ в $(x.y)$ $(\lambda x.xy)$.

Вхождение переменной x в Y связано тогда и только тогда, когда x находится внутри части Y, имеющей вид $\lambda x.z$. В противном случае x считается свободной переменной.

Говорят, что x входит свободно в Y (или $x \in Y$), если x имеет свободное вхождение в Y.

Систематическое применение символа λ позволяет очень точно различать свободные и связанные переменные. Именно в контексте исчисления математически строго уточняется использованное ранее правило подстановки.

Следует отметить, что x может быть свободной и связанной в одном и том же терме. Примером может служить $(x, (\lambda x.x))$. Для произвольных термов N, M и для любой переменной результат $[N/x]M$ подстановки N для каждого свободного вхождения x в M определяется индукцией по построению:

1) $[N/x]x \equiv N$;
2) $[N/x]a = a$ для всех атомов $a \not\equiv x$;
3) $[N/x](M_1M_2) \equiv ([N/x]M_1)([N/x]M_2)$;
4) $[N/x](\lambda x.Y) \equiv \lambda x.Y$;
5) $[N/x](\lambda y.Y) \equiv (\lambda y.[N/x]Y)$, если $y \equiv x$, $y \in N$ и $x \notin Y$;
$\equiv (\lambda z[N/x][z/y]Y)$, если $y \not\equiv x$, $y \in N$ и $x \in Y$.

В последнем случае Z — первая переменная, несвободная ни в N, ни в Y. Этот случай введен для предотвращения интуитивной зависимости подстановки $[N/x](\lambda y.Y)$ от связанной переменной y. Например, если не учесть такую специфику, то для всех y и w получится $[w/x](\lambda y.x) \equiv \lambda yw$.

Это выражение представляет собой постоянную функцию, если $y \not\equiv w$, и функцию-тождество, если $y \equiv w$. С учетом случая 5 для $y \equiv w$ имеет место выражение

$$[w/x](\lambda w.x) \equiv \lambda z.w,$$

которое все еще остается постоянной функцией.

С интуитивной точки зрения функция $\lambda x.M$, примененная к N, приобретает значение $[N/x]M$, а поэтому редукция в этом смысле может считаться процессом вычисления значения функций, которые входят в терм.

Пример 10.5. Рассмотрим следующие редукции:

1) $(\lambda x.xy)\,F$ red Fy;
2) $(\lambda x.y)\,F$ red y;
3) $(\lambda x.(\lambda x.yx)\,z)v$ red $[v/x]\,((\lambda y.yx)\,z) \equiv (\lambda y.yv)\,z$ red $[z/y]\,(yv) \equiv zv$;
4) $(\lambda x.xxy)\,(\lambda x.xxy)$ red $(\lambda x.xxy)\,(\lambda x.xxy)\,y$ red $(\lambda x.xxy)\,(\lambda x.xxy)\,yy$ и т. д;
5) $(\lambda x.xx)\,(\lambda x.xx)$ red $(\lambda x.xx)\,(\lambda x.xx)$ и т. д.

Говорят, что терм X находится в нормальной форме тогда и только тогда, когда X не содержит редексов.

Грубо говоря, нормальная форма U — это простейший терм, который имеет ту же самую интуитивную интерпретацию, что и U. В частности, в редукции 3 zv является нормальной формой для $(\lambda x.(\lambda y.yx)\,z)\,v$.

В редукциях 4 и 5 термов нет нормальных форм. Отметим, что многие термы редуцируются несколькими способами, например

$$(\lambda x.(\lambda y.yx)\,z)\,v \text{ red } (\lambda x.zx)\,v \text{ red } zv.$$

Семантику λ-исчисления, как это обычно принято в математической логике, связывают с интерпретациями. Точная интерпретация λ-термов зависит от конкретного применения теории. Вообще говоря, λ-термы представляют собой одноместную функцию, причем ее значение и аргументы в свою очередь могут быть функциями. Переменные представляют собой произвольные (одноместные) функции, (XY) означает результат применения функции X к аргументу Y. Если Y — терм, в который x входит свободно, то $(\lambda x.Y)$ представляет собой функцию, значение которой для аргумента A является результатом подстановки A вместо X в Y. Например, λx, xy представляет собой такую операцию приложения (аппликации) функции к некоторому аргументу таким образом, чтобы для всех термов F выполнялось

$$(\lambda x.xy)\,F = Fy$$

в том смысле, что обе части равенства имели бы одинаковую интерпретацию. Заметим, что $\lambda x.y$ представляет собой постоянную функцию, принимающую значение y для любого аргумента: $(\lambda x.y)F = y$. Отметим также, что терм (xx) означает результат приложения функции к самой себе. Очень часто такую разновидность термов оставляют неинтерпретированной.

10.5. СВЯЗЬ С ОБЫЧНОЙ ФОРМАЛЬНОЙ ЛОГИКОЙ

Оказывается, что λ-исчисление можно использовать для построения обычной системы формальной логики. На первый взгляд может показаться, что следует просто ввести термы, представляющие собой связки и кванторы, и постулировать обычные правила. Но такой способ приводит к противоречивости системы в том смысле, что имеет место для любого терма.

В действительности в качестве объектов выбирают дополнительно некоторые операции Π, Σ, E. В этом случае можно определить

$$\forall x X = \Pi\,(\lambda x.x)\,X \text{ для всех } x;$$

$$\exists xX = \Sigma\,(\lambda x.x)\,X \text{ для некоторого } x;$$

$$X \to Y \equiv \mathrm{E}\,(\lambda x.X,\ \lambda x.Y)\,Y \text{ для всех таких } x, \text{ что } Y \text{ существует.}$$

Здесь символ → использован для обозначения импликации. Это означает, что функциональная абстракция является единственной операцией, которая связывает переменные.

10.6. ТИПОВОЕ λ-ИСЧИСЛЕНИЕ

Определение 1. *Типовое λ-исчисление* (обозначенное λ^{τ}) — это теория [30, 82], определяемая следующим образом:

1. λ^{τ} имеет алфавит:

v_0^{σ}, v_1^{σ} ... — переменные для каждого типа $\sigma \in$Тур;

λ , (,) — вспомогательные символы.

2. Множество *термов типа* σ (обозначаемое через Λ^{τ}) индуктивно определяется следующим образом:

$$v_i^{\sigma} \subseteq \Lambda_{\sigma};$$

$$M \in \Lambda_{\sigma\to\tau};\ N \in \Lambda_{\sigma} \Rightarrow (MN) \in \Lambda_{\tau};$$

$$M \in \Lambda_{\tau};\ x \in \Lambda_{\sigma} \Rightarrow (\lambda x.M) \in \Lambda_{\sigma\to\tau},$$

где x пробегает переменные. Очень часто пишется $M \in \sigma$ вместо $M \in \Lambda_{\sigma}$. Множество типовых λ-термов (обозначаемое через Λ^{τ}) — это $U\{\Lambda_{\sigma} \mid \sigma \in \text{Тур}\}$.

3. Формулы системы λ^{τ} — это равенства $M = N$, где M, $N \in \Lambda_{\sigma}$ и тип $\sigma \in$Тур произволен.

4. Понятия свободной и связанной переменных, замкнутых термов и подстановки определяются очевидным образом: Λ_{σ}^{0} означает множество замкнутых термов типа σ.

5. Система λ^{τ} аксиоматизируется правилами и аксиомами для равенства (включая правило [82]) и схемой аксиом

$$(\beta)\,(\lambda x.M)\,N = M[x := N],$$

где типы таковы, что термы корректно построены.

6. Система $\lambda\eta^{\tau}$ — это λ^{τ} плюс схема

$$(\eta)\,\lambda x.Mx = M,$$

если $x \notin FV(M)$. Равенство в теориях λ и $\lambda\eta$ можно анализировать с помощью понятий редукции β и $\beta\eta$ соответственно. Нечто похожее верно для λ^{τ} и $\lambda\eta^{\tau}$.

Предложение. Имеется понятие редукции $\beta(\eta)$, адекватное для $\lambda(\eta)^{\tau}$, т. е. обладающее свойством Черча—Рассера. Приведем теорему, которая отмечает различие между λ и λ^{τ}.

Теорема 10.1. 1. Любой терм $M \in \Lambda^{\tau}$ имеет нормальную форму (НФ).

2. $\lambda(\eta)^{\tau}$ сильно нормализуема, т. е. любая $\beta(\eta)$-редукция λ^{τ}-терма заканчивается. Некоторым бестиповым λ-термам, например $\lambda x.x$, можно назначить типы $\lambda x^{\sigma} \cdot x^{\sigma} \in \sigma \to \sigma$. Это возможно для любого σ. В этой ситуации говорят, что $\sigma \to \sigma$ — возможный тип для $\lambda x.x \in \Lambda$. Другие λ-термы, например $\lambda x.xx$, не имеют типов.

Определение 1. Пусть $M \in \Lambda_{\sigma}$. Тогда $|M| \in \Lambda$ — это результат стирания всех типов в M. Говорят, что $|M|$ *типизируем* (или стратифицируем) и что σ — возможный тип для $|M|$.

2. Если $\sigma \in$ Тур, то частный случай σ^* типа σ — это любой результат замены всех вхождений типа 0 в σ любым другим типом. Например, $\sigma \to \sigma$ — частный случай типа $0 \to 0$ и типа 0.

10.7. ПРИМЕНЕНИЕ λ-ИСЧИСЛЕНИЯ В ЯЗЫКАХ ПРОГРАММИРОВАНИЯ

В тех разделах кибернетики, где рассматривают теорию программирования, используют λ-исчисление. Так, на его основе разработан язык программирования LISP (по англ. List Proccessing — обработка списков), который является языком функций. Это означает, что каждая конструкция, которая может быть записана на языке LISP и является функцией (в математическом смысле), выполняется в языке LISP как функция.

Пример 10.6. Выражение $\lambda AB(A^2 - B^2)$ записывается в языке LISP как

```
LAMBDA (AB) (DIFFERENCE (TIMES AA) (TIMES BB)).
```

Значением обращения к LAMBDA является само описание функции. В частности, если задать на языке LISP входную строку (LAMBDA(AB) (DIFFERENCE(TIMES AA) (TIMES BB))), то результатом работы LISP-системы будет число 27. Действительно,

$$AB.\,(A^2 - B^2)\,(6\ 3) = 6^2 - 3^2 = 27.$$

Пример 10.7. Пусть требуется записать программу вычисления квадрата некоторой величины: $\lambda x.x^2$, тогда ALGOL-процедура имеет вид.

```
integer procedure  p (x);
 integer x:  p : = x × x;
integer procedure  p (x);
  integer x: p : x × x.
```

Соответствующая запись на языке FORTRAN имеет вид

```
INTEGER FUNCTION p (x)
p = x × x
RETURN

END
```

LISP-выражение этой программы выглядит следующим образом:

```
DEFUN (((P (LAMBDA (X) TIMES  XX)))).
```

Тогда, например, **P**(7) = 49, т. е. $\lambda x.x^2\,(7) = 7^2 = 49$.

Не вдаваясь в подробности, заметим, что λ-исчисление (и, стало быть, LISP) служит основой для определения рекурсивных функций, которые имеют все возрастающее значение в искусственном интеллекте (например, при обработке с помощью ЭВМ фраз естественного языка).

Пример 10.8. Для простоты покажем, как в различных языках программирования рекурсивно определяется факториал:

$$0! = 1$$

$$n! = n\,(n-1)!$$

FORTRAN

```
INTEGEN FUNCTION FACT (N)
IF (N. EQO) GOTO 1
FACT = N • FACT (N — 1)
RETURN
END
```

ALGOL

```
integer procedure fact (n)
integer n;  fact: = if n = 0
then 1 else n × fact (n — 1)
```

LISP

```
(COND ((ZEROPN) 1) T (TIMESN (FACT (DIFFERENCE N1)))).
```

Здесь функция ZEROP проверяет, является ли ее аргумент нулем. Буква P в ZEROP означает predicate (предикат). Функция COND специфична тем, что каждый из ее аргументов — пара: первым элементом пары является условие, а вторым — величина или действие; FACT означает функцию-факториал, а T — true (истина). В данном случае в качестве аргументов функции COND выступают две пары.

И наконец, функция FACT определяется следующим образом:

```
DEFUN (((FACT (LAMBDA (N) (COND ((ZEROP N) 1) ×
    × (T (TIMES N (FACT (DIFFERENCE N1))))))))).
```

Глава 11

КОМБИНАТОРНАЯ ЛОГИКА. АЛГЕБРА И ИСЧИСЛЕНИЕ

11.1. ОБЩИЕ СВЕДЕНИЯ

Если в какой-нибудь системе имеются формальные переменные, то получается, что утверждения этой системы относятся к некоторым объектам, называемым переменными, и с формальной точки зрения так оно и есть. Но когда такую систему интерпретируют естественным образом, никаких содержательных объектов, соответствующих этим переменным, не обнаруживается. Другими словами, у таких систем нет естественной прямой интерпретации. Утверждения, включающие в себя формальные переменные, интерпретируются как утверждения о функциях, полученные из интерпретируемых утверждений с помощью некоторого содержательного аналога функциональной абстракции. Если можно было бы найти способ определения функциональной абстракции лишь в терминах обычных операций, то при формулировке системы формальные переменные не понадобились бы [30, 73, 76, 82].

Такое определение было предложено в комбинаторной логике, в которой изучаются системы комбинаторов. Системы комбинаторов решают ряд задач систем λ-исчисления, но не используют связанных переменных. Таким образом, преодолеваются трудности, связанные с операцией подстановки. Тем не менее в жертву приносится интуитивная простота системы λ-исчисления.

Закон коммутативности в арифметике определяется как утверждение «= для всех x, y; $x+y=y+x$».

Однако этот закон можно выразить без использования связанных переменных x и y, определив

$$A(x,\ y) = x+y \text{ для всех } x,\ y$$

и дополнительно представив оператор C:

$$(C(f))(x,\ y) = f(y,\ x) \text{ для всех } f,\ x,\ y.$$

В этом случае закон коммутативности записывается как $CA = A$.

11.2. ОСНОВНЫЕ ПОНЯТИЯ

Оператор C может быть назван комбинатором. Приведем другие примеры комбинаторов:

B — действует на две функции $(-B(f,\ g))(x) = f(g(x))$;
I — оператор тождества $I(f) = f$;
K — формирует постоянные функции $(K(a))(x) = a$;
S — определяется свойством $(S(f,\ g))(x) = f(x,\ g(x))$.

Не будем строго определять комбинатор в этом неформальном контексте, а построим формальную систему (термов), в которой комбинатор может быть представлен.

Предположим, что существует бесконечная последовательность переменных и конечная или бесконечная последовательность констант, включающая упомянутые комбинаторы.

Эти переменные и константы называются *атомами*. Множество комбинаторных термов определяется по индукции:

каждый атом есть комбинаторный терм;

если X и Y — комбинаторные термы, то (XY) — комбинаторный терм.

Комбинатор определяется как терм, содержащий только лишь атомы I, K, S. Понятие вхождения определяется естественным образом. В комбинаторных термах нет вхождений связанных переменных. Фраза «X имеет вхождение в M» может быть записана как $x \in M$.

Для любых термов M, N и любой переменной x результат $[N/x]M$ подстановки N для каждого свободного вхождения x в M определяется индукцией по построению M:

$$[N/x]\,x \equiv N;$$

$$[N/x]\,a = a \text{ для всех атомов } a \not\equiv x;$$

$$[N/x]\,(M_1M_2) \equiv ([N/x]\,M_1)\,([N/x]\,M_2).$$

Для различных переменных $x_1, \ldots, x_n$ и для любых термов $N_1, \ldots, N_n$ определим $[N_1/x_1, \ldots, N_n/x_n]M$ как результат одновременной подстановки N_1 вместо x_1, N_2 вместо x_2, N_n вместо x_n в M.

Чтобы комбинаторные термы выполняли те же функции, что и λ-термы, определим редукцию red комбинаторного терма и для каждой переменной определим операцию абстракции, записываемую в виде $[x]$ и соответствующую (λx) таким образом, что

$$([x]\,M)\,N \text{ red } [N/x]\,M.$$

Отношение редукции X red Y (читается как «X редуцируется к Y») определяется по индукции следующим образом.

Схемы аксиом:

$(I)\, IX$ red X для всех X;

$(K)\, KXY$ red X для всех X; Y;

$(S)\, SXYZ$ red $XZ(YZ)$ для всех X, Y, Z;

ρX red X для всех X.

Правила вывода:

$$(\mu) = \frac{X \text{ red } X'}{ZX \text{ red } ZX'};$$

$$(\nu) = \frac{X \text{ red } X'}{XZ \text{ red } X'Z};$$

$$(\tau) = \frac{X \text{ red } Y \text{ и } Y \text{ red } Z}{X \text{ red } Z}.$$

Термы IX, KX, Y, $SXYZ$ называются *редексами*, а соответственно X, X, $XZ(YZ)$ — контрактами. Получение *контракта* терма X означает замену одного вхождения редекса в X его контрактом.

Говорят, что X находится в нормальной форме тогда и только тогда, когда X не содержит редексов. Если U редуцируется к X, находящемуся в нормальной форме, то X называется нормальной формой U.

11.3. СИНТАКСИС КОМБИНАТОРНОЙ ЛОГИКИ

Основные редукции. Покажем, что комбинаторы **B**, **C** и **I** выражаются через **K** и **S**.

Пример 11.1. Можно определить, что $B = S(KS)K$, так как $BXYZ$ red $X(YZ)$. Действительно,

$$BXYS = S\,(KS)\,KXYZ \text{ red};$$

red $KSX(KX)YZ$ применением контрактов $S(KS)KX$;

red $S(KX)YZ$ применением контрактов KSX;

red $KXY(YZ)$ применением контрактов $S(KX)YZ$;

red $X(YZ)$ применением контрактов KXZ.

Пример 11.2. Определим $C = S(BBS)(KK)$. Тогда $CXYZ$ red XZY, поскольку

$$CXYZ = S\,(BBS)\,(KK)\;XYZ \text{ red};$$

red $BBSX(KKX)YZ$ применением контрактов $S(BBS)(KK)X$;

red $BBSXKYZ$ применением контрактов KKX;

red $B(SX)KYZ$ согласно примеру 11.1;

red $SX(KYZ)$ согласно примеру 11.1;

red $XZ(KYZ)$ red XZY.

Пример 11.3. $SKK\,X$ red (KX) red X. Следовательно, можно было бы определить $I = SKK$, а не выбирать I в качестве атома.

Для каждого терма M и переменной x определим терм $[x]M$ (функциональную абстракцию) индукцией по построению M:

1) $[x]x \equiv I$;

2) $[x]M \equiv KM$, если $x \in M$;

3) $[x]Ux \equiv U$, если $x \in U$ и не применимы случаи 2 и 3;

4) $[x]Uv \equiv S([x]U)([x]V)$.

Например, $[x]xy \equiv S([x]x)([x]y)$ по п. 4; $\equiv SI(Ky)$ по п. 1 и 2. Можно доказать, что $([x]M)N$ red $[N/x]M$.

Это легко выполняется разбором случаев 1—4 определения, приведенного ранее.

Для переменных $x_1, \ldots, x_m$, обязательно различных, абстракция определяется посредством

$$[x_1, \ldots, x_m]\,M \equiv [x_1]\,([x_2]\,(\ldots([x_m]\,M)\ldots)).$$

Приведем примеры.

Пример 11.4.

$$\begin{aligned}[x, y]\,x &\equiv [x]\,([y]\,x)\\ &\equiv [x]\,(Kx) \quad \text{по п.2;}\\ &\equiv K \quad \text{по п.3.}\end{aligned}$$

Пример 11.5.

$$\begin{aligned}[x,\ y,\ z]\, xz\,(yz) &\equiv [x]\,([y]\,([z]\, xz\,(yz)))\\ &\equiv [x]\,([y]\,(\mathrm{S}\,([z]\, xz)\,([z]\, yz))) \quad \text{по п.4;}\\ &\equiv [x]\,([y]\,\mathrm{S}_{xy}) \quad \text{по п.3;}\\ &\equiv [x]\, S_x \quad \text{по п.3;}\\ &\equiv \mathrm{S} \quad \text{по п.3.}\end{aligned}$$

Заметим, что из равенства X и Y еще не следует равенство $[x]X$ и $[x]Y$, поскольку $\mathrm{S}xyz=xz(yz)$, но $[x]\mathrm{S}xyz\neq[x]xz(yz)$, так как

$$[x]\ \mathrm{S}\ xyz = \mathrm{S}\,(\mathrm{SS}\,(\mathrm{K}\ y))\,(\mathrm{K}\ Z) \ \text{ и } \ [x]\ xz\,(yz) = \mathrm{S}\,(\mathrm{SI}\,(\mathrm{K}\ Z))\,(\mathrm{K}\ yz).$$

11.4. СВЯЗЬ λ-ИСЧИСЛЕНИЯ И ТЕОРИИ КОМБИНАТОРОВ

Было установлено, что комбинаторы и λ-термы решают одну и ту же задачу. Заметим, что обе теории, как ранее было показано, не являются в точности эквивалентными. Эквивалентность теорий достигается специальным введением так называемого расширенного равенства. Выразим комбинаторные термы посредством λ-термов (подробности см. в [82, 38]).

Для любого комбинаторного терма X определим λ-преобразование X, обозначаемое X_λ, со следующими свойствами

если X — атом, отличающийся от I, K, S, то $X_\lambda \equiv X$;

$$I_\lambda = \lambda u,\ u;$$
$$K_\lambda = \lambda uv.u;$$
$$S_\lambda = \lambda uvw.uw\,(vw);$$

если $X\equiv YZ$, то $X_\lambda = Y_\lambda Z_\lambda$ (здесь u, v, w — переменные).

Приведем примеры.

Пример 11.6. Если $Z\equiv X$, то

$$([x]\ Z)_\lambda = \mathrm{I}_\lambda = \lambda u.u = \lambda x.x = \lambda x.Z_\lambda.$$

Пример 11.7. Если $x\in Z$, то

$$([x]\ Z)_\lambda \equiv (\mathrm{K}\ Z)_\lambda \equiv (\lambda uv.u)\ Z_\lambda = (\lambda yx.y)\ Z_\lambda = \lambda x.Z_\lambda.$$

Пример 11.8. Если $Z = UV$, то можно показать, что

$$[x]\ Z = \mathrm{S}\,([x]\ U)\,([x]\ V)$$

и $$([x]\ Z)_\lambda = \mathrm{S}_\lambda\,([x]\ U)_\lambda\,([x]\ V)_\lambda = \mathrm{S}_\lambda\,(\lambda x.U_\lambda)\,(\lambda x.U_\lambda) = (\lambda yzx.yx\,(zx))\,(\lambda x.U_\lambda)\times$$
$$\times\,(\lambda x.V_\lambda) = \lambda x.\,((\lambda x.U_\lambda)\ x)$$
$$((\lambda x.V_\lambda)\ x) = \lambda x.U_\lambda V_\lambda = \lambda x.Z_\lambda.$$

Семантика комбинаторной логики. Интерпретация термов. Интерпретация комбинаторных термов, так же как и λ-термов, зависит от конкретных применений теории. Вообще говоря,

термы предполагают представления одноместной функции, и XY представляет собой результат применения X к аргументу Y. Некоторые из термов могут быть оставлены неинтерпретированными (возможна зависимость понимания XX от сферы применения).

Комбинаторы I, K, S можно интерпретировать так же, как в начале параграфа, хотя, например, K — это абстрактная запись построения постоянной функции из любого объекта, а не операция над множеством в обычном понимании. Такие конструкции с трудом допускают теоретико-множественную интерпретацию.

В настящее время теория комбинаторов — это передний край математической логики. Комбинаторы используются, во-первых, для построения логических систем, основанных на абстрактном понимании операции, во-вторых, для записи конструктивных функций в теории доказательств, в-третьих, при создании и анализе языков программирования в кибернетике.

11.5. ТИПОВАЯ КОМБИНАТОРНАЯ ЛОГИКА

(*i*) Типовая комбинаторная логика (CL^τ) — это комбинаторный вариант системы λ^τ. Она имеет исходные константы

$$K_{\sigma\tau} \in \sigma\to(\tau\to\sigma),$$

$$S_{\sigma\tau\rho} \in (\sigma\to(\tau\to\rho))\to(\sigma\to\tau)\to(\sigma\to\rho)$$

для всех типов σ, τ, $\rho\in$Тур и схемы аксиом

$$K_{\sigma\tau}MN = M;\quad S_{\sigma\tau\rho}MNL = ML\,(NL),$$

где M, N, L имеют подходящие типы.

(*ii*) Заметим, что типы комбинаторов K и S переходят в истинные высказывания, если → понимается как импликация [82].

(*iii*) При наличии экстенсиональности теории λ^τ и CL^τ становятся эквивалентными.

Понятие экстенсиональности. Лямбда-термы обозначают процессы. Различные термы могут обозначать один и тот же процесс. Например, $\lambda x.Mx$ и M' дают один и тот же результат MN в применении к терму N. Поэтому вводится следующее правило.

Определение. *Экстенсиональность* — это следующее правило вывода:

$$M_x = N_x \Rightarrow M = N\ (\mathrm{ext}),$$

где $x\notin\mathrm{FV}(MN)$.

Определение. (*i*) Множество FV(*M*) свободных переменных терма *M* можно индуктивно определить следующим образом:

$$FV(x)=\{x\};$$

$$FV(\lambda x.M)=FV(M)-\{x\};$$

$$FV(MN)=FV(M)\cup FV(N).$$

(*ii*) M называется *замкнутым термом или комбинатором,* если $FV(M)=\varnothing$.

$$(iii)\ \Lambda^0=\{M\in\Lambda/M\ \text{замкнут}\}.$$

$$(iV)\ \Lambda^0(\vec{x})=\{M\in\Lambda/FV(M)\subseteq\{\vec{x}\}\}.$$

(*iii*) *Замыкание терма* $M\in V$ есть по определению $\lambda\vec{x}.M$, где $\{x\}=EV(M)$. Например, $\lambda xy.xy\in\Lambda^0$; $\lambda xy.xyz\subset\Lambda^0(z)$.

2. Теория λ, расширенная этим правилом, обозначается через λ+ext.

Возникают вопросы: какая разница между λ+ext и λ и что можно доказать в λ+ext, но не в λ? Один пример уже был приведен выше — $\lambda x.Mx=M$. Можно доказать, что это, по существу, единственное различие.

Определение. Рассмотрим следующую схему аксиом η: $\lambda xMx=M$ (η — конверсия), где $x\not\subset FV(M)$. Пусть λη — это теория λ, расширенная схемой η.

Теорема 11.1 (Карри). Теории λ+ext и λη эквивалентны.

Доказательство. Покажем сначала, что $\lambda+\text{ext}\vdash\eta$. Действительно, $(\lambda x.Mx)$. Следовательно, если $x\in FV(M)$, то по правилу ext имеем $\lambda x.Mx=M$.

Наоборот, теория λη замкнута относительно правила ext. Пусть $Mx=Nx$, где $x\in FV(MN)$. Тогда по правилу ζ

$$M=N\Rightarrow\lambda x.M=\lambda x.N$$

имеем $\lambda x.Mx=\lambda x.Nx$. Отсюда $M=N$ по схеме η. Заметим, что правило ζ играет существенную роль при установлении эквивалентности между λ+ext и λη. Поэтому ζ иногда называют правилом *слабой экстенсиональности.* Экстенсиональное λ-исчисление обычно обозначается λη. Одна из причин рассмотрения λη состоит в том, что она обладает определенным свойством полноты.

11.6. КАТЕГОРИАЛЬНАЯ АБСТРАКТНАЯ МАШИНА (АБСТРАКТНЫЕ МАШИНЫ КОМБИНАТОРНОЙ ЛОГИКИ И λ-ИСЧИСЛЕНИЯ)

11.6.1. Общие сведения

В настоящее время имеются несколько механизмов, описывающих вычислимость. К ним относятся последовательности инструкций (алгоритмы) машин Тьюринга и Поста, нормальные алгоритмы Маркова, аппарат λ-конверсии.

Подобно тому как машина Тьюринга представляет собой абстрактный автомат, реализующий алгоритмы, аналогичные «машины» реализуют вычисления λ-выражений и термов комбинаторной логики. К ним следует отнести, прежде всего, машину SK-редукции Тюрнера [40], обеспечивающую редукцию термов комбинаторной логики в нормальную форму, и SECD-машину Ландака [40]. Машина SK-редукции действует путем применения к вычисляемому выражению последовательности переписывающих правил, подобно тому как это делается в нормальных алгоритмах Маркова.

SECD-машина представляет собой вариант стековой машины, состоящей из четырех компонентов: стека (*S* — Stack), окружения (E — enviroment), кода (C — code) и Dampa (*D* — (Dump). Действие машины состоит в применении кода вычисляемого λ-выражения к окружению (т. е. «оценке» выражения), промежуточные результаты при этом заносятся в стек и дамп.

Рассмотрим один из вариантов машины, обеспечивающей вычисление λ-выражений и, следовательно, аппаратную поддержку для функциональных языков программирования, — Категориальную Абстрактную Машину (КАМ) — Jhe Categorical Abstract machine (CAM). Преимущества этой машины по сравнению с SECD состоят в более простой структуре (три компонента против четырех), в простоте доказательства корректности вычисленного выражения. Кроме того, КАМ — более фундаментальная машина: комбинаторные термы есть не что иное, как последовательности инструкций такой машины, и КАМ позволяет реализовывать аппарат теории категорий, что само по себе уже является существенным расширением.

В основе представления λ-выражения, предназначенных для вычисления в КАМ, лежит формализм Де Брюйна, позволяющий в принципе избавиться от имен переменных. Выражения в нотации Де Брюйна преобразуются путем введения специальных комбинаторов в более удобную для вычислений форму (в частности, от синтаксико-семантических выражений осуществляется переход к чисто синтаксическим), после чего такие выражения преобразуются в последовательность машинных инструкций.

11.6.2. Категориальная комбинаторная логика

А. Формализм Де Брюйна. Необходимость введения формализма Де Брюйна диктуется проблемой коллизии имен переменных, возникающей в λ-исчислении. Действительно, в процессе применения правила β-редукции

$$(\beta)\ (\lambda x.M)\vec{y} = [y \to x]\, M$$

не допускается подстановка $[y \to x]$, в результате которой свободная переменная y становится связанной вследствие того, что она попала в область действия абстракции по другой переменной с тем же именем. Единственный выход в таком случае, позволяющий избежать коллизии, — переименование связанной переменной y путем применения правила α

$$(\alpha)\ \lambda y.M \equiv \lambda z\, [(z \to y]\, M)$$

для всех y, свободно входящих в M.

Так, выражение $(\lambda xy.x)\, y \equiv (\lambda x(\lambda y.x))\, y$ ни в коем случае не редуцируется в $\lambda y.y$.

Проблему коллизии имен переменных можно решить, избавившись от самих имен. Действительно, существенная информация о переменной в *замкнутом* (т. е. не содержащем свободных переменных!) терме — это только ее *глубина связанности*, т. е. количество абстракций, в область действия которых она попадает.

Определение. *Глубиной связанности* переменной x в замкнутом терме M называется количество букв λ (в развернутой записи λ-выражения) между данной переменной, входящей в терм, и связывающей ее буквой λ (исключая саму λ).

Прием Де Брюйна состоит в том, что переменные в теле терма заменяются числами, выражающими их глубину связанности, а имена переменных в абстракциях (стоящих при λ) удаляются.

Пример 11.9. Выражение $\lambda y(\lambda xy.x)\, y \equiv (\lambda y.(\lambda x.(\lambda y.x)))\, y$ преобразуется в $\lambda(\lambda\lambda.1)\, y$ путем видоизмененного правила β.

Это выражение приводится к виду $\lambda\lambda.1$ без всякого переименования переменных.

Нотация Де Брюйна записи λ-выражений интересна тем, что не только позволяет элегантно обрабатывать β-конверсию, но и может рассматриваться как комбинаторная логика специального вида с правилами, подобными правилам S и K комбинаторной логики.

Б. Означивание выражения. Необходимость означивания выражений возникает в λ-исчислении при переходе от синтаксиса к семантике, т. е. в процессе приписывания всем свободным

переменным и выражениям формального языка значений из некоторой области, называемой *окружением*. Окружение описывает связь значений переменных с идентификаторами.

Всякий терм M имеет в качестве означивания функцию (M), которая каждый элемент окружения сопоставляет с некоторым значением. В теории моделей λ-исчисления приняты следующие семантические равенства [82], описывающие означивание термов:

$$\left.\begin{aligned} &\widehat{\text{ev}}\,(x)\,\rho = \rho\,(x); \\ &\widehat{\text{ev}}\,(c)\,\rho = e; \\ &\widehat{\text{ev}}\,(MN)\,\rho = \widehat{\text{ev}}\,(M)_\rho(\widehat{\text{ev}}\,(N)_\rho); \\ &\widehat{\text{ev}}\,(\lambda x.M)\,\rho d = \widehat{\text{ev}}\,(M)\,\rho\,[x \leftarrow d], \end{aligned}\right\} \qquad (11.1)$$

где ρx — значение, ассоциированное с переменной x в окружении ρ; c — константа, обозначающая значение c; $\rho[x \leftarrow d]$ — окружение, в котором значение для переменной x заменено значением d.

Теперь установим вид этих семантических равенств в нотации Де Брюйна. Для этого положим, что окружение имеет вид $\rho \rightleftarrows (\ldots((),\ v_n),\ \ldots,\ v_0)$, где v_i — значение, ассоциированное с переменной, обозначенной в терме номером i. Теперь очевидным образом получаем следующую группу семантических равенств:

$$\left.\begin{aligned} &\widehat{\text{ev}}\,(0)\,(\rho,\ d) = d; \\ &\widehat{\text{ev}}\,(c)\,\rho = c; \\ &\widehat{\text{ev}}\,(MN)\,\rho = \widehat{\text{ev}}\,(M)_\rho(\widehat{\text{ev}}\,(N)_\rho); \\ &\widehat{\text{ev}}\,(\lambda.M)\,\rho d = \widehat{\text{ev}}\,(M)\,(\rho,\ d); \\ &\widehat{\text{ev}}\,(n+1)\,(\rho,\ d) = \widehat{\text{ev}}\,(n)\,\rho. \end{aligned}\right\} \qquad (11.2)$$

Легко видеть, что смысл первой пары равенств состоит в «извлечении» из списка значения под номером n. Второе и третье равенства остаются практически без изменения. Смысл последнего равенства заключается в том, что значением переменной с номером (именно по ней производится абстракция в $\lambda.M$) становится d.

В дальнейшем означивания как таковые нас интересовать не будут, так как имеем дело исключительно с замкнутыми термами. Поэтому следующий шаг будет состоять в переходе к чисто синтаксическим равенствам.

В. Комбинаторные равенства. Введем три комбинатора: S-арности 2, Λ-u'-арности 1. Кроме того, введем бесконечно много комбинаторов $n!$-арности 1. Положим по определению

$$\left.\begin{aligned} &\widehat{\mathrm{ev}}\,(n)\rightleftarrows n!;\\ &\widehat{\mathrm{ev}}\,(c)\rightleftarrows {}'c;\\ &\widehat{\mathrm{ev}}\,(MN)=S\left(\widehat{\mathrm{ev}}\,(M),\ \widehat{\mathrm{ev}}\,(N)\right);\\ &\widehat{\mathrm{ev}}\,(\lambda.M)=\widehat{\mathrm{ev}}\,(\Lambda)\,M.\end{aligned}\right\}\qquad(11.3)$$

Введенное определение позволяет преобразовать проведенную выше группу равенств п.В в чисто синтаксические:

$$\left.\begin{aligned} &0!\,(x.y)=y;\quad (n+1)!\,(x.y)=n!x;\\ &({}'x)\,y=x;\\ &S'\,(x,\ y)=xz\,(yz);\\ &\Lambda\,(x)\,yz=x\,(y,\ z).\end{aligned}\right\}\qquad(11.4)$$

Полученная система достаточно близка к классической комбинаторной логике. Отличие состит в том, что в нашем случае допускаются явные произведения. Так, правило для S записано в некоррированной форме, а последнее правило описывает операцию *коррирования,* т. е. замены функции двух аргументов функцией первого аргумента, которая действует на второй аргумент.

Недостатком полученной системы является бесконечное множество комбинаторов $n!$, поэтому наша задача — получить систему из конечного числа комбинаторов.

Г. Категориальные комбинаторы. Тот факт, что построенная в п.В комбинаторная логика допускает явные произведения типа (4, $(\lambda x.x)$ 3), что ведет к необходимости введения комбинатора, преобразующего пару значений в произведение.

Определение. *Спаривающим комбинатором* называется комбинатор с характеристикой

$$\langle\widehat{ev}\,(M),\ \widehat{ev}\,(N)\rangle=\widehat{ev}\,(M,\ N).$$

Смысл спаривающего комбинатора состоит в образовании из двух функций: $f:D-E$ и $g:D-F$ их сцепления $h\rightleftarrows<f,\ g>$ $D\rightarrow E\times F$, имеющего пару значений этих двух функций.

Не следует путать *пару* и *сцепление*: сцепление — функция $(D\Rightarrow E)\times(D\Rightarrow F)$, а пара — значение из $(E\times F)$! Для этого комбинатора определяются два комбинатора проекций.

Определение. *Комбинаторами проекций* называются комбинаторы с характеристиками

$$\mathrm{Fst}\,(\widehat{\mathrm{ev}}\,(M,\ N))=\widehat{\mathrm{ev}}\,(M);$$

$$\mathrm{Snd}\,(\widehat{\mathrm{ev}}\,(M,\ N))=\widehat{\mathrm{ev}}\,(N).$$

Возвращаясь к синтаксической форме, имеет три равенства:

$$\left.\begin{aligned} &\mathrm{Fst}\,(x,\ y)=x;\\ &\mathrm{Snd}\,(x,\ y)=y;\\ &\langle x,\ y\rangle\,z=(xz,\ yz).\end{aligned}\right\}\qquad(11.5)$$

Теперь, чтобы перейти от системы (11.4) к единой системе комбинаторов с учетом (11.5), введем новый, чисто категориальный комбинатор *композиции,* т. е. основной инструмент теории категорий.

Определение. *Комбинатором композиции* называется комбинатор с характеристикой

$$(\widehat{\mathrm{ev}}\,(M\circ N))_\rho=\widehat{\mathrm{ev}}\,(M)\,(\widehat{\mathrm{ev}}\,(N))_\rho.$$

Введение этого комбинатора позволяет нам выразить часть равенств системы (11.4) через комбинаторы системы (11.5).

Положим по определению в (11.4)

$$\left.\begin{aligned} &S\,(x,\ y)\rightleftarrows \mathrm{App}\circ\langle,\rangle;\\ &n!\rightleftarrows \mathrm{Snd}\circ\mathrm{Fst}^n,\end{aligned}\right\}\qquad(11.6)$$

где $\mathrm{Fst}^{n+1}\rightleftarrows\mathrm{Fst}\circ\mathrm{Fst}^n$.

Предполагается, что

$$\mathrm{App}\,(x,\ y)=xy.\qquad(11.7)$$

Теперь система (11.4) преобразуется в систему характеристик категориальных комбинаторов:

$$\left.\begin{aligned} &(\mathrm{ass})\,(x\circ y)\,z=x\,(yz);\\ &(\mathrm{dpair})\,\langle x,\ y\rangle\,z=(xz,\ yz);\\ &(\mathrm{fst})\,\mathrm{Fst}\,(x,\ y)=x;\\ &(\mathrm{snd})\,\mathrm{Snd}\,(x,\ y)=y;\\ &(\mathrm{guote})\,({}'x)\,y=x;\\ &(\mathrm{ac})\,\mathrm{App}\,(\Lambda-(x)\,y,\ z)=x\,(y,\ \ z).\end{aligned}\right\}\qquad(11.8)$$

Построенная система комбинаторов, вообще говоря, шире, чем первоначальная (11.4). Равенство (ass) устанавливает связь полторации и аппликации подобно тому, как это осуществляется (dpair) со спариванием и сцеплением. Правило для S преобразуется в $S(x.y)z=\mathrm{App}(xz,yz)$.

Заметим, что M можно рассматривать как сокращение для $\Lambda(M \circ \mathrm{Snd})$, так как

$$\Lambda(x \circ \mathrm{Snd})\, yz = (x \circ \mathrm{Snd})(y,\ z) = x\, \mathrm{Snd}\,(y,\ z) = xz$$

и

$$('x)\, yz = xz.$$

Такое преобразование можно использовать для функциональных констант типа $+$.

Здесь мы вплотную подошли к структуре абстрактной категориальной машины, так как комбинаторы построенной системы уже можно использовать в качестве инструкций такой машины, применяя замкнутые категориальные термы к пустому окружению.

Термин «категориальная» оправдан тем, что у нас используются основные понятия теории категорий, такие, как композиция и (вводимая дальше) тождественность.

Декартовы категории предполагают произведения (комбинаторы $<,>$, Fst, Snd и декартово-замкнутые категории — кодирование (Λ и App).

11.6.3. Структура абстрактной категориальной машины

А. Компоненты абстрактной категориальной машины. Покажем, что введенные выше комбинаторы, образующие систему (11.8), могут быть использованы в качестве машинных инструкций. Действительно, пусть вычисляемое выражение находится в форме M_v, где v — значение, т. е. терм, находящийся в нормальной с точки зрения правил (11.8) форме, а M — код, действующий на v. Тогда комбинаторы Fst и Snd действуют на значение $(v_1,\ v_2)$ посредством доступа к его первому или второму компоненту.

Сцепление $<M,\ N>$ действует на v как $<M,\ N>v = (M_v,\ N_v)$. Из этого следует, что в ходе вычисления $<M,\ N>v$ необходимо отдельно вычислять M_v и N_v и затем соединять их вместе. Для этого следует выбрать сначала, что вычислять первым (допустим, M_v, имеющее результатом v_1). Вместе с тем необходимо хранить и v, чтобы использовать его при выполнении N_v, результатом которого будет v_2. Остается соединить вместе v_1 и v_2, однако при этом предполагается, чтобы было сохранено и значение v_1, причем сохранить его надо сразу же после восстановления v.

Теперь мы пришли к трем компонентам, из которых должна состоять категориальная машина. Машина должна иметь потенциально бесконечный *стек* для накопления промежуточных значений вычислений, *код*, в котором будут последовательно храниться машинные инструкции (комбинаторы), и, наконец, *терм*, в котором будет храниться значение, — объект действия комбинатора. Таким образом, состояние машины в каждый момент времени описывается тройкой T, C, S.

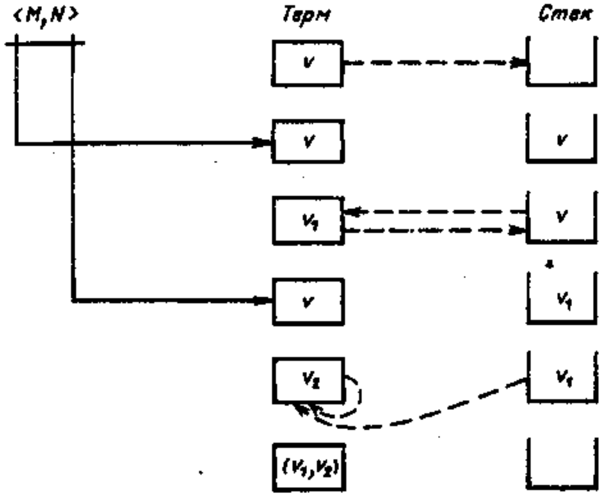


Рис. 11.1. Вычисление выражения $<M,\ N>v$: $M_v = v_1$; $N_v = v_2$.

На рис. 11.1 показано, как в такой машине может быть вычислено выражение $<M,\ N>v$.

Из рисунка видно, что перед началом вычислений необходимо продвинуть терм в вершину стека. Далее код для M применяется к терму и в него же заносится, после чего происходит обмен вершины стека и терма. Затем код для N применяется по-прежнему к выражению v, находящемуся в стеке, и в заключение создается пара из вершины стека и терма, которая помещается в терм. Коды для M и N могут, конечно, иметь сложную структуру и сами быть последовательностями кодов.

Здесь мы вплотную подошли к описанию инструкций КАМ.

Б. Набор инструкций абстрактной категориальной машины. Из изложенного выше вытекает, что эффект от действия, $<,>$ должен быть следующим:

$<$: проталкивает терм в вершину стека;

,: обменивает терм и вершину стека;

$>$: создает пару «вершины стека и терм», заменяет терм только что построенной парой и продвигает стек.

В дальнейшем последовательности вычислений в машине будем изображать в виде таблиц, характеризующих состояние трех ее компонентов:

T	C	S
v	$\langle M,\ N\rangle$	[ ]
v	$M,\ N\rangle$	v

v_1	$, N\rangle$	v
v	$N\rangle$	v_1
v_2	$\rangle$	v_1
$(v_1,\ v_2)$		$[\ ]$

Здесь $[\ ]$ обозначает пустой список, $[e_1, \ldots, e_n]$ — список из n элементов $e_1, \ldots, e_n$.

Теперь, зная структуру машины, опишем инструкции для Fst и Snd: Fst предполагает терм вида $(s,\ t)$ и заменяет его на s; Snd предполагает терм вида $(s,\ t)$ и заменяет его на t.

Код для n! составлен из n инструкций Fst, за которыми следует инструкция Snd, т. е. имеет место композиция этих комбинаторов. Для композиции нет необходимости введения специальной инструкции, так как для $y \cdot x$ будем использовать в качестве синтаксиса в выражении xy, что обеспечит применение кода y к результату действия кода x при последовательном выполнении инструкций слева направо.

Для комбинаторов коррирования Λ код $\Lambda(M)$ заменяется кодом $\Lambda(c)$, где c — код для M: Λ заменяет терм s на $C:s$, где C есть код, инкапсулированный в Λ. Отсюда видно, что Λ не производит никаких вычислений, а обеспечивает «заморозку» кода M для дальнейших вычислений.

Операция аппликации должна обеспечивать следующую последовательность действий:

$$(\mathrm{App} \cdot \langle M,\ N\rangle)\, v = \mathrm{App}\,(\langle M,\ N\rangle)\, v = \mathrm{App}\,(M_v,\ N_v) = M_v\,(N_v).$$

Так, пусть $(M_v,\ N_v) \rightleftarrows (v_1,\ v_2)$. Инструкция App предполагает, что $v_1 = \Lambda(p)\, v_1'$, и выполняет

$$\mathrm{App}\,(\Lambda\,(p)\, v_1',\ v_2) = p\,(v_1',\ v_2).$$

Имеем: App предполагает терм $(C:s,\ t)$ и заменяет его на $(s,\ t)$ и предваряет остаток кода символом e. Тем самым инструкция App обеспечивает «разморозку» кода, «замороженного» инструкцией Λ, и вычисление выражения, задаваемого кодом C. Обработку комбинатора′ будем производить в зависимости от вида «цитируемой» константы. Для базовых понятий, таких, как целые числа, знак = заменяет терм на инконсультированную константу. Для остальных случаев, к которым относятся функциональные константы, подобные +, применим кодирование: код для $'e$ есть $\Lambda(\mathrm{Snd}\, C)$.

В заключение объединим все инструкции в таблицу, введя новые обозначения, более точно характеризующие их поведение:

$$\mathrm{Fst} \rightleftarrows \mathrm{car}$$
$$\mathrm{Snd} \rightleftarrows \mathrm{cdr}$$
$$\langle \quad \rightleftarrows \mathrm{pust}$$
$$, \quad \rightleftarrows \mathrm{swap}$$
$$\rangle \rightleftarrows \mathrm{cons}$$
$$\mathrm{App} \rightleftarrows \mathrm{app}$$
$$\Lambda \rightleftarrows \mathrm{cur}$$
$$\rightleftarrows \quad \mathrm{guote}$$

Через aL обозначено a, предваряющее L, а через $L1@L2$ — сцепление $L2$ и $L1$.

Таблица 11.1

Исходная конфигурация			Результат		
Терм	Код	Стек	Терм	Код	Стек
$(s,\ t)$	car.C	S	s	C	S
$(s,\ t)$	cdr.C	S	t	C	S
s	(guote c).C	S	c	C	S
s	(cur C).C_1	S	$(C:s)$	$C1$	S
s	pust.C	S	s	C	s, S
t	swap.C	s, S	s	C	t, S
t	cons.C	s, S	$(s,\ t)$	C	S
$(C:s,\ t)$	app.$C1$	S	$(s,\ t)$	C @ $C1$	S

11.6.4. Некоторые подходы к оптимизации вычислений

В целом ряде случаев при вычислении выражений на абстрактной категориальной машине некоторые последовательности кодов можно свести к более коротким с тем же эффектом. Это дает возможность оптимизации кодов до начала вычислений.

А. Оптимизация с использованием β-редукции. Код, соответствующий выражению

$$\Lambda\,(M)\, N = \mathrm{App} \circ \langle \Lambda\,(M),\ N\rangle = \langle \Lambda\,(M),\ N\rangle\, \mathrm{App},$$

есть push, за которым следует cur C (C — код для M) за ним следует swap, за ним — $C1$ — код для N и далее cons и app. В предположении, что применение кода $C1$ к терму s есть v, имеем следующую последовательность вычислений:

s	push. (cur C).swap.C1● [cons; app]	S
s	(cur C).swap.C1● [cons; app]	$s.S$
$C{:}s$	swap.C1● [cons; app]	$s.S$
s	C1● [cons; app]	$(C{:}s).S$
v	[cons; app]	$(C{:}s).S$
$C{:}s{:}v$	[app]	S
$(s.v)$	C	S

Тот же результат можно получить, оптимизировав код push. (cur C). swap. C1● [cons : app] как push. C1● cons. C. Действительно,

s	push. $C1$● cons. C	S
s	$C1$● cons.C	$s.S$
v	cons.C	$s.S$
$(s,\ v)$	C	S

Чтобы обосновать правомерность проведения такой оптимизации, придется ввести комбинатор *тождественности* из теории категорий. Основой оптимизации является правило

$$\text{(Beta) App} \circ \langle \Lambda (x),\ y\rangle = x \circ \langle \text{Id} y\rangle.$$

Тогда код для $(\lambda x.M)N$ будет:

$$\text{push}\cdot\text{ship}\cdot\text{swap}● C1 ●\,\text{cons}\cdot C,$$

где ship используется в качестве Id и выступает в роли «пустой» инструкции. Однако, конечно, эффект от [push; ship; swap] менее, чем эффект от [push].

Б. Оптимизация выражений, содержащих функции. Для функциональных понятий типа + мы ввели выражение $\Lambda(+\ \circ \text{Snd})$. Если код $M+N$ будет иметь вид

$$\text{App} \circ \langle \Lambda(+\ \circ \text{Snd}),\ \ \langle M,\ N\rangle\rangle = (+\ \circ \text{Snd}) \circ \langle \text{Id},\ \ \langle M,\ N\rangle\rangle =$$
$$= +\ \circ \langle M,\ N\rangle = \langle M,\ N\rangle + .$$

Здесь мы воспользовались оптимизацией из п. А. Тем самым код для $M+N$ следует компилировать в код для $\langle M,\ N\rangle$, за которым следует +. Кроме того, последовательность [cons; plus] может быть сжата в [plus], если предполагать, что plus — настоящая бинарная функция, использующая в качестве аргументов терм и вершину стека, а не терм, в котором находится пара аргументов.

11.6.5. Расширенный набор инструкций

А. Вычисление выражений, содержащих рекурсию. Начнем обсуждение с примера вычисления факториала — наиболее характерной функции, содержащей рекурсию:

$$\text{letrec fact } n = \text{if}\quad n = 0 \text{ then } 1$$
$$\text{else}\quad n * \text{fact}\,(n-1) \text{ infact } 1.$$

Запишем эту функцию в виде λ-выражения, использовав комбинатор неподвижной точки:

$$\text{(fix)} = YM = M\,(YM);$$
$$\text{fact} = (\lambda g.g1)\,(Y\,(\lambda fn \text{ if } n = 0 \text{ then } 1 \text{ else } n * f\,(n-1))).$$

Из этого примера видно, что для вычисления подобных выражений необходимо определять кодирование условного оператора if M then N else P. Определим для него соответствующую последовательность инструкций как push, за которым следует код для M, за которым следует branch ($C1$, $C2$), где $C1$, $C2$ — коды для N и P. Инструкция branch дает следующий эффект: branch в зависимости от значения терма true или false выполняет соответственно $C1$ или $C2$. Правило (fix) для комбинатора неподвижной точки будет иметь вид

$$YM = M\,(YM) = \text{App}\,(M,\ \ YM) = \text{App} \circ \langle M,\ YM\rangle.$$

Продолжим рассмотрение примера. В нашем случае M имеет вид ($M = \lambda\lambda.\,P$), при этом

$$YM = M\,(YM) = \lambda\lambda.P\,(YM) = \text{App}\,(\lambda\lambda.P,\ YM) =$$
$$= \text{App} \circ \langle \Lambda\,(\Lambda\,(P)),\ YM\rangle = \Lambda\,(P) \circ \langle \text{Id},\ YM\rangle.$$

На последнем шаге мы использовали оптимизацию по правилу (Beta).

Из приведенных преобразований видно, что код $C1$ для Y оказывает на произвольный терм s такой же эффект, как

$$[\text{puch};\ \ C1;\ \ \text{cons};\ \ (\text{cur}\,C)],$$

где C — код для P. Очевидно, что если результат действия C_1 (кода для YM) на s есть t, то результат от [push; $C1$; cons; (cur C)] есть $C:(s,\ t)$. Ясно, что значение терма t должно быть таково, что $t = C:(s,\ t)$, и представляется в виде графа, изображенного на рис. 11.2, *а*.

Отсюда очевидным образом вытекает решение: в качестве кода для $Y(\lambda\lambda.\ P)$ следует использовать (Rec cur C), где C — код для P, действие которого на терм s будет состоять в замене этого терма графом, как это показано на рисунке стрелкой.

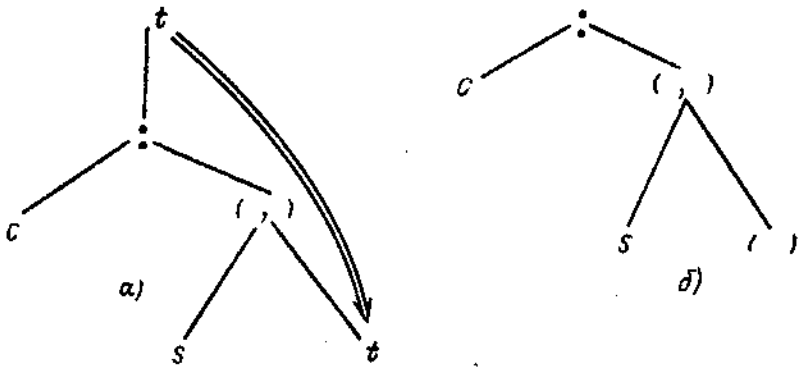


Рис. 11.2. Пояснение к вычислению, содержащему рекурсию

Действие (Rec cur C) можно декомпозировать на два этапа: построение объекта (рис. 11.2, *б*), где ( ) — фиктивное значение, и замена правого элемента пары на целый объект.

Изложенное приводит к необходимости ввести новую инструкцию mkloop: заменяет правую часть сцепления (адрес которого находится в верхнем элементе стека) термом и удаляет этот верхний элемент из стека. В момент начала выполнения этой инструкции вершина стека содержит указатель на окружение, первый элемент которого (доступный по cdr), является фиктивным значением, а терм содержит связи с этим окружением так, что выполнение mkloop имеет в качестве побочного эффекта занесение его в правую циклическую структуру.

Покажем, как выполняется код для $Y(\lambda\lambda.\ P)$, который запишется так (C — код для P):

s	[push; guote ( ); cons; push] @ (cur *C*1).[mkloop]	*S*
s	[guote ( ); cons; push] @ (cur *C*1). [mkloop]	*s*, *S*
( )	[cons; puch] @ (cur *C1*). [mkloop]	*s*, *S*
(*s*, ( ))	[push] @ (cur *C1*). [mkloop]	*S*
(*s*, *l*)	(cur *C1*). [mkloop]	(*s*, ( )).*S*
C1:(*s*, *l*)	[mkloop]	(*s*, ( )).*S*
t	[ ]	*S*

Подведем итоги. К первоначальному набору инструкций добавим еще три инструкции КАМ для обработки условных операторов и рекурсии (табл. 11.2).

Таблица 11.2

Исходная конфигурация			Результат		
Терм	Код	Стек	Терм	Код	Стек
true	(branch (l_1, l_2)).*l*	*s*, *S*	*s*	*C1* @ *C*	*S*
false	(branch (l_1, l_2)).*l*	*s*, *S*	*s*	*C2* @ *C*	*S*
s	mkloop.*l*	(*s1*, *s2*).*S*	*s′*	*C*	*S*

В последнем случае имеется в виду, что *s* содержит указатели на *s1* и *s2* как на подтермы и что *s2* заменяется на *s*, в результате чего *s* превращается в *s′*.

Б. «Заморозка» вычислений. Все рассмотренные до сих пор методики вычислений относятся к категории так называемой «активной оценки» (eger evaluation) вычисляемых выражений, которая состоит в том, что аргументы как функций, так и сцеплений вычисляются всякий раз полностью. Однако в целом ряде случаев подобные вычисления оказываются совершенно бесполезными, так как не вносят вклад в результат оценки всего выражения в целом.

Избежать бесполезных вычислений позволяет механизм «ленивой оценки» (lazy evaluation), кроме того, этот механизм позволяет оперировать с бесконечными структурами. Суть его состоит в том, чтобы ввести явную задержку, «замораживание» выполнения инструкции, отложить его. В этом смысле действие инструкции замораживания **freeze** сходно с действием **cur**, **freeze**: заменяет терм *s* структурой (*c*. *s*), где *c* — код, инкапсулированный в **freeze**, Созданную и помещенную в терм структуру (*c*. *s*) будем называть **laze**. При введении **laze** изменяется природа терма: не создается новых значений, но возможно введение неоцениваемых выражений.

В ходе вычисления выражения может возникнуть ситуация, когда инструкции **car**, **cdr**, **app** и **plus** выполнены и уже нельзя ожидать, что их аргументом будет сцепление, а термом **laze**. В этом случае надо обеспечить оценку **laze**. Для этого предназначен код «разморозки», действие которого противоположно действию **freeze**: **unfreeze**: не выполняет никаких действий до тех пор, пока терм не станет (*c*, *s*), в этом случае перед кодом (включающим **unfreeze**) ставится *c* и терм становится равным *s*.

Обобщим вновь введенные инструкции, обеспечивающие механизм «ленивой оценки», в виде табл. 11.3.

Третье правило применяется в случае, когда второе не может быть применено.

В ходе реализации КАМ имеются две возможности:

Таблица 11.3

Исходная конфигурация			Результат		
Терм	Код	Стек	Терм	Код	Стек
s	(freeze C).Cl	S	C, s	Cl	S
$c.s$	unfreeze.Cl	S	s	c @ unfreeze.Cl	S
s	unfreeze.C	S	s	C	S

1) неявно обеспечить выполнение механизма «ленивой оценки» при композиции выражений в машинные инструкции. Например, можно возложить на компилятор в выражении App ∘ $\langle M, N\rangle$ функцию «замораживать» коды для M и N с помощью инструкции;

2) явно специфицировать механизм «ленивой оценки» в языке более высокого уровня путем введения примитива **freeze**. Мы следуем этому варианту.

И в том и в другом случае компилятору поручается вводить инструкции unfreeze там, где это необходимо.

Рассмотрим небольшой пример. Необходимо вычислить

$$\text{let } z = 2 \text{ in } (\lambda x.z)\ (\text{freeze } (\lambda y.y)\ 1).$$

Обозначив $f \rightleftarrows$ freeze, $B \rightleftarrows (\lambda y.\, y)\ 1)$, вычислим:

$(\ )$	$\langle '2\rangle\ \ \langle\Lambda(FS)\, f(B)\rangle A$	$[\ ]$
$((\ ),\ 2)$	$\langle\Lambda(FS),\ f(B)\rangle A$	$[\ ]$
$((\ ),\ 2)$	$f(B)\rangle A$	$[FS{:}((\ ),\ 2)]$
$B((\ ),\ 2)$	$\rangle A$	$[FS{:}((\ ),\ 2)]$
$(((\ ),\ 2),\ \ B((\ ),\ 2))$	FS	$[\ ]$
2	$[\ \ ]$	$[\ ]$

Мы избежали бесполезного, как и ожидалось, выполнения выражения для B.

Рассмотрим, как может быть использована «ленивая оценка» для обработки рекурсивно определенных списков. Вычислим **letrec** $x = (1,$ **freeze** $x)$ **in fst** (**snd**(x)).

Обозначим **u** $\rightleftarrows$ **unfreeze**, $F \rightleftarrows$ **fst**, $s \rightleftarrows$ **snd**, $m \rightleftarrows$ **mkloop**. Компилятор транслирует **fst** (M) в код для M, за которым следуют **unfreeze** и **cur**:

$(\ )$	$\mid\langle '(\ )\rangle\langle\langle '1,\ f(\ )\rangle m\ \text{и}\ s\ \text{и}\ F$	$[\ ]$
$((\ ),\ (\ ))$	$\langle '1,\ f(S)\rangle m\ \text{и}\ s\ \text{и}\ F$	$[((\ ),\ (\ ))]$
$((\ ),\ (\ ))$	$f(S)\rangle m\ \text{и}\ s\ \text{и}\ F$	$[1;\ ((\ ),\ (\ ))]$
$S.((\ ),\ (\ ))$	$\rangle m\ \text{и}\ s\ \text{и}\ F$	$[1;\ ((\ ),\ (\ ))]$
$(1,\ (s.((\ ),\ (\ ))))$	$m\ \text{и}\ s\ \text{и}\ F$	$[((\ ),\ (\ ))]$
$s = (1,\ (s.((\ ),\ s)))$	$\text{и}\ s\ \text{и}\ F$	$[\ ]$
$s = (1,\ (s.((\ ),\ s)))$	$s\ \text{и}\ F$	$[\ ]$
$t = (s.((\ ),\ (1,\ t)))$	$\text{и}\ F$	$[\ ]$
$((\ ),\ s) = (1,\ (s.((\ ),\ s)))$	$s\ \text{и}\ F$	$[\ ]$
$s = (1,(s.((\ ),.\ s)))$	$\text{и}\ F$	$[\ ]$
$s = (1,(s.((\ ),\ s)))$	F	$[\ ]$
1	$[\ \ ]$	$[\ ]$

11.6.6. Вопросы реализации и эффективности

Сделаем несколько заключительных замечаний. Абстрактная категориальная машина может быть использована, например, для поддержки функционального языка программирования. На протяжении данной главы мы частично использовали синтаксис ML для записи λ-выражений. Ниже приводится краткая спецификация подмножества ML. В этом подмножестве мы пренебрегли описанием ссылок, присваиваний и исключительных ситуаций, а также декларациями конкретных и абстрактных типов.

```
rectype ML =
mplus| . . . |mequal||mefst|mesmd
|mlint of int
|mlbool of bool
|mlvar of striny
|mlcond of ML # ML # ML
|mlpair of ML # ML
|ml in of MLD # ML
|mlabstr of MLPAT # ML
|ml app of ML # ML
and MLD =
mlet of MLPAT # ML
|mlletrec of MLPAT # ML
and MLPAT =
nullpat
|varpat of string
|pairpat of MLPAT # MLPAT
```

Три объявленных типа *ML*, *MLD* и *MLPAT* соответствуют выражениям *ML*, декларациям *ML* и образцам *ML*. Далее приводится описание инструкций и значений *ML*:

```
rctrype instruction = plus | . . . |eq
    and value = nullvalue
              |inf of int
              |bool of bool
              |pair of value # value
```

В литературе приводится текст компилятора с языка *ML* в коды КАМ, написанный также на *ML* [40]. Программы на *ML* и программы КАМ преобразуются при этом в объекты *ML*, а процесс компиляции преобразуется в функцию *ML*.

КАМ описывается как функция Exec типа confry→confry, где тип confry есть совокупность состояний кода и стека. Эта функция задает все возможные изменения конфигураций в результате выполнения инструкций.

Заметим также, что в *ML* используются имена переменных, а не кодирование Де Брюйна, поэтому компилирующая функция транслирует имена переменных в некоторый код доступа, по которому во время вычислений ее значение можно отыскать в окружении.

Сделаем несколько замечаний об эффективности реализации КАМ. Может показаться, что не слишком эффективно представлять окружение в виде связанных списков, так как с точки зрения скорости доступа для этой цели более подошли бы массивы. Однако это не так. Во-первых, в *ML* допускается обработка списков путем прямого доступа к их значениям. Во-вторых, окружения содержат только свободные переменные, и длина их очень ограничена: эту длину еще можно уменьшить путем некоторой оптимизации. Наконец, окружения в виде связанных списков можно разделить и строить замыкания, а это очень важно при использовании механизма «ленивой оценки».

Более существенным представляется следующее. Правило аппликации предполагает обязательное сцепление аргумента с окружением. Но даже для сравнительно простых функций это приводит к большой потере времени на размещение и затем сборку мусора. Представляется разумным, чтобы избежать этого, ввести специальный регистр для хранения головы окружения.

С точки зрения эффективности очень важен набор инструкций, обеспечивающих полную мощность β-редукции; это делает возможным оптимизацию времени компиляции так, как это было показано в п. 11.6.4. Классические комбинаторы испытывают недостаток в таких правилах.

Вызывает большой интерес также возможность аппаратной реализации описанной в этой главе машины.

Глава 12

ТЕНЗОРНАЯ АЛГЕБРА И ИСЧИСЛЕНИЕ

12.1. ВВЕДЕНИЕ В ТЕНЗОРНУЮ АЛГЕБРУ И ИСЧИСЛЕНИЕ

Как уже отмечалось в гл. 1, тензорный аппарат предлагается как основной для математического образования инженера-математика или специалиста по прикладной математике. В Японии в свое время этим активно занималась группа проф. Кондо, в СССР над этой темой работали П. Г. Кузнецов и Л. Т. Кузин.

Более ранние работы в этом направлении связаны с именами Крона и математика Веблена [83—91].

В значительной степени тензорная идеология начинает немного устаревать и перекрываться другими разделами математики, в частности теорией категорий. С позиции теории категорий тензоры — это определенный тип категории Ten. Однако тензорный аппарат развивался в течение многих десятилетий и является определенным «сплавом» из многих разделов математики, представляющим интерес для инженера любого профиля. Остановимся на некоторых важных для кибернетика свойствах тензоров.

Существует очень много профессиональных точек зрения и определений тензоров, некоторые будут рассмотрены ниже. Для нас важны следующие точки зрения на понятие тензора:

тензор как «экономное» средство описания систем большой размерности;

тензор как алгебраический объект;

тензор как преобразователь;

тензор как геометрический объект;

тензор как физический объект.

Первое свойство поясняется следующим образом. В 60-х годах в теории систем автоматического регулирования стала популярной векторная запись системы алгебраических и дифференциальных уравнений. Благодаря этому вместо систем дифференциальных уравнений писалось одно векторное дифференциальное уравнение. Далее, несколько векторных уравнений описываются одним тензорным уравнением и т. д. Налицо и более экономичный способ записи математических выражений, освобождение исследователя от рутинных выкладок и предоставление ему больше времени для истинно научных исследований. Но это только первое, наиболее видимое преимущество тензорного представления.

Главное преимущество, конечно, заключается в более серьезных вопросах, а именно в новых способах и средствах описания семантических, смысловых и интеллектуальных свойств кибернетических систем. Количество как бы переходит в качество, так как объединение нескольких векторных уравнений в одно тензорное происходит «немеханически», синтаксически, а с учетом родственных, семантических свойств близости векторных уравнений. Так, к примеру, для электромеханических систем, которым мы обязаны становлением инженерных тензорных методов с помощью работ Крона, одно тензорное уравнение объединяет все двигатели постоянного тока, а для генераторов переменного тока — другое тензорное уравнение и т. д. Именно это свойство тензоров позволило Крону создать модель обобщенной электрической машины, из которой как частные случаи следовали описания любых известных и даже новых, неизвестных электромеханических систем, а в дальнейшем вообще любых кибернетических систем. В основу своей общей теории относительности Эйнштейн положил тензорные уравнения, которые приведем ниже.

Он объединил ряд уравнений механики, электромагнитных явлений, тяготения, геометрии и т. д. и получил закон, лежащий в основе новой (релятивистской) теории тяготения.

В руководствах по математике тензор известен как объект, используемый в преобразованиях одних алгебраических систем в другие. Вот здесь снова появляется свойство, очень важное для кибернетики и аналогичное таким же свойствам других математических разделов,— это процедуральность, динамика, способность описывать движение. Тензор обладает свойством дуальности, т. е. это и статический, и динамический объект одновременно.

Далее, тензор — это геометрический объект, такой же, как вектор, скаляр и т. д. Эта точка зрения, которой мы обязаны, главным образом, Крону и Веблену, как бы открывает еще одну дверь кибернетику-инженеру в бесконечный мир геометрии. Тензор как геометрический объект обладает достаточно специфическими, может быть, уникальными свойствами.

Тензор существует как объект, инвариант в любом геометрическом пространстве, но реализуется он в виде своих компонентов.

Для сравнения тензорного и матричного описаний кибернетических систем можно привести пример Крона с объемным изображением физического объекта — скульптуры человека, животного и т. д. и фотографиями этой скульптуры. Скульптуру (объемное изображение) как бы описывает тензор, а его фотографии (плоские изображения) описывают матрицы. Точка зрения, что физические объекты описываются тензорами, берет свое начало в теории упругости, гидродинамике и аэродинамике, откуда и произошло название *тензор* (давление, напряжение). Оказывается, если взять маленькую площадку в сплошной среде, то давление на эту площадку носит тензорный характер и не описывается вектором. Подчеркиваем, что дело не в математике, а в физике явлений, т. е. явление, которое мы изучаем,— давление на элементарную (малую) площадку — носит тензорный характер. И таких физических, геометрических и кибернетических явлений можно привести очень много. Для их описания применяется математический формальный аппарат, называемый *тензорной алгеброй и исчислением.*

Важность физической сущности тензора не позволяет освободиться от электрической интерпретации (токов и напряжений), которую использует Крон. Несколько попыток в этом направлении делалось, но при этом теряется очень много ценных физико-математических свойств тензора. И наконец, глобальное значение тензорного подхода к исследованию всех физических явлений природы связано с тенденцией геометризации физических дисциплин, когда физические законы следуют из геометрических (примат геометрии перед физикой). Это направление исторически началось в основном с работы Н. Лобачевского, потом было развито в общей теории относительности А. Эйнштейна и завершилось тензорным уравнением, описывающим структуру мира. Для нас важно, что тензорный аппарат может быть использован для описания семантики процессов управления.

12.2. ОПРЕДЕЛЕНИЕ ТЕНЗОРА

Прежде чем излагать основы тензорного исчисления, рассмотрим на примере двухмерного случая свойства контравариантных и ковариантных векторов.

Пример 12.1. В двухмерном векторном пространстве контравариантный вектор (рис. 12.1) разлагается по координатным векторам $\bar{e}_1$ и $\bar{e}_2$ по формуле

$$\bar{a} = a^1\bar{e}_1 + a^2\bar{e}_2 = a^\alpha e_\alpha. \tag{12.1}$$

Переход от одной аффинной системы координат $(o, \bar{e}_1, \bar{e}_2)$ к другой $(o', \bar{e}_{1'}, \bar{e}_{2'})$ определяется вектором $oo'=h$ и формулами преобразований к новым базисным векторам от старых:

$$\left.\begin{aligned} \bar{e}_{1'} &= p^1_{1'}\bar{e}_1 + p^2_{1'}\bar{e}_2; \\ \bar{e}_{2'} &= p^1_{2'}\bar{e}_1 + p^2_{2'}\bar{e}_2, \end{aligned}\right\} \tag{12.2}$$

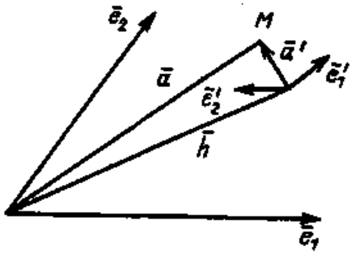


Рис. 12.1. Двухмерное пространство

или

$$\bar{e}_{i'} = p^i_{i'}\bar{e}_i; \qquad i=1,\ 2; \quad i'=1',\ 2'. \tag{12.3}$$

Из рис. 12.1 имеем, что

$$\bar{a} = \bar{a}' + \bar{h}. \tag{12.4}$$

Координаты точки M по отношению к старой системе (x^1, x^2) связаны с координатами той же точки в новой системе $(x^{1'}, x^{2'})$ формулами

$$x^i\bar{e}_i = x^{i'}\bar{e}_{i'} + \bar{h}. \tag{12.5}$$

Разлагая вектор $\bar{h}$ по векторам $\bar{e}_1$ и $\bar{e}_2$ и заменяя по (12.2), получаем

$$x^i\bar{e}_i = x^{i'}p^i_{i'}\bar{e}_i + h^i\bar{e}_i. \tag{12.6}$$

Приравнивая коэффициенты при линейно независимых векторах $\bar{e}_1$ и $\bar{e}_2$ в (12.6), получаем

$$x^i = x^{i'}p^i_{i'} + h^i, \quad i=1,\ 2; \quad i'=1',\ 2'. \tag{12.7}$$

Итак, при переходе от старой системы координат к новой координаты любого вектора преобразуются по правилам

$$a^i = p^i_{i'}a^{i'} \qquad i=1,\ 2; \quad i'=1',\ 2', \tag{12.8}$$

где a^1, a^2 — старые координаты вектора $\bar{a}$; $a^{1'}$, $a^{2'}$ — новые координаты вектора $\bar{a}$.

Заметим, что формулы (12.2) выражают новые базисные векторы через старые, а формулы (12.8) — старые координаты вектора через новые значения. Кроме того, суммирование в (12.2) происходит по верхнему индексу, соответствующему строке матрицы преобразований

$$\mathbf{P} = \begin{pmatrix} p^1_{1'} & p^2_{1'} \\ p^1_{2'} & p^2_{2'} \end{pmatrix}, \tag{12.9}$$

а в формуле (12.8) — по нижнему индексу, соответствующему столбцу. Для получения выражения, соответствующего (12.8), для новых координат через старые замечаем, что так как векторы $\bar{e}_1$ и $\bar{e}_2$ линейно независимы, то детерминант

$$\Delta = \begin{vmatrix} p^1_{1'} & p^2_{1'} \\ p^1_{2'} & p^2_{2'} \end{vmatrix} \neq 0. \tag{12.10}$$

Решив уравнения (12.8) относительно $a^{1'}$ и $a^{2'}$, получим

$$a^{i'} = q^{i'}_i a^i, \qquad i=1,\ 2; \quad i'=1',\ 2', \tag{12.11}$$

где

$$\left.\begin{aligned} q^{1'}_1 &= \frac{p^2_{2'}}{\Delta}; \qquad q^{1'}_2 = \frac{p^1_{2'}}{\Delta_1}; \\ q^{2'}_1 &= \frac{p^2_{1'}}{\Delta}; \qquad q^{2'}_2 = \frac{p^1_{1'}}{\Delta}. \end{aligned}\right\} \tag{12.12}$$

В этих выражениях $p^i_{i'}$ — приведенные миноры детерминанта Δ, задаваемого формулой (12.10), т. е. частное от деления адъюнкты элемента детерминанта на значение детерминанта. Вектор $\bar{a}=\{a^1, a^2\}$, который преобразуется по стандартным законам, называется в аффинном пространстве контравариантным вектором, а вектор, который изменяется как базисный, ковариантным [см. (12.2)]. Важно подчеркнуть, что матрицы преобразований **P** контравариантного вектора связаны с матрицей преобразования **Q** ковариантного вектора как прямая и обратная матрицы, т. е. в соответствии с (12.12). Кроме того, нетрудно заметить, что при преобразовании контравариантного вектора он умножается справа на прямую матрицу преобразований, а соответствующий ковариантный вектор при преобразовании умножается слева на обратную матрицу преобразований. Поэтому, если векторы $\bar{a}$, $\bar{u}$ принадлежат одному пространству, то существует их скалярное произведение a^iu_i, которое не изменяется при преобразованиях координат, т. е. является инвариантом.

Первое определение тензора

Тензорное исчисление относится к понятию преобразующих аппаратов математики, приводящих к геометризации физических процессов, что помогает математической обработке их. Тензор органически связан с понятием пространства и его преобразованиями.

Прежде всего, вводится понятие аффинного пространства, которое состоит из множества точек, и пространства векторов, соединяющих эти точки (с подходящей аксиоматикой). Геометрическими объектами в этом пространстве называются объекты, инвариантные относительно группы аффинных преобразований. Точкам аффинного пространства $A_n = \{M_i\}$ можно поставить во взаимно однозначное соответствие систему n чисел $X^1, X^2, \ldots, X^n$. Эта система чисел сопоставляема с данной точкой и называется системой координат этой точки. Система координат X^i называется аффинной, если она связана с другой системой координат $X^{i'}$ преобразованиями вида

$$x^i = p^i_{i'} x^{i'} + h^i, \qquad i = 1, 2, \ldots, n; \qquad i' = 1', 2', \ldots, n'. \tag{12.13}$$

Считается, что суммирование происходит по повторяющемуся индексу. Преобразование, задаваемое формулой (12.13), называется *аффинным преобразованием.*

Контравариантный тензор определяется совокупностями чисел, которые преобразуются по закону

$$X^i = p^i_\alpha X^{*\alpha}, \qquad i, \ \alpha = 1, \ 2, \ldots, \ n. \tag{12.14}$$

Считается, что детерминант преобразования отличен от нуля:

$$\Delta = |p^i_\alpha| \neq 0. \tag{12.15}$$

Решая уравнения (12.14) относительно $X^{*\alpha}$ получаем

$$X^{*\alpha} = q^\alpha_i X^i, \qquad i = 1, 2, \ldots, n.$$

Здесь q^α_i — приведенный минор элемента p^i_α в детерминанте (12.15), так что

$$P^i_\beta q^\alpha_i = \delta^\alpha_\beta,$$

где

$$\delta^\alpha_\beta = \begin{cases} 1, & \text{если} \quad \alpha = \beta; \\ 0, & \text{если} \quad \alpha \neq \beta, \end{cases}$$

— дельта-функция Кронекера.

Очевидно, что $|p^\alpha_\beta| \, |q^\alpha_i| = 1$ и $|q^\alpha_i| = \Delta^{-1} \neq 0$.

Другим геометрическим объектом в аффинном пространстве является ковариантный вектор, координаты которого u_i преобразуются по закону

$$\overset{*}{u}_\alpha = p^i_\alpha u_i. \tag{12.16}$$

Разрешив уравнения (12.16) относительно u_i, получим

$$u_i = q^\alpha_i \overset{*}{u}_\alpha,$$

где $q^\alpha_i = p^i_\alpha / \Delta$ — приведенные миноры детерминанта (12.15), причем

$$q^\alpha_j p^i_\alpha = \frac{\sum\limits_{\alpha=1}^{n} q^\alpha_j p^i_\alpha}{\Delta} = \delta^i_j \quad \text{и} \quad q^\alpha_j \overset{*}{u}_\alpha = q^\alpha_j p^i_\alpha u_i = \delta^i_j u_i.$$

Ковариантный тензор валентности r определяется заданием чисел $a_{i_1 i_2 \ldots i_r}$ которые преобразуются по закону

$$\overset{*}{a}_{\alpha_1 \alpha_2 \ldots \alpha_r} = P^{i_1}_{\alpha_1} P^{i_2}_{\alpha_2} \ldots P^{i_r}_{\alpha_r} a_{i_1 i_2 \ldots i_r}. \tag{12.17}$$

В силу того что каждый из индексов $i_1, i_2, \ldots, i_n$ пробегает независимо друг от друга значения $1, 2, \ldots, n$, число возможных значений ковариантного тензора валентности r равно n^r.

Контравариантный тензор валентности r определяют такие n^r чисел $a^{i_1 i_2 \ldots i_r}$, которые изменяются при преобразовании пространства по закону

$$a^{*\alpha_1 \alpha_2 \ldots \alpha_r} = q^{\alpha_1}_{i_1} q^{\alpha_2}_{i_2} \ldots q^{\alpha_r}_{i_r} a^{i_1 i_2 \ldots i_r} \tag{12.18}$$

или

$$a^{i_1 i_2 \ldots i_r} = P^{i_1}_{\alpha_1} P^{i_2}_{\alpha_2} \ldots P^{i_r}_{\alpha_r} a^{*\alpha_1 \alpha_2 \ldots \alpha_r}. \tag{12.19}$$

Тензор, определяемый тензорным преобразованием, тесно связан с аффинным пространством, так как по существу тензорные преобразования совпадают с аффинными с точностью до постоянных коэффициентов h^i [см. (12.13)].

Вводятся в рассмотрение смешанные тензоры, например смешанный тензор ковалентности 3 и контравалентности 2 (общей валентности 5) определяется как совокупность n^5 чисел, которая преобразуется по закону

$$C^{*\lambda\mu}_{\alpha\beta\gamma} = q^\lambda_e q^\mu_m p^i_\alpha P^j_\beta P^k_\gamma C^{em}_{ijk}. \tag{12.20}$$

Ковалентность таких тензоров равна числу нижних индексов, контравалентность — числу верхних индексов.

ТЕНЗОРЫ В ЕВКЛИДОВЫХ ПРОСТРАНСТВАХ

Тензорный аппарат тесно связан с математическим анализом и дифференциальной геометрией. Считается, что любое преобразование локально линейно. Поэтому преобразование координат

$$\overline{X}^i = f^i(X^1, X^2, \ldots, X^n) \tag{12.21}$$

переводит координаты двух близких точек $(X^1, X^2, \ldots, X^n)$ и $(X^1+dX^1, X^2+dX^2, \ldots, X^n+dX^n)$ в две близкие точки $(\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^n)$ и $(\bar{x}^1+d\bar{x}^1, \ldots, \bar{x}^n+d\bar{x}^n)$, причем

$$d\bar{x}^i = \frac{\partial \bar{x}^i}{\partial x^j} dx^j. \tag{12.22}$$

Объект, заданный в векторной системе координат своими дифференциалами $d\bar{x} = (dx^1, dx^2, \ldots, dx^n)$ и системой преобразований типа (12.22), при переходе к другой системе координат является контравариантным вектором. В дифференциальной геометрии вводится в рассмотрение ковариантный вектор, определяемый как градиент скалярной функции $A(x)$. Частные производные этой функции равны:

$$A_i = \frac{dA}{\partial x^i}.$$

С переходом к другой системе координат, связанной с исходной соотношениями

$$x^i = q^i(\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^n), \tag{12.23}$$

функция $A(\bar{X})$ преобразуется в функцию

$$\bar{A}(\bar{X}) = A(q^1(\bar{X}), q^2(\bar{X}), \ldots, q^n(\bar{X})). \tag{12.24}$$

Дифференцируя эти формулы, получаем

$$\frac{\partial A}{\partial \bar{x}^i} = \frac{\partial A}{\partial x^\alpha} \frac{\partial x^\alpha}{\partial \bar{x}^i}. \tag{12.25}$$

Введем в рассмотрение функции

$$\bar{A}_i(X) = \frac{\partial \bar{A}}{\partial \bar{X}^i}, \tag{12.26}$$

которые связаны соотношениями

$$\bar{A}_i = A_1 \frac{\partial x^1}{\partial \bar{x}^i} + A_2 \frac{\partial x^2}{\partial \bar{x}^i} + \ldots + A_n \frac{\partial x^n}{\partial \bar{x}^i}. \tag{12.27}$$

Это преобразование определяет систему функций $(A_1, A_2, \ldots, A_n)$, являющихся градиентом скаляра $A(x)$ как ковариантного вектора. Частным случаем аффинного пространства является евклидово пространство, которое определяется как такое аффинное пространство, в котором определена евклидова метрика заданием метрического тензора g_{ij} или евклидова скалярного произведения. Компоненты метрического тензора представляют собой попарные скалярные произведения единичных попарно ортогональных базисных векторов $\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n$. Система $\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n$ таких векторов называется *репером*. Допустим, в некоторой системе аффинных координат выбрано $\frac{1}{2} n(n+1)$ чисел g_{ij}, таких, что $g_{ij} = g_{ji}$ $(i, j = 1, 2, \ldots, n)$ и $|g_{ij}| \neq 0$, $g_{ij}a^i a^j > 0$, для любой системы чисел $a^1, a^2, \ldots, a^n$, из которых хотя бы одно отлично от нуля.

Преобразование объекта, описываемого g_{ij}, задается выражением

$$g_{i'j'} = p^i_{i'} p^j_{j'} g_{ij}. \tag{12.28}$$

Геометрический объект g_{ij}, заданный в любой системе координат, называется *метрическим тензором евклидова n-мерного пространства.*

Вводится в рассмотрение группа аффинных преобразований, которая оставляет метрический тензор

$$g_{\alpha\beta} = p^i_\alpha p^j_\beta g_{ij} \tag{12.29}$$

неизменным.

Матрица преобразований, удовлетворяющая условию (12.29), имеет детерминант, равный единице со знаком плюс или минус, и определяет так называемую группу движений и отражений в n-мерном пространстве. Движение, перемещение имеет смысл в пространстве, в котором определена метрика и можно ввести понятие расстояния, длины пути. В евклидовом пространстве существует очень важный инвариант — скалярное произведение.

Введем в рассмотрение два контравариантных вектора: $\bar{a}$ и $\bar{b}$.

При некотором преобразовании из группы движений и отражений им соответствуют векторы a^* и b^*, определяемые соотношением (12.29). С помощью этой формулы можно получить соотношение

$$g_{\alpha\beta} a^{*\alpha} b^{*\beta} = p^i_\alpha a^\alpha p^j_\beta b^\beta g_{ij}, \tag{12.30}$$

или

$$g_{\alpha\beta} a^{*\alpha} b^{*\beta} = g_{ij} a^i b^j. \tag{12.31}$$

Это означает, что выражение $g_{ij}a^i b^j$ является инвариантом относительно группы движений. Этот инвариант называется скалярным произведением векторов $\bar{a}$ и $\bar{b}$. Таким образом, в евклидовом пространстве имеется скалярное произведение любых двух векторов, которое определяет расстояние, путь и т. д. В этом пространстве ортонормированный базис со своим дуальным базисом связаны единичной матрицей, и поэтому исчезает различие между контравариантными и ковариантными векторами.

Скалярное произведение вектора $\bar{a}$ самого на себя называется *скалярным квадратом*, а $\sqrt{a^2}$ — длиной вектора $\bar{a}$.

Тензор g_{ij} поэтому и называется *метрическим*, что он определяет такие понятия в евклидовом пространстве, как длина вектора, путь (длина кривой) и т. д. В евклидовом пространстве длина кривой, соединяющей две точки A и B, определяется формулой

$$\int\limits_A^B dS = \int\limits_{t_0}^{t_1} \sqrt{\left(\frac{dy^1}{dt}\right)^2 + \left(\frac{dy^2}{dt}\right)^2 + \left(\frac{dy^3}{dt}\right)^2}\, dt, \qquad (12.32)$$

где y^1, y^2, y^3 — прямоугольные декартовы координаты; кривая записывается в параметрическом виде $y^i = f^i(t)$; точке A соответствует значение параметра $t = t_0$, а точке B — значение параметра $t = t_1$.

В произвольной системе координат формула (12.32) запишется в виде

$$\int\limits_A^B dS = \int\limits_{t_0}^{t_1} \sqrt{g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt}}\, dt, \qquad (12.33)$$

где

$$g_{ij} = \frac{\partial y^1}{\partial x^i} \frac{\partial y^1}{\partial x^j} + \frac{\partial y^2}{\partial x^i} \frac{\partial y^2}{\partial x^j} + \frac{\partial y^3}{\partial x^i} \frac{\partial y^3}{\partial x^j} \qquad (12.34)$$

— метрический тензор.

При переходе к другой системе координат $\bar{X}$ величины g_{ij} в формуле (12.33) заменяются на следующие:

$$\bar{g}_{ij} = \frac{\partial y^a}{\partial \bar{x}^i} \frac{\partial y^a}{\partial \bar{x}^j} = \frac{\partial y^a}{\partial x^a} \frac{\partial x^a}{\partial \bar{x}^i} \frac{\partial y^a}{\partial x^p} \frac{\partial x^p}{\partial \bar{x}^j} = g_{pq} \frac{\partial x^p}{\partial \bar{x}^i} \frac{\partial x^q}{\partial \bar{x}^j}.$$

Из этого соотношения следует, что величины g_{ij} являются компонентами ковариантного тензора второй валентности и коэффициентами дифференциальной квадратичной формы:

$$dS = g_{ij} dx^i dx^j = (dy^1)^2 + (dy^2)^2 + (dy^3)^2. \qquad (12.35)$$

Если ввести в рассмотрение квадратную матрицу g^{ij} с элементами

$$g^{ij} = G^{ij}/g, \qquad (12.36)$$

где G^{ij} — алгебраическое дополнение элемента g_{ij}; $g = |g_{ij}|$ — детерминант матрицы (g_{ij}), то g^{ij} будет *контравариантным метрическим тензором*, в то время как g_{ij} называется *ковариантным метрическим тензором.*

Очевидно, что

$$g_{ij} g^{ik} = \delta_j^k. \qquad (12.37)$$

Кроме того, $g_{pj} G^{pk} = g \delta_i^k$. Полагая в этом соотношении $k = j$ и дифференцируя, получаем

$$G^{ij} = \frac{\partial g}{\partial g_{ij}}, \qquad (12.38)$$

так как G^{pf} не зависит от g_{pf}. Иначе это равенство может быть записано в виде

$$g^{ij} = \frac{1}{g} \frac{\partial g}{\partial g_{ij}}. \qquad (12.39)$$

В соотношениях производная от определителя, элементы которого являются функциями $(x^1, x^2, \ldots, x^n)$, задается формулой

$$\frac{\partial g}{dx^j} = \frac{\partial g_1^{i_1}}{\partial x^j} G_{i_1}^1 + \frac{\partial g_2^{i_2}}{\partial x^j} G_{i_2}^2 + \ \ldots\ = \frac{\partial g_c^j}{\partial x^j} G_j^c, \qquad (12.40)$$

где G_f^c — алгебраическое дополнение элемента g_c^j, стоящего на пересечении столбца с номером f и строки с номером c.

Нетрудно убедиться в том, что в любом евклидовом пространстве в силу условия $g_{ij} = g_{ji}$ имеем

$$|g^{ij}| = \frac{1}{|g_{ij}|} = \frac{1}{g}. \qquad (12.41)$$

Для любой прямоугольной декартовой системы координат эти два тензора имеют одинаковые компоненты:

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}. \qquad (12.42)$$

В системе сферических координат имеем

$$dS^2 = dr^2 + r^2 d\theta^2 + r^2 \cos^2 \theta d\varphi^2, \qquad (12.43)$$

и длина кривой определяется по формуле

$$\int\limits_{t_1}^{t_2} \sqrt{\left[\left(\frac{dr}{dt}\right)^2 + r^2 \left(\frac{d\theta}{dt}\right)^2 + r^2 \cos^2 \theta \left(\frac{d\varphi}{dt}\right)^2\right]}\, dt \qquad (12.44)$$

Соответствующие метрические тензоры имеют вид

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\cos^2\theta \end{vmatrix}; \tag{12.45}$$

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & \frac{1}{r^2\cos^2\theta} \end{vmatrix}. \tag{12.46}$$

Если обозначить соответственно тензоры второй валентности (12.45) как ковариантный тензор $g_{i\alpha}$, (12.46) как контравариантный $g^{\alpha j}$, а (12.42) как смешанный $S_i{}^j$, то нетрудно получить соотношение

$$g_{i\alpha}g^{\alpha j}=S_i^j. \tag{12.47}$$

Следует различать пространства и системы координат в этом пространстве. Евклидова геометрия n измерений определяется как теория, которая занимается квадратичной формой типа

$$dS^2=(dy^1)^2+(dy^2)^2+\;.\;\;.\;\;.\;+(dy^n)^2.$$

Каждая система координат, в которой квадратичная форма имеет такой вид, будет называться *ортонормированной* системой [84—91].

В евклидовом пространстве можно ввести несколько систем координат (рис. 12.2). Декартовой системой координат называется такая, в которой коэффициенты g_{ij} постоянны: $g_{ij}=\mathrm{const}$. В множестве декартовых систем координат можно выделить подмножество ортогональных декартовых систем, таких, для которых

$$g_{ij}=\delta_j^i. \tag{12.48}$$

Если ввести репер системы координат, задаваемый ортами $(\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n)$, то в этом случае

$$g_{ij}=\bar{e}_i\bar{e}_j=\delta_j^i. \tag{12.49}$$

Частично под влиянием теории относительности, где используется четырехмерное пространство Минковского (x_1, x_2, x_3, —jt), вводятся некоторые разновидности ортогональных декартовых координат [90], а именно полагают, что

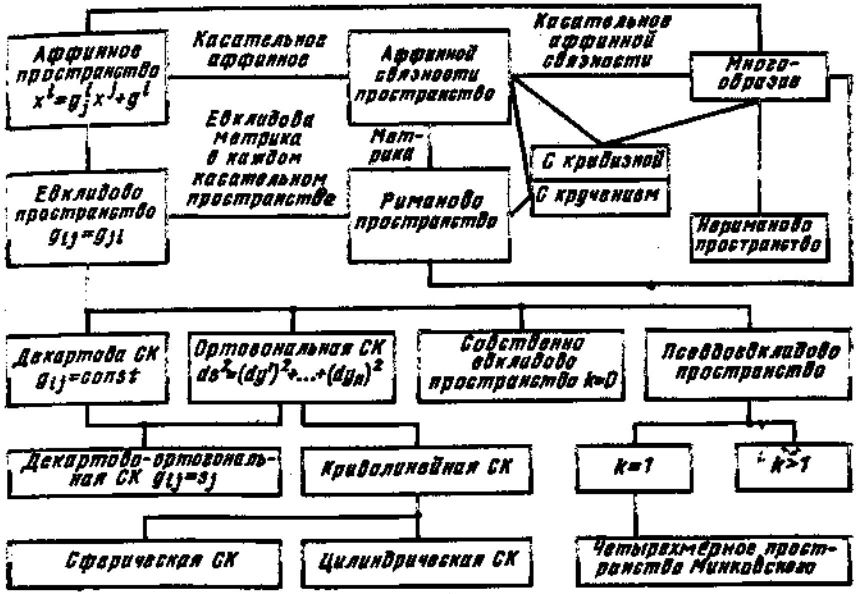


Рис. 12.2. Классификация геометрических пространств

$$g_{ij}=0, \qquad i\neq j; \tag{12.50}$$

$$g_{11}=g_{22}=\;.\;\;.\;\;.\;=g_{kk}=-1;$$

$$g_{k+1,\,k+1}=\;.\;\;.\;\;.\;=g_{nn}=1$$

и называют k-индексным пространством. Для такого пространства связь между ковариантными и контравариантными векторами будет задаваться формулами

$$x_i=-x^i, \qquad i=1,\,2,\;.\;\;.\;\;.\;,\,k;$$

$$x_i=x^i, \qquad i=k+1,\;.\;\;.\;\;.\;,\,n.$$

В связи с этим вводятся в рассмотрение следующие подмножества евклидовых ортогональных декартовых координат:

а) пространство с индексом k;

б) собственно евклидово пространство с $k=0$;

в) псевдоевклидово пространство с $k=1$ и др. Последний тип пространства используется в теории относительности [85, 90].

Матрица ортогонального преобразования, т. е. преобразования ортонормированного репера в ортонормированный, обладает тем свойством, что ее транспонированная совпадает с обратной к ней матрицей. Такая матрица называется *ортогональной*. Теперь имеем

$$(g_{ij}) = (g^{ij}). \tag{12.51}$$

Здесь приравниваются элементы матриц с одинаковыми, не поменявшимися местами индексами. И наоборот с помощью матриц ортогональных преобразований получим ортонормированные реперы. Из определения ортогональной матрицы $\mathbf{A}$ следует, что $\mathbf{A}\cdot\mathbf{A}'=0$, где $\mathbf{A}'$ — матрица, транспонированная к $\mathbf{A}$. Поскольку определители матриц $\mathbf{A}$ и $\mathbf{A}'$ совпадают, то $|\mathbf{A}||\mathbf{A}'|=|\mathbf{A}^2|=1$. Поэтому $|\mathbf{A}|=\pm 1$, т. е. определитель ортогональной матрицы равен ± 1. Если детерминант $(g_{ij})=1$, то такое преобразование называется *собственным движением*, при $(g_{ij}) = -1$ — *несобственным движением*.

Бескоординатное определение тензора

Для корректного введения тензорных понятий необходимым является наличие линейного векторного пространства. Следующим шагом в построении обычных тензоров является определение дуального пространства V^* как пространства линейных функций на V. Если f — линейная функция на V, то обозначим значение этой функции на векторе $x \in V$ через (f, x). Поскольку любой вектор $x \in V$ раскладывается по базису e_i, а любой вектор $f \in V^*$ — по базису f^k, то достаточно задать значения функций f^k на векторах e_i, чтобы описать переход от пространства V к V^*. Легко показать, что это всегда можно сделать так, чтобы выполнялось равенство:

$$f^k(e_i) = \delta^k_i. \tag{12.52}$$

Тогда базисы f^k и e_i называются *взаимными* или *дуальными*. Причем, если задан базис e_i, то взаимный к нему базис определяется однозначно. Если базисы e_i и $e_{i'}$ преобразуются по формуле (12.2), то, используя (12.52), легко показать, что взаимные базисы и f^k и $f^{k'}$ преобразуются по формуле

$$f^k = B^k_{l'} f^{l'}, \tag{12.53}$$

где $B^k_{l'}$ равно $P^l_{l'}$ из (12.2).

Разложив произвольный вектор $b \in V^*$ по базисам f^k и $f^{k'}$, получим

$$b = \beta_k f^k; \qquad b = \beta_{k'} f^{k'}, \tag{12.54}$$

а затем, подставив (12.53) в (12.54), найдем закон преобразования координат β_k вектора b:

$$\beta_l = B^k_l \beta_k. \tag{12.55}$$

Прежде чем перейти к бескоординатному способу определения тензора, введем сначала конструкцию тензорного произведения векторных пространств, имеющую и самостоятельное значение.

Пусть $V_1, V_2, \ldots, V_k$ — конечномерные векторные пространства над полем (или телом) F; k — положительное целое число и V — векторное пространство, натянутое на множество $V_1 \times V_2 \times \ldots \times V_k = \{(e_1, e_2, \ldots, e_k) : e_i \in V_i, i = 1, 2, \ldots, k\}$ как на базис. Очевидно, что V есть множество всех линейных комбинаций символов $(e_1, \ldots, e_2)$ с коэффициентами из поля (или тела) F.

Пусть H — подпространство пространства V, порожденное всеми элементами вида

$$(e_1, \ldots, e_i + v_i, \ldots, e_k) - (e_1, \ldots, e_i, \ldots, e_k) - \\ -(e_1, \ldots, v_i, \ldots, e_k); \tag{12.56}$$

$$\alpha(e_1, \ldots, e_i, \ldots, e_k) - (e_1, \ldots, \alpha e_i, \ldots, e_k), \tag{12.57}$$

где $e_i, v_i \in V_i$, $i = 1, \ldots, k$, и $\alpha \in F$.
Тогда фактор-пространство V/H обозначается символом $V_1 \otimes \ldots \otimes V_k$ и называется *тензорным произведением векторных пространств* $V_1, \ldots V_k$.

Если $e^i \in V_i$, $i = 1, 2, \ldots, k$, то класс $(e_1, \ldots, e_k) + H$ обозначается символом $e_1 \otimes e_2 \otimes \ldots \otimes e_k$ и называется *тензорным произведением векторов* $e_1, e_2, \ldots, e_k$.

Понятно, что отображение φ: $V_1 \times V_2 \times \ldots \times V_k \to V_1 \otimes V_2 \otimes \otimes V_k$, определяемое по правилу

$$\varphi(e_1, e_2, \ldots, e_k) = e_1 \otimes e_2 \otimes \ldots \otimes e_k,$$

однозначно. В силу принадлежности элементов (12.56) и (12.57) подпространству H пространство $V_1 \otimes \ldots \otimes V_k$ удовлетворяет системе тождеств:

$$e_1 \otimes \ldots \otimes (e_i + \\ + v_i) \otimes \ldots \otimes e_k = e_1 \otimes \ldots \otimes e_i \otimes \ldots \otimes e_k + \\ + e_1 \otimes \ldots \otimes v_i \otimes \ldots \otimes e_k;$$

$$\alpha(e_1 \otimes \ldots \otimes e_i \otimes \ldots \otimes e_k) = e_1 \otimes \ldots \otimes (\alpha e_i) \otimes \ldots \otimes e_k,$$

где $e_i, v_i \in V_i$, $i = 1, \ldots, k$ и $\alpha \in F$.

Из этих тождеств следует, что отображение полилинейно, т. е. линейно по каждому аргументу.

Переходим теперь к построению тензорной алгебры векторного пространства V. Пусть V^* — векторное пространство, дуальное к V. Пусть

$$\Pi_i^k = V^* \otimes V^* \otimes \ . \ . \ . \ \otimes V^* \otimes V \otimes \ . \ . \ . \ \otimes V,$$

где V^* и V берутся соответственно i и k раз. (Предполагаем, что размерность пространства V равна r, где r — натуральное число).

Семейство

$$\Pi(V) = \{\Pi_i^k : (k, \ i) \in Z \times Z, \ \ k, \ \ i \geqslant 0\}$$

можно превратить в градуированную алгебру[1].
Задаем на $\Pi(V)$ операцию умножения по правилу

$$(b^1 \otimes \ . \ . \ . \otimes b^i \otimes a_1 \otimes \ . \ . \ . \otimes a_k)(d^1 \otimes \ . \ . \ . \otimes d^e \otimes c_1 \otimes \\ \otimes \ . \ . \ . \otimes c_m) = b^1 \otimes \ . \ . \ . \otimes b^i \otimes d^1 \otimes \ . \ . \ . \otimes d^e \otimes a_1 \otimes \\ \otimes \ . \ . \ . \otimes a_k \otimes c_1 \otimes \ . \ . \ . \otimes c_m.$$

Полученная алгебра $\Pi(V)$ называется *тензорной алгеброй векторного пространства V.*

Иногда $\Pi(V)$ записывают в виде прямой суммы: $\Pi(V) = \sum\limits_{k, \, i > 0} \Pi_i^k$. Элементы алгебры $\Pi(V)$ называются тензорами. Если тензор принадлежит пространству $\Pi_i{}^k$, он называется *тензором валентности k и ковалентности i* (или i раз ковариантным и k раз контравариантным).

Операции сложения, умножения тензоров и умножения их на скаляры из F — это основные операции, определенные в градуированной алгебре $\Pi(V)$.

Определим теперь операцию свертки. Если $e \in V$ и $u \in V^*$, то образ элемента e при отображении u обозначаем через ue. Очевидно, что $ue \in F$.

Сверткой по индексам p и q называется отображение $\partial_{pq} : \Pi_j^k \to \Pi_{i-1}^{k-1}$, определяемое по правилу

$$\partial_{pq}(b^1 \otimes \ . \ . \ . \otimes b^{q-1} \otimes b^q \otimes b^{q+1} \otimes \ . \ . \ . \otimes b^i \otimes a_1 \otimes \ . \ . \ . \otimes \\ \otimes a_{p-1} \otimes a_p \otimes a_{p+1} \otimes \ . \ . \ . \otimes a_k) = b^q a_p (b^1 \otimes \ . \ . \ . \otimes b^{q-1} \otimes b^{q+1} \otimes \\ \otimes \ . \ . \ . \otimes b^i \otimes a_1 \otimes \ . \ . \ . \otimes a_{p-1} \otimes a_{p+1} \otimes \ . \ . \ . \otimes a_k),$$

где

$$a_m \in V, \quad b^l \in V^*, \quad m = 1, \ . \ . \ . \ , k; \quad l = 1, \ . \ . \ . \ , i.$$

---

[1] Градуированной алгеброй над полем (или телом) F называется семейство векторных пространств над F $\{M_{ik} : (i, \ k) \in Z \times Z, \ i, \ k > 0\}$ вместе с выделенным элементом $1 \in M_{0\,0} = F$ и функцией (умножением), которая сопоставляет каждую пару элементов $\lambda \in M_{ik}$, $\mu \in M_{ml}$ с произведением $\lambda\mu \in M_{i+m, \, k+l}$, F — билинейным и удовлетворяющим соотношениям $(\lambda\mu)V = \lambda(\mu V)$; $1\lambda = \lambda 1 = \lambda$.

В алгебре $\Pi(V)$ выделяют некоторые градуированные подалгебры, представляющие значительный интерес. Во-первых, это алгебра

$$C(V) = \left\{\underbrace{V \otimes \ \ . \ . \otimes V}_{m \text{ раз}} : m \in z, \quad m \geqslant 0\right\},$$

называемая *контравариантной*, и, во-вторых,

$$C(V^*) = \left\{\underbrace{V^* \otimes \ . \ . \ . \otimes V^*}_{l \text{ раз}} : l \in z, \quad l \geqslant 0\right\},$$

называемая *ковариантной*. Элементы алгебр $C(V)$ и $C(V^*)$ называются соответственно *контравариантными* и *ковариантными тензорами*.

В алгебрах $C(V)$ и $C(V^*)$ можно ввести операции симметрирования и альтернирования тензоров. Рассмотрим эти операции в алгебре $C(V)$.

Пусть $e_1, \ \ldots, \ e_r$ — базис векторного пространства V. Для каждого $V^{\otimes m} = \underbrace{V \otimes V \otimes \ldots \otimes V}_{m \text{ раз}}$, $m \neq 0$, через σ обозначаем подстановку

$$\begin{pmatrix} 1 & 2 & . \ . \ . & m \\ \sigma(1) & \sigma(2) & . \ . \ . & \sigma(m) \end{pmatrix}$$

и строим линейное отображение $\sigma : V^{\otimes m} \to V^{\otimes m}$, полагая сначала

$$\sigma(e_1 \otimes e_2 \otimes \ . \ . \ . \otimes e_m) = e_{\sigma(1)} \otimes e_{\sigma(2)} \otimes \ . \ . \ . \otimes e_{\sigma(m)}$$

для всех элементов базиса пространства $V^{\otimes m}$ и продолжая затем полученное отображение по линейности на все $V^{\otimes m}$. Для $m = 0$, т. е. для F, полагаем по определению $\sigma(\alpha) = \alpha$ для всякого $\alpha \in F$.

Операции симметрирования (x) и альтернирования $[x]$ на $V^{\otimes m}$ определяются следующим образом:

$$(x) = \sum_{\sigma} \sigma(x);$$

$$[x] = \sum_{\sigma} (-1)^{[\sigma]} \sigma(x),$$

где $[\sigma]$ — четность подстановки σ, и суммирование производится по всем подстановкам σ степени m.

Покажем теперь, каким образом от бескоординатного задания тензора перейти к его координатному заданию.

Пусть, как прежде,

$$e_1, \ e_2, \ . \ . \ . \ , \ e_r \tag{12.58}$$

— базис пространства V и $e^1, \ldots, e^r$ — дуальный к (12.58), т. е. $e^i(e_j)=\delta_j{}^i$, где $\delta_j{}^i$ — символ Кронекера. Тогда тензорное произведение $\Pi_l^k=V^{*\otimes l}\otimes V^{\otimes k}$ имеет базис

$$\{e^{j_1}\otimes\ldots\otimes e^{j_l}\otimes e_{e_1}\otimes\ldots\otimes e_{e_k}:$$

$$:j_1\ldots j_l,\quad e_1,\ldots,e_k=1,\ldots r\},$$

состоящий из r^{l+k} элементов. Если $x\in\Pi_l{}^k$, т. е. является тензором ковалентности l и валентности k, то

$$x=x^{e_1\ldots e_k}_{j_1\ldots j_l}e^{j_1}\otimes\ldots\otimes e^{j_l}\otimes e_{e_1}\otimes\ldots\otimes e_{e_k},$$

где знаки суммирования опущены и введено условие суммирования[1], которое заключается в том, что индекс, встречающийся дважды, вверху и внизу, подразумевает суммирование по всем его значениям от 1 до r. Элементы $x^{e_1\ldots e_k}_{j_1\ldots j_l}$ называются координатами (или компонентами) тензора x.

Пусть

$$e_1', e_2', \ldots, e_r' \tag{12.59}$$

— новый базис пространства V;

$$\begin{pmatrix} C_1^{1'} & \ldots & C_r^{1'} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ C_1^{r'} & \ldots & C_r^{r'} \end{pmatrix} \tag{12.60}$$

— матрица перехода от базиса (12.58) к базису (12.59) и

$$\begin{pmatrix} b_1^1 & \ldots & b_r^1 \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ b_1^r & \ldots & b_r^r \end{pmatrix}$$

— матрица обратного перехода. Переход от (12.58) к (12.59) с матрицей перехода (12.60) индуцирует преобразование базиса в тензорном произведении $\Pi_l{}^k$. Легко проверить, что при этом координаты тензора x преобразуются по правилу

$$x'^{e_1\ldots e_k}_{j_1\ldots j_l}=C^{u_1}_{j_1}\ldots C^{u_l}_{j_l}b^{e_1}_{v_1}\ldots b^{e_k}_{v_k}x^{v_1\ldots v_k}_{u_1\ldots u_l}, \tag{12.61}$$

частные случаи которого рассмотрены выше.

В геометрической и физической литературе закон преобразования (12.61) обычно берется за основу при определении тензора. А именно, тензором ковалентности l и валентности k

---

[1] Это условие суммирования впервые было введено Эйнштейном и поэтому называется правилом Эйнштейна.

называется семейство систем, состоящих из r^{l+k} элементов вида $x^{v_1\ldots v_k}_{u_1\ldots u_l}\in F$, заданных по одной в каждом базисе, которые при преобразовании базисов преобразуются по закону (12.61). Это определение, очевидно, эквивалентно бескоординатному определению тензора. Чтобы установить, является ли заданный в координатах набор величин тензором, достаточно показать, что этот набор преобразуется по закону (12.61). Именно по этой причине приведенное здесь определение тензора удобно для применений и широко используется в геометрии и физике. Недостаток координатного определения тензора состоит в том, что в нем затемняется алгебраическая сущность понятия тензора.

Определение операций над тензорами, заданными в бескоординатной форме, можно переформулировать для тензоров, заданных в координатах. Примеры такой переформулировки рассмотрены ниже.

Теорема о частном. Пусть величина $P_{\lambda\mu\nu}$ такова, что $A^\lambda P_{\lambda\mu\nu}$ является тензором для любого вектора A^α. Тогда $P_{\lambda\mu\nu}$ — тензор.

Доказательство. Полагаем $Q_{\mu\nu}=A^\lambda P_{\lambda\mu\nu}$. По условию $Q_{\mu\nu}$ является тензором. Поэтому согласно правилу (12.61)

$$Q'_{\mu\nu}=c^\beta_\mu c^\gamma_\nu Q_{\beta\gamma}.$$

Тогда $A'^\lambda P'_{\lambda\mu\nu}=c^\beta_\mu c^\gamma_\nu A^\alpha P_{\alpha\beta\gamma}$. Так как A^α является вектором, то имеем

$$A^\alpha=C^\alpha_\lambda A'^\lambda.$$

Таким образом,

$$A'^\lambda P'_{\lambda\mu\gamma}=C^\alpha_\lambda C^\beta_\mu C^\gamma_\nu A'^\lambda P_{\alpha\beta\gamma}.$$

Это равенство должно выполняться для всех A'^α. Поэтому

$$P'_{\lambda\mu\nu}=C^\alpha_\lambda C^\beta_\mu C^\gamma_\nu P_{\alpha\beta\gamma},$$

т. е. согласно (12.61) $P_{\lambda\mu\nu}$ является тензором.

Теорема остается справедливой для величины с произвольным числом нижних и верхних индексов.

Пример 12.2. Даны тензоры $V^\alpha_{\mu\nu}$ и $W^\alpha_{\mu\nu}$ одинакового строения, т. е. принадлежащие пространству $\Pi_2{}^1$ тензорной алгебры $\Pi(V)$. Покажем, что набор величин

$$P^\alpha_{\mu\nu}=V^\alpha_{\mu\nu}+W^\alpha_{\mu\nu}$$

также является тензором, принадлежащим пространству $\Pi_2{}^1$. Поскольку $V^\alpha_{\mu\nu}$ и $W^\alpha_{\mu\nu}$ являются тензорами, то согласно правилу (12.61)

$$W^{\alpha}_{\mu\nu} = C^{u}_{\mu}C^{v}_{\nu}b^{\alpha}_{\rho}W^{\rho}_{uv} \text{ и } V^{\alpha}_{\mu\nu} = C^{u}_{\mu}C^{v}_{\nu}b^{\alpha}_{\rho}V^{\rho}_{uv}.$$

Отсюда

$$P'^{\alpha}_{\mu\nu} = V'^{\alpha}_{\mu\nu} + W'^{\alpha}_{\mu\nu} = c^{u}_{\mu}c^{v}_{\nu}b^{\alpha}_{\rho}V^{\rho}_{uv} + c^{u}_{\mu}c^{v}_{\gamma}b^{\alpha}_{p}W^{\rho}_{uv} = c^{u}_{\mu}c^{v}_{\nu}b^{\alpha}_{\rho}\left(V^{\rho}_{uv} + W^{\rho}_{uv}\right) =$$
$$= c^{u}_{\mu}c^{V}_{\nu}b^{\alpha}_{p}P^{\rho}_{u\delta}.$$

Теперь согласно (12.61) $P^{\alpha}_{\mu\nu}$ — тензор.

Пример 12.3. Пусть даны тензоры $V^{\alpha}_{\mu\nu}$ и $W_{p}{}^{kl}$. Их произведение определяется по правилу

$$P^{\alpha kl}_{\mu\nu p} = V^{\alpha}_{\mu\nu}W^{kl}_{p}.$$

Покажем, что $P^{\alpha kl}$ является тензором. Поскольку по условию $V'^{\alpha}_{\mu\nu}$ и $W_{p}{}^{kl}$ являются тензорами, то они согласно (12.61) преобразуются по правилам $V'^{\alpha}_{\mu\nu} = c^{t}_{\mu}c^{r}_{\nu}b^{\alpha}_{\alpha}V^{\alpha}_{tr}$ и $W'^{kl}_{p} = c^{\beta}_{\rho}b^{k}_{\omega}b^{l}_{\pi}W^{\omega\pi}_{\beta}$. Тогда $P'^{\alpha kl} = V'^{\alpha}_{\mu\nu}W'^{kl}_{p} = c^{t}_{\mu}c^{r}_{\nu}b^{\alpha}_{\alpha} \times$
$\times V^{\alpha}_{tr}C^{\beta}_{p}b^{k}_{\omega}b^{l}_{\pi}W^{u\pi}_{\beta} = c^{t}_{\mu}c^{r}_{\nu}c^{\beta}_{p}b^{\alpha}_{\alpha}b^{k}_{\omega}b^{l}_{\pi}V^{\alpha}_{tr}W^{\omega\pi}_{\beta} = c^{t}_{\mu}c^{r}_{\gamma}c^{\beta}_{\rho}b^{\alpha}_{\alpha}b^{k}_{\omega}b^{l}_{\pi}P^{\alpha\omega\pi}_{tr\beta}.$

Согласно (12.61) $P^{\alpha kl}_{\mu\nu p}$ является тензором.

Пример 12.4. Найти свертку тензора $A^{\mu\nu}_{\alpha\rho\delta}$ по индексам ν и ρ. Имеем

$$\partial_{\nu\rho}\left(A^{\mu\nu}_{\alpha\rho\delta}\right) = A^{\mu\nu}_{\alpha\nu\delta} = \sum_{\nu=1}^{r} A^{\mu\nu}_{\alpha\nu\delta},$$

или получаем

$$C^{\mu}_{\alpha\delta} = \sum_{\nu=1}^{r} A^{\mu\nu}_{\alpha\nu\delta}.$$

12.3. ТЕНЗОР И ПРЕОБРАЗОВАНИЕ ДИФФЕРЕНЦИАЛЬНЫХ ФОРМ

При применении тензорного метода широко используется математический аппарат преобразований в однородных пространствах.

Преобразование аффинных реперов в системе составляет группу квазиаффинных преобразований, которая действует среди многообразия реперов, а не в самом пространстве.

Переменные x^{i} в области Ω аффинного пространства будем называть криволинейными координатами, если они связаны с аффинными координатами в области Ω обратимым и в обе стороны однозначным и непрерывно дифференцируемым преобразованием:

$$x^{i'} = f^{i'}(x^{1}, \ldots, x^{n}); \qquad x^{i} = \varphi^{i}(x'^{1}, \ldots, x'^{n}). \quad (12.62)$$

Якобианы их имеют отличные от нуля детерминанты

$$\left|\frac{\partial x^{i'}}{\partial x^{i}}\right| \neq 0; \qquad \left|\frac{\partial x^{i}}{\partial x^{i'}}\right| \neq 0. \quad (12.63)$$

Рассмотрим подробное тензорное преобразование дифференциальных форм в криволинейной системе координат линейного аффинного пространства. В евклидовом пространстве примером криволинейных координат являются цилиндрические и полярные координаты.

Преобразование криволинейных координат влечет за собой преобразование локального репера в каждой точке M, причем векторы нового репера разлагаются по векторам старого с коэффициентами $\dfrac{\partial x^{i}}{\partial x^{i'}}$, $\left|\dfrac{\partial x^{i}}{\partial x^{i'}}\right| \neq 0$ и $g^{i}_{i'} = \dfrac{\partial x^{i}}{\partial x^{i'}}$,

а роль $\bar{e}_{i}, \bar{e}_{i'}$ играют $\dfrac{\partial}{\partial x^{i}}$ и $\dfrac{\partial}{\partial x^{i'}}$.

Рассмотрим произвольное тензорное поле. В каждой точке возникает локальный репер, и координаты берутся относительно именно этого репера. Поэтому

$$V^{i}_{jk}(M) = V^{i}_{jk} \quad (x^{1}_{|||}x'') \quad (12.64)$$

являются координатами тензора в системе криволинейных координат x^{i}.

Если преобразуются криволинейные координаты, то происходит и преобразование локального репера в каждой точке, а следовательно, и координат тензора по тензорному закону:

$$V^{i'}_{j'k'}(M) = g^{i'}_{i}g^{j}_{j'}g^{k}_{k'}V^{i}_{jk}(M), \quad (12.65)$$

при этом матрица $g^{i}_{i'}$ совпадает с матрицей с элементами $\dfrac{\partial x^{i}}{\partial x^{i'}}$, а

$$g^{i'}_{i} = \frac{\partial x^{i'}}{\partial x^{i}}.$$

Тогда

$$V^{i'}_{j'k'}(M) = \frac{\partial x^{i'}}{\partial x^{i}}(M)\frac{\partial x^{j}}{\partial x^{j'}}(M)\frac{\partial x^{k}}{\partial x^{k'}}(M)V^{i}_{jk}(M). \quad (12.66)$$

Все алгебраические операции переносятся и на тензорные поля (но в каждой точке). С абсолютным дифференциалом тензорных полей дело обстоит сложнее.

Пусть задан касательный вектор $\bar{\xi}_{0}$ в точке M_{0} к аффинному пространству своими координатами (относительно репера в M_{0}). Перенесем его в точку M_{1}. Надо установить, как следует изменить $\bar{\xi}_{0}{}^{i}$, чтобы выразить тот же $\bar{\xi}_{0}$ относительно репера точки M_{1}.

Перенос осуществляется не скачком, а непрерывно вдоль каждого бесконечно малого участка пути $M_{0}M_{1}$. Текущая точка пути будет непрерывно дифференцируемой функцией t, $x^i=x^i(t)$. В каждой точке пути откладываем постоянный вектор ξ_0, координаты которого изменяются ввиду изменения репера:

$$\xi^i=\xi^i(t). \qquad (12.67)$$

Вектор $\bar{\xi}_0$ в локальном репере в точке $M(t)$ равен:

$$\bar{\xi}_0=\xi^i(t)\,\bar{x}_i(x^1\ .\ .\ .\ x^n), \qquad (12.68)$$

где $\bar{x}_i(x^1...x^n)$ — векторы локального репера в точке $M(t)$, а $x^1...x^n$ сами зависят от t. Дифференцируя и учитывая, что $\bar{\xi}_0=$const, имеем

$$0=d\xi^i\bar{x}_i+\xi^i d\bar{x}_i. \qquad (12.69)$$

Далее векторы $d\bar{x}_i$ надо разложить по векторам локального репера.

По формуле полного дифференциала

$$d\bar{x}_i(x^1\ .\ .\ .\ x^n)=\bar{x}_{ij}dx^j, \qquad (12.70)$$

где

$$\bar{x}_{ij}=\frac{\partial^2\bar{x}(x^1\ .\ .\ .\ x^n)}{\partial x^i\partial x^j}. \qquad (12.71)$$

Эти векторы определены для каждой точки. Их можно разложить по векторам локального репера в этой точке:

$$\bar{x}_{ij}=\Gamma^k_{ij}\bar{x}_k, \qquad (12.72)$$

где $\Gamma_{ij}{}^k$ — коэффициенты разложения, зависящие от точки. Равенство $\bar{x}_{ij}=\bar{x}_{ji}$ влечет за собой $\Gamma_{ij}{}^k=\Gamma_{ji}{}^k$. Величины $\Gamma_{ij}{}^k$, определенные в данной системе координат x^i для каждой точки M, называются *коэффициентами связности.*

После подстановки (12.72) в (12.70) получим

$$d\xi^k=-\Gamma^k_{ij}\xi^j dx^i. \qquad (12.73)$$

Эту формулу можно пояснить следующим образом: если в данной точке $M(x^i)$ вектор имеет координаты ξ^k, то такие координаты будет иметь тот же вектор в близкой точке $M(x^i+dx^i)$. С помощью $\Gamma_{ij}{}^k$ связываются векторы в точках M.

Если координаты x^i — аффинные, т. е. $\bar{x}(x^1...x^n)=x_i\bar{e}_i$, $\bar{x}_i=\bar{e}_i$, то согласно (10.72) $\Gamma_{ij}{}^k=0$, и наоборот.

При переходе к другой криволинейной системе $\Gamma_{ij}{}^k$ не будет преобразовываться в $\Gamma^{k'}_{i'j'}$ по тензорным законам, которые задаются формулами $\bar{x}_{ij}=\Gamma_{ij}{}^k\bar{x}_k$; $\bar{x}_{i'j'}=\Gamma^k_{i'j'}\bar{x}_{k'}$. Соответствующие преобразования запишутся следующим образом:

$$\Gamma^{k'}_{i'j'}=\frac{\partial^2x^k}{\partial x^i\partial x^j}\,\frac{\partial x^{k'}}{\partial x^k}+\frac{\partial x^{i'}}{\partial x^i}\,\frac{\partial x^{j'}}{\partial x^{j'}}\,\frac{\partial x^{k'}}{\partial x^k}\Gamma^k_{ij}. \qquad (12.74)$$

Наличие первого члена делает этот закон нетензорным. Если в данной точке M для каждой системы криволинейных координат при переходе к другой системе указана система чисел $\Gamma_{ij}{}^k$, преобразующихся по закону (12.74), то говорят, что в точке M задан объект связности.

Объект связности рассматривается в каждой точке области и образует поле объекта связности. В евклидовом пространстве задан тензор метрики

$$g_{ij}=\bar{e}_i\bar{e}_j, \qquad (12.75)$$

В криволинейных координатах он относится к каждой точке

$$g_{ij}(M)=g_{ij}(x^1\ .\ .\ .\ x^n). \qquad (12.76)$$

При переходе к другой системе криволинейных координат он преобразуется по формуле

$$g_{i'j'}=\frac{\partial x^i}{\partial x^{i'}}\,\frac{\partial x^j}{\partial x^{j'}}g_{ij}. \qquad (12.77)$$

Объект связности $\Gamma_{ij}{}^k(M)$ евклидова пространства можно вычислить, зная метрический тензор $g_{ij}(M)$, в какой-либо криволинейной системе координат. Действительно, если

$$\bar{x}_{ij}=\Gamma^k_{ij}\bar{x}_k$$

умножить скалярно на $\bar{x}_l$, то получим

$$\bar{x}_l\bar{x}_{ij}=\Gamma^k_{ij}g_{kl}. \qquad (12.78)$$

Опустив верхний индекс $\Gamma_{ij}{}^k$, получим

$$\Gamma_{l,\,ij}=g_{lk}\Gamma^k_{ij}. \qquad (12.79)$$

Наоборот, $\Gamma_{ij}{}^k$ получаются из $\Gamma_{l,\,ij}$ поднятием первого индекса:

$$\Gamma^k_{ij}=g^{kl}\Gamma_{l,\,ij}. \qquad (12.80)$$

Для вычисления $\Gamma_{ij}{}^k$ достаточно вычислить $\Gamma_{l,\,ij}$:

$$\left.\begin{aligned}\Gamma_{l,\,ij}&=\bar{x}_l\bar{x}_{ij};\\ \Gamma_{l,\,ij}&=\Gamma_{l,\,ji}.\end{aligned}\right\} \qquad (12.81)$$

Продифференцировав $\bar{x}_l\bar{x}_k$ по x^m почленно, получим

$$\left.\begin{aligned} \bar{x}_{lm}\bar{x}_k+\bar{x}_l\bar{x}_{km}&=\frac{\partial g_{kl}}{\partial x^m};\\ \Gamma_{k,\,lm}+\Gamma_{l,\,km}&=\frac{\partial g_{kl}}{\partial x_m}. \end{aligned}\right\} \qquad (12.82)$$

Сделав попарную перестановку индексов, получим еще два уравнения:

$$\left.\begin{aligned} \Gamma_{l,\,mk}+\Gamma_{m,\,lk}&=\frac{\partial g_{lm}}{\partial x^m};\\ \Gamma_{m,\,kl}+\Gamma_{k,\,ml}&=\frac{\partial g_{mk}}{\partial x^l}. \end{aligned}\right\} \qquad (12.83)$$

Складывая, вычитая и учитывая, что $\Gamma_{l,\,ij}$ по индексам симметричен, получаем

$$\Gamma_{l,\,mk}=\frac{1}{2}\left(\frac{\partial g_{lk}}{\partial x^m}+\frac{\partial g_{lm}}{\partial x^k}-\frac{\partial g_{mk}}{\partial x^l}\right) \qquad (12.84)$$

или

$$\Gamma_{ij}^k=\frac{1}{2}g^{kl}\left(\frac{\partial g_{il}}{\partial x^j}+\frac{\partial g_{lj}}{\partial x^i}+\frac{\partial g_{ij}}{\partial x^l}\right). \qquad (12.85)$$

Полученные выражения называются символами Кристоффеля 1-го и 2-го рода. Если $\bar{x}^i$ — аффинные координаты, то $\bar{x}=x^i\bar{e}_i$, $\bar{x}_i=\bar{e}_i=\text{const}$, $g_{kl}=\bar{x}_k\bar{x}_l=e_ke_l=\text{const}$ и $\Gamma_{kl}{}^i=0$.

12.4. ТЕНЗОР И ТЕОРИЯ МНОГООБРАЗИЙ. ДИФФЕРЕНЦИАЛ КРИСТОФФЕЛЯ

Рассмотрим некоторые вопросы тензорного исчисления с позиций так называемой теории многообразий. Многообразием называется топологическое пространство, всякая область Ω которого гомеоморфна[1] некоторой области n-мерного евклидова пространства R^n. Поэтому в Ω могут быть введены криволинейные системы координат функциями преобразований: $x^1, \ldots, x^n$; $y^1, \ldots, y^n$ с

$$\left.\begin{aligned} T: y^i&=y^i(x^1, \ldots, x^n);\\ T^{-1}: x^i&=x^i(y^1, \ldots, y^n). \end{aligned}\right\} \qquad (12.86)$$

Если функции $y^i(x)$ и $x^i(y)$ — только непрерывны, то многообразие называется *топологическим*. Если же функции $y^i(x)$ и $x^i(y)$ непрерывны вместе со своими первыми частными производными в некоторой области пространства и якобиан $y=\left|\frac{\partial y^i}{\partial x^i}\right|\neq 0$ во всех точках этой области, то многообразие называется *гладким*. В теории гладких многообразий вводится в рассмотрение касательное пространство. Пусть заданы в двух системах координат в точке M системы чисел $x^1, \ldots, x^n$ и $x^{1'}, \ldots, x^{n'}$ и смешенный тензор третьей валентности $V_{jk}{}^i(M)$ и $V_{j'k'}^{i'}(M)$. Тогда преобразование тензоров будет производиться в соответствии с формулой

$$V_{j'k'}^{i'}(M)=\frac{\partial x^{i'}}{\partial x^i}(M)\frac{\partial x^j}{\partial x^{j'}}(M)\frac{\partial x^k}{\partial x^{k'}}(M)V_{jk}^i(M). \qquad (12.87)$$

Роль коэффициентов $g_i^{i'}$ и $g_{i'}^i$ аффинного преобразования

$$x'=g_i^{i'}x^i+g^{i'}$$

играют производные $\frac{\partial x^i}{\partial x^{i'}}$ и $\frac{\partial x^{i'}}{\partial x^i}$. В этой связи появляется концепция касательных пространств. В основу этого направления положена идея множества точек кривой, для которой задано поле реперов в пространстве.

Рассмотрим закон преобразования объектов, определяемых системами частных производных скаляра. Комплект частных производных скаляра $f(x^1, \ldots, x^n)$ будет

$$\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \ldots, \frac{\partial f}{\partial x^n} \text{ или } \{f_{x^i}\}. \qquad (12.88)$$

Их можно рассматривать как компоненты градиента вектора в системе X. Если координаты x^i преобразовать по закону

$$T: x^i=x^i(y^1, \ldots, y^n),$$

то комплект не преобразуется автоматически, а должен быть указан закон преобразования для данного случая. Например, по правилу дифференцирования сложных функций в одной системе координат

$$G: \frac{\partial f}{\partial y^i}=\frac{\partial f}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial y^i}, \qquad (12.89)$$

или в другой системе координат

$$\frac{\partial f}{\partial z^i}=\frac{\partial f}{\partial x^{\alpha}}\frac{\partial x^\alpha}{\partial z^i} \qquad (12.90)$$

и т. д. Можно представить комплекты $\left\{\frac{\partial f}{\partial x^i}\right\}$; $\left\{\frac{\partial f}{\partial y^i}\right\}$; $\left\{\frac{\partial f}{\partial z^i}\right\}$

---

[1] Гомеоморфизмом топологического пространства X в топологическое пространство Y называется взаимно однозначное и взаимно непрерывное отображение X на Y.

и т. д. как один и тот же математический объект, взятый в различных системах координат.

Если имеется комплект n функций $A_1(x), \ldots, A_n(x)$ в системе X и если величины $B_1(y), \ldots, B_n(y)$ в системе y вычисляются по ковариантному закону

$$B_i(y) = \frac{\partial x_\alpha}{\partial y_i} A_\alpha(x), \tag{12.91}$$

то комплект $\{A_i(x)\}$ представляет компоненты ковариантного вектора в системе X. Комплект $B_i(y)$ представляет тот же вектор в системе y.

Закону вычисления коварианта

$$B^i(y) = \frac{\partial y^i}{\partial x^\alpha} A_\alpha(x) \tag{12.92}$$

соответствуют компоненты контравариантного вектора. Комплект величин $A_{i_1, i_2, \ldots, i_r}(x)$ в системе X представляет компоненты ковариантного тензора валентности r, если комплект n^r величин $B_{i_1, \ldots, i_r}(y)$ в системе y находится как

$$B_{i_1, i_2, \ldots, i_r} = \frac{\partial x^{\alpha_1}}{\partial y^{i_1}} \frac{\partial x^{\alpha_2}}{\partial y^{i_2}} \cdots \frac{\partial x^{\alpha_r}}{\partial y^{i_r}} A_{\alpha_1, \ldots, \alpha_r}, \tag{12.93}$$

и контравариантного тензора валентности r, если

$$B^{i_1, i_2, \ldots, i_r} = \frac{\partial y^{i_1}}{\partial x^{\alpha_1}} \cdots \frac{\partial y^{i_r}}{\partial x^{\alpha_r}} A^{\alpha_1, \ldots, \alpha_r}, \tag{12.94}$$

или r раз ковариантного и S раз контравариантного, если

$$B^{j_1, j_2, \ldots, j_r}_{i_1, i_2, \ldots, i_s} = \frac{\partial x^{\alpha_1}}{\partial y^{i_1}} \cdots \frac{\partial x^{\alpha_r}}{\partial y^{i_r}} \frac{\partial y^{j_1}}{\partial x^{\beta_1}} \cdots \frac{\partial y^{j_s}}{\partial x^{\beta_s}} A^{\beta_1, \ldots, \beta_s}_{\alpha_1, \ldots, \alpha_r}. \tag{12.95}$$

Расстояние между точками $\rho_{(x)}$ и $\rho_{(x+dx)}$ в декартовых координатах (элемент дуги)

$$ds^2 = dx^i dx^j.$$

При изменении координатной системы

$$x^i = x^i(y^1, \ldots, y^n);$$

$$ds^2 = \frac{dx^i}{dy^\alpha} \frac{\partial x^j}{\partial y^\beta} dy^\alpha dy^\beta = g_{\alpha\beta} dy^\alpha dy^\beta, \tag{12.96}$$

где $g_{\alpha\beta}(y) = \dfrac{\partial x^i}{\partial y^\alpha} \dfrac{\partial x^i}{\partial y^\beta}$ — метрический тензор, симметричный относительно α и β. Если система имеет решение, то существует преобразование, приводящее квадратичную форму к сумме квадратов, и $g_{\alpha\beta}$ определяет евклидово пространство.

Сумма частных производных тензора $g_{ij}(x)$, обозначаемая символом

$$[ij, k] = \frac{1}{2}\left(\frac{\partial y_k}{\partial x^j} + \frac{\partial y_k}{\partial x^i} - \frac{\partial y_{ij}}{\partial x^k}\right), \quad i, j, k = 1, \ldots, h, \tag{12.97}$$

называется символами Кристоффеля первого рода (трехиндексными), а выражения $\left\{ {k \atop ij} \right\} = g^{k\alpha} [ij, \alpha]$ называются трехиндексными символами Кристоффеля второго рода. Тензор $g^{k\alpha}$ находится из соотношения взаимности

$$g^{k\alpha}, g_{ki} = \delta^i_\alpha. \tag{12.98}$$

Число независимых символов Кристоффеля определяется как

$$N = \frac{1}{2} n^2 (n+1). \tag{12.99}$$

Законы преобразования для $[ij, k]$ и $\left[{k \atop ij} \right]$ следующие. Функции $g_{ij}(x)$ преобразуются при переходе к системе y в функции $h_{ij}(y)$ так, что

$$h_{ij} = \frac{\partial x^\alpha}{\partial y^i} \frac{\partial x^\beta}{\partial y^j} g_{\alpha\beta}. \tag{12.100}$$

Символы Кристоффеля в y $[ij, k]$

$$y[ij, k] = \frac{1}{2}\left(\frac{\partial h_{ik}}{\partial y^j} + \frac{\partial h_{jk}}{\partial y^i} - \frac{\partial h_{ij}}{\partial y^k}\right). \tag{12.101}$$

Дифференцируя h_{ij}, получаем

$$\frac{\partial h_{ij}}{\partial y^k} = g_{\alpha\beta}\left(\frac{\partial^2 x^\alpha}{\partial y^i \partial y^k} \frac{\partial x^\beta}{\partial y^j} + \frac{\partial^2 x^\beta}{\partial y^j \partial y^k} \frac{\partial x^\alpha}{\partial y^i}\right) + \frac{\partial x^\alpha}{\partial y^i} \frac{\partial x^\beta}{\partial y^j} \times$$

$$\times \frac{\partial x^\nu}{\partial y^k} \frac{\partial g_{\alpha\beta}}{\partial x^\nu}. \tag{2.102}$$

Учитывая равенство $g_{\alpha\beta} = g_{\beta\alpha}$ и подставляя в (12.81), получаем

$$y[ij, k] = \frac{\partial x^\alpha}{\partial y^i} \frac{\partial x^\beta}{\partial y^j} \frac{\partial x^\nu}{\partial y^k} [\alpha\beta, \nu] + \frac{\partial^2 x^\alpha}{\partial y^i \partial y^j} \frac{\partial x^\beta}{\partial y^k} g_{\alpha\beta}. \tag{12.103}$$

Эта формула показывает, что символы Кристоффеля преобразуются по тензорному закону при условии, что второй член в (12.103) равен нулю, т. е. когда преобразование координат линейно:

$$y^i = c^i_j x^j \text{ и } c^i_j = \text{const}.$$

Аналогично можно показать, что и символы Кристоффеля второго рода не являются тензорами. Не преобразуется по тензорному закону также комплект производных. Комплект частных производных $\frac{\partial f}{\partial x^i}$ скалярной функции $f(x^1, \ldots, x^n)$ согласно (12.93) представляет ковариантный тензор (вектор), поскольку

$$\frac{\partial f}{\partial y^i}=\frac{\partial f}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial y^i}. \qquad (12.104)$$

Но комплект вторых производных $\left\{\frac{\partial^2 f}{\partial y^i \partial y^j}\right\}$, равных (12.104),

$$\frac{\partial^2 f}{\partial y^i \partial y^j}=\frac{\partial^2 f}{\partial x^\alpha \partial x^\beta}\frac{\partial x^\beta}{\partial y^i}\frac{\partial x^\alpha}{\partial y^j}+\frac{\partial f}{\partial x^\alpha}\frac{\partial^2 x^\alpha}{\partial y^i \partial y^j},$$

из-за наличия второго члена не преобразуется по тензорному закону, если только преобразование координат не будет аффинным. Вторая производная (12.102) получена после дифференцирования ковариантного вектора $\left\{\frac{\partial f}{\partial x^i}\right\}$. Можно показать, что комплект частных производных любого ковариантного вектора не будет тензором.

Пусть $B_i(y)=\frac{\partial x^\alpha}{\partial y^i}A_\alpha(x)$, где вектор A_α ковариантный. Тогда

$$\frac{\partial B_i}{\partial y^j}=\frac{\partial x^\alpha}{\partial y^i}\frac{\partial x^\beta}{\partial y^j}\frac{\partial A_\alpha}{\partial x^\beta}+\frac{\partial^2 x^\alpha}{\partial y^i \partial y^j}A_\alpha. \qquad (12.105)$$

Учитывая формулу Кристоффеля

$$\frac{\partial^2 x^m}{\partial y^i \partial y^j}=\left\{{\nu \atop ij}\right\}\frac{\partial x^m}{\partial y^\nu}-\left\{{m \atop \alpha\beta}\right\}\frac{\partial x^\alpha}{\partial y^i}\frac{\partial x^\beta}{\partial y^j}, \qquad (12.106)$$

получаем

$$\frac{\partial B_i}{\partial y^j}=\frac{\partial x^\alpha}{\partial y^i}\frac{\partial x^\beta}{\partial y^j}\frac{\partial A_\alpha}{\partial x^\beta}+\left\{{\nu \atop ij}\right\}\frac{\partial x^\alpha}{\partial y^\nu}A_\alpha-\left\{{\alpha \atop \nu\beta}\right\}\frac{\partial x^\nu}{\partial y^i}\frac{\partial x^\beta}{\partial y^j}A_\alpha; \qquad (12.107)$$

или

$$\frac{\partial B_i}{\partial y^j}-\left\{{\nu \atop ij}\right\}B_\nu=\left(\frac{\partial A_\alpha}{\partial x^\beta}-\left\{{\nu \atop \alpha\beta}\right\}A_\nu\right)\frac{\partial x^\alpha}{\partial y^i}\frac{\partial x^\beta}{\partial y^j}. \qquad (12.108)$$

Отсюда видно, что совокупность n^2 функций $\frac{\partial A_i}{\partial x_j}-\left\{{\alpha \atop ij}\right\}A_\alpha$ подчиняется тензорному закону ковариантного вектора и называется *ковариантной производной* $\underline{A}$. Для вычисления ковариантной производной необходимо иметь совокупность символов Кристоффеля, т. е. должен быть задан тензор g_{ij}. Совокупность n^2 величин $\frac{\partial A^i}{\partial x_j}+\left\{{i \atop \alpha j}\right\}A^\alpha$ называется *производной контравариантного вектора*. Если A — тензор нулевой валентности, то получим обычную производную. Если $g_{ij}=\mathrm{const}$, то символы Кристоффеля исчезают, имеет место обычная производная. Ковариантная производная фундаментальных тензоров g_{ij} и g^{ij} согласно теореме Риччи равна нулю [85], вследствие чего фундаментальные тензоры могут быть вынесены за знак ковариантного дифференцирования:

$$(g_{\alpha i}A^\alpha_{ik})_l=g_{\alpha i}A^\alpha_{k,l}. \qquad (12.109)$$

Вторая ковариантная производная вектора A_i приводит к выражению

$$A_{i,jk}-A_{i,kj}=\left[\frac{\partial\left\{{\alpha \atop ik}\right\}}{\partial x^j}-\frac{\partial\left\{{\alpha \atop ij}\right\}}{\partial x^k}\right]+\left\{{\beta \atop jk}\right\}\left\{{\alpha \atop \beta j}\right\}-\left\{{\beta \atop ij}\right\}\left\{{\alpha \atop \beta k}\right\}A_\alpha. \qquad (12.110)$$

Выражение в скобках представляет собой смешанный тензор четвертой валентности R^α_{ijk} Если порядок дифференцирования не имеет значения, то $R^\alpha_{ijk}=0$. В общем случае порядок дифференцирования существен. Тензор $R^i{}_{jkl}$ называется тензором Римана—Кристоффеля второго рода:

$$R^i_{jkl}=\begin{vmatrix}\frac{\partial}{\partial x^k} & \frac{\partial}{\partial x^l}\\ \left\{{i \atop jk}\right\} & \left\{{i \atop jl}\right\}\end{vmatrix}+\begin{vmatrix}\left\{{i \atop \alpha k}\right\} & \left\{{i \atop \alpha l}\right\}\\ \left\{{\alpha \atop jk}\right\} & \left\{{\alpha \atop jl}\right\}\end{vmatrix}. \qquad (12.111)$$

Ковариантный тензор Римана—Кристоффеля (первого рода) равен:

$$R_{ijkl}=\begin{vmatrix}\frac{\partial}{\partial x^k} & \frac{\partial}{\partial x^l}\\ [jk,\ i] & [jl,\ i]\end{vmatrix}+\begin{vmatrix}\left\{{\alpha \atop jk}\right\} & \left\{{\alpha \atop jl}\right\}\\ [ik,\ \alpha] & [il,\ \alpha]\end{vmatrix}. \qquad (12.112)$$

Оба тензора связаны равенством

$$R_{ijkl}=g_{i\alpha}R^\alpha_{ijkl}. \qquad (12.113)$$

Тензоры Римана—Кристоффеля обладают свойствами:

$$R_{jikl}=-R_{ijkl};\quad R_{ijlk}=-R_{ijkl};\quad R_{ijkl}=R_{klij};\quad R_{ijkl}+R_{iklj}+R_{iljk}=0. \qquad (12.114)$$

Тензор $R^\alpha_{ij\alpha}=R_{ij}$ называется *тензором Риччи.*

Найдено тождество Бьянки:

$$R^i_{jkl,m}+R^i_{jlm,k}+R^i_{jmk,l}=0. \qquad (12.115)$$

Из тензора Риччи получается тензор Эйнштейна:

$$G^i_j=R^i_j-\frac{1}{2}\delta^i_jR, \qquad (12.116)$$

где $R\equiv g^{ij}R_{ij}$.

Все формулы, полученные для тензоров, справедливы и для тензорных полей.

Римановым пространством называется гладкое многообразие, в котором задано поле тензора

$$g_{ij}(M)=g_{ij}(x^1, \quad . \quad . \quad . \quad , \quad x^n),$$

два раза ковариантного, симметрического, т. е. $g_{ij}=g_{ji}$, невырожденного, т. е. $|g_{ij}|\neq 0$, т. е. задана метрика с помощью метрического тензора аналогично евклидову пространству. Каждой точке M многообразия соответствует касательное аффинное пространство. Представляется возможным, располагая гладким тензорным полем класса $C^\infty g_{ij}(M)$, превращать каждое аффинное пространство A_n в евклидово R_n, вводя в нем скалярное произведение любых двух векторов по формуле

$$\bar{\xi}\bar{\eta}=g_{ij}(M)\,\xi^i\eta^j. \qquad (12.117)$$

Это означает, что риманово пространство V_n — это гладкое многообразие, в котором в каждое касательное пространство A_n внесена евклидова метрика.

В силу полного свертывания (12.116) является инвариантом. Риманово пространство будет собственно римановым или псевдоримановым в зависимости от того, будут ли его касательные пространства собственно- или псевдоевклидовыми.

Все свойства R_n справедливы для касательных пространств в каждой точке риманова пространства A_n (см. рис. 12.2).

12.5. ПРИМЕНЕНИЕ ТЕНЗОРНОГО МЕТОДА В ФИЗИКЕ

При тензорном методе проектирования широко используются тензорные уравнения, которые представляют собой специальный тип алгебраических уравнений. Тензорные уравнения характеризуются следующими особенностями:

а) каждый член тензорного уравнения является тензором;

б) каждое слагаемое представляет собой тензор одной и той же валентности;

в) если в тензорном уравнении каждое слагаемое является 0-тензором, уравнение называется формой: линейной, если каждый член уравнения имеет

вид $L_\alpha i^\alpha$; квадратичной, если каждый член уравнения имеет вид $L_{\alpha\alpha}i^\alpha i^\alpha$; билинейной, если каждый член уравнения имеет вид $L_{\alpha\beta}i^\alpha i^\beta$; полилинейной, если каждый член уравнения имеет вид $L_{\alpha\beta/\delta i_1}i^\alpha i^\beta$. Каждый из индексов пробегает значения от 1 до n, где n — размерность базиса. Эти формы являются алгебраическими, вместо переменных могут быть их производные или дифференциалы, тогда эти формы становятся дифференциальными;

г) если каждое слагаемое в тензорном уравнении 1-тензор, то это слагаемое будет представлять множество форм, систему форм. Например, $Z_{\alpha\beta}i^\beta$ есть множество (оно равно базису) линейных форм; $\Gamma_{\alpha\beta\gamma}i^\alpha i^\beta$ — множество билинейных форм и т. д.

Каждое слагаемое может быть и 2-тензором. Тогда будем иметь систему билинейных форм. Если тензорных уравнений может быть множество, то разнообразие их составных частей — форм ограничено. По существу, современная линейная алгебра посвящена теории форм.

В качестве примера рассмотрим запись в виде тензорных уравнений уравнения электромагнитного поля Максвелла [84]. В электромеханике уравнения Максвелла играют такую же функциональную роль, как уравнения Лагранжа — Гамильтона.

Во-первых, если ввести для механических переменных аналоги электромагнитным переменным $\bar{E}$, $\bar{H}$, $\bar{D}$, $\bar{B}$ и т. д., то и механические и электромагнитные процессы будут описаны уравнениями Максвелла.

Во-вторых, в уравнениях Максвелла заключены свойства оптимальности, в том числе свойство минимакса, так как электромагнитная волна распространяется по геодезической кривой в пространстве.

При использовании формализма Максвелла постулируется, что в любой точке пространства (не особой) векторы поля подчинены следующим уравнениям Максвелла:

$$\left.\begin{aligned}
&\operatorname{rot}\bar{H}=\frac{4\pi}{c}\,j+\frac{1}{c}\frac{\partial D}{\partial t};\\
&\operatorname{rot}\bar{E}=-\frac{1}{c}\frac{\partial B}{\partial t};\\
&\operatorname{div}\bar{B}=0;\\
&\operatorname{div}D=4\pi\rho;\\
&\operatorname{div}j+\frac{\partial P}{\partial t}=0,
\end{aligned}\right\} \qquad (12.118)$$

где $\bar{H}$ — вектор напряженности магнитного поля; $\bar{E}$ — вектор напряженности электрического поля; $\bar{B}$ — вектор магнитной индукции; $\bar{D}$ — вектор электрической индукции; $\bar{j}$ — вектор плотности тока; ρ — плотность электрического заряда; c — скорость света; div A — дивергенция вектора, которая для трехмерного декартова пространства как скалярная величина определяется формулой

$$\operatorname{div}\bar{A}=\frac{\partial A_x}{\partial x}+\frac{\partial A_y}{\partial y}+\frac{\partial A_z}{\partial z}; \qquad (12.119)$$

$\mathrm{rot}\,\bar{A}$ — ротор вектора, который для трехмерного декартова пространства как вектор определяется формулой

$$\mathrm{rot}\,\bar{A} = \begin{vmatrix} \bar{i} & \bar{j} & \bar{k} \\ \partial/\partial x & \partial/\partial y & \partial/\partial z \\ A_x & A_y & A_y \end{vmatrix}. \qquad (12.120)$$

Эти уравнения Максвелла инвариантны при преобразованиях координатных систем только декартова пространства при условии их равномерного движения со скоростью, далекой от скорости света. Это ограничение снимается, если уравнения (12.118) записать в другой форме, а именно в четырехмерном пространстве Минковского. К основным трем измерениям x, y, z добавляется четвертое — время t, затем вводятся новые тензоры переменных в соответствии с таблицами:

$$x^\alpha = \begin{array}{c|c|c|c|c|} \alpha & 1 & 2 & 3 & 4 \\ \hline & x & y & z & jct \\ \hline \end{array}$$

$$\varphi^\alpha = \begin{array}{c|c|c|c|c|} \alpha & 1 & 2 & 3 & 4 \\ \hline & A_x & A_y & A_z & i\varphi \\ \hline \end{array}$$

$$S_\alpha = 4\pi \begin{array}{c|c|c|c|c|} \alpha\backslash\beta & 1 & 2 & 3 & 4 \\ \hline & \rho_0 V^x & \rho_0 V^y & \rho_0 V^z & j\rho_0 \\ \hline & \overline{\beta c} & \overline{\beta c} & \overline{\beta c} & \beta \\ \hline \end{array}$$

$$F^{*\alpha\beta} = \begin{array}{c|c|c|c|c|} \backslash\alpha & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & E_z & -E_y & jB_x \\ \hline 2 & -E_z & 0 & E_x & jB_y \\ \hline 3 & E_y & -E_x & 0 & jB_z \\ \hline 4 & -jB_x & -jB_y & -jB_z & 0 \\ \hline \end{array} \qquad (12.121)$$

$$F^{\alpha\beta} = \begin{array}{c|c|c|c|c|} \alpha\backslash\beta & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & B_z & -B_y & -jE_x \\ \hline 2 & -B_z & 0 & B_x & -jE_y \\ \hline 3 & B_y & -B_x & 0 & -jE_z \\ \hline 4 & jE_x & jE_y & jE_z & 0 \\ \hline \end{array}$$

$$H^{\alpha\beta} = \begin{array}{c|c|c|c|c|} \alpha\backslash\beta & 1 & 2 & 3 & 4 \\ \hline 1 & 0 & H^z & -H^y & -jD^x \\ \hline 2 & -H^z & 0 & H^x & -jD^y \\ \hline 3 & H^y & -H^x & 0 & -jD^z \\ \hline 4 & jD^x & jD^y & jD^z & 0 \\ \hline \end{array}$$

где V — скорость движения заряда; c — скорость света; β — относительная скорость.

С помощью этих тензорных переменных система уравнений Максвелла будет записываться в виде

$$\left.\begin{aligned} &\frac{\partial H^{\alpha\beta}}{\partial x^\beta} = S^\alpha; \qquad \frac{\partial S^\alpha}{\partial x^\alpha} = 0; \\ &\frac{\partial F^{\alpha\beta}}{\partial x^\beta} = 0; \qquad F_{\alpha\beta} = \frac{\partial \varphi_\alpha}{\partial x^\beta} - \frac{\partial \varphi_\beta}{\partial x^\alpha}. \end{aligned}\right\} \qquad (12.122)$$

Эти тензорные уравнения имеют те же ограничения, что и исходные уравнения (12.118), т. е. они остаются инвариантными только относительно преобразования X координат в декартовом пространстве при равномерном движении. Для обеспечения инвариантности тензорных уравнений Максвелла в криволинейных системах координат, движущихся с ускорением, необходимо постулировать справедливость уравнений (12.122) с заменой частных производных тензоров на абсолютные или ковариантные. В результате получим

$$\left.\begin{aligned}&\frac{\delta H^{\alpha\beta}}{\partial x^{\beta}} = S^{\alpha}; \qquad \frac{\delta S^{\alpha}}{\partial x^{\alpha}} = 0;\\ &\frac{\delta F^{\alpha\beta}}{\partial x^{\beta}} = 0; \qquad F_{\alpha\beta} = \frac{\delta\varphi_{\alpha}}{\partial x^{\beta}} - \frac{\delta\varphi_{\beta}}{\partial x^{\alpha}}.\end{aligned}\right\} \tag{12.123}$$

Если использовать символы Кристоффеля, то в развернутом виде эти уравнения запишутся следующим образом:

$$\left.\begin{aligned}&\frac{\partial H^{\alpha\beta}}{\partial x^{\beta}} + \Gamma^{\alpha}_{\nu\beta}H^{\nu\beta} + \Gamma^{\beta}_{\nu\beta}H^{\alpha\nu} = S^{\alpha};\\ &\frac{\partial F^{\alpha\beta}}{\partial x^{\beta}} + \Gamma^{\alpha}_{\nu\beta}F^{\nu\beta} + F^{\beta}_{\nu\beta}F^{\alpha\nu} = 0;\\ &\frac{\partial S^{\alpha}}{\partial x^{\alpha}} + \Gamma^{\alpha}_{\beta\alpha}S^{\beta} = 0,\end{aligned}\right\} \tag{12.124}$$

где S — символ ковариантной производной; $\Gamma^{\alpha}_{\gamma\beta}$ — символ Кристоффеля.

Различие между (12.122) и (12.123) связано с тем обстоятельством, что ковариантные производные тензора только в аффинном пространстве совпадают с частными производными. В криволинейном пространстве появляются члены, содержащие коэффициенты связности $\Gamma_{\beta\gamma,\alpha}$, зависящие от координатной системы.

Сущность тензорного метода анализа технических систем заключается в следующем:

а) подлежащую изучению сложную систему надо анализировать на базе другой, родственной системы, уравнения которой уже известны. Переход от описания простой известной системы к описанию сложной анализируемой системы осуществляется стандартными преобразованиями, правила для которых дает тензорный анализ;

б) более простая система выбирается путем либо разбиения сложной системы на несколько составляющих с помощью удаления некоторых связей, либо принятия новых, более удобных координатных осей нового базиса.

Переход от уравнения простой системы к уравнениям изучаемой сложной составляет содержание теории преобразования и представляет собой основу *тензорного метода.*

Согласно первому принципу теории относительности в физике все физические величины являются тензорными, т. е. они инварианты относительно определенного класса преобразований координат. Это значит, что в различных системах координат этого класса изменяется не сама физическая величина, а ее проекция на оси данной системы.

Одни и те же тензорные уравнения справедливы для целого класса координатных систем, преобразующихся одна в другую с помощью дробно-линейной функции, уравнения каждой конкретной координатной системы будут матричными.

Основой для построения тензорного метода в электромеханике являются два обобщающих постулата [85].

Первый заключается в том, что если имеется один, к примеру, электрический контур (неподвижный, подвижный), то он характеризуется набором величин (R, L, C...), определенным образом между собой связанных алгебраическим уравнением. Если имеется несколько контуров, то они характеризуются матрицами, связанными в матричное уравнение.

Второй постулат говорит, что эти наборы контуров (для отдельных элементов) различными способами могут быть соединены. И такое множество описывается тензорным уравнением.

Все электрические цепи могут быть описаны одним тензорным уравнением:

$$(u) = (z)\,(i), \tag{12.125}$$

все множество электромагнитных систем — парой тензорных уравнений:

$$\left.\begin{aligned}&(u) = (z)\,(i);\\ &(M_{\text{эм}}) = (i)\,(B),\end{aligned}\right\} \tag{12.126}$$

где (B) — индукция; $(M_{\text{эм}})$ — электромагнитный момент.

Токи принимаются за базис. Кроме того, предполагается что вектор тока i контравариантный, а вектор напряжения u — ковариантный, z — дважды ковариантный тензор сопротивления.

12.6. РАСШИРЕННАЯ РЕЛЯЦИОННАЯ И ТЕНЗОРНАЯ АЛГЕБРЫ

Аппаратом для работы с тензорами является тензорная алгебра. Ее сигнатуру составляют операции сложения, умножения, обращения, а также некоторые другие специальные операции. Специфика выполнения операций в тензорной алгебре по сравнению с матричной алгеброй состоит в том, что операции выполняются не в соответствии с порядковыми номерами строк и столбцов, а по ассоциированным со строками и столбцами индексам. Операция сложения выполняется только в том случае, если списки скользящих индексов у операндов совпадают. При этом совпадение фиксированных индексов не требуется, а результат формируется в виде тензора с объединенными списками соответствующих фиксированных индексов. В остальном выполнение операций над тензорами аналогично выполнению операций над матрицами [78, 90, 91].

Для того чтобы продемонстрировать селективную мощность операций расширенной тензорной алгебры, достаточно показать, что эта алгебра является реляционно полной, т. е. что любой операции реляционной алгебры соответствует определенная последовательность операций расширенной тензорной. В силу информационной специфики и селективной направленности операций реляционной алгебры для их покрытия тензорная алгебра расширяется двумя специальными операциями, которыми являются:

θ-операция, которая применяется для определения выполнимости обобщенных θ-условий над множествами фиксированных индексов и компонентов тензоров;

ψ-операция, которая применяется для раздвоения списка скользящих индексов у тензора.

Вводятся две разновидности θ-операции. Это обусловлено, во-первых, тем, что посредством тензоров задаются связи лишь определенных подмножеств скользящих индексов, а в общем случае может потребоваться установить связь между множествами скользящих индексов, на которых построены различные тензоры. Во-вторых, предлагаемая алгебра не является булевой, поэтому в некоторых случаях при выполнении аналогов теоретико-множественных операций могут получаться неоднозначно интерпретируемые результаты. Чтобы избежать этого, вводится θ-операция первого рода. В-третьих, наличие компонентов тензоров, которые не имеют аналога в реляционном подходе, приводит к необходимости их интерпретации, которая заключается в следующем. Если компонент не равен нулю, это означает, что существует связь между соответствующими фиксированными индексами. Значение компонента в этом случае задает количественную меру и этой связи. Если значение компонента равно нулю, то соответствующая связь отсутствует. С точки зрения реляционного подхода величина ненулевого компонента не имеет значения, а набор фиксированных индексов, идентифицирующих компонент в тензоре, соответствует кортежу отношения.

θ-операция первого рода определяется следующим образом. Пусть $I_1, I_2, \ldots, I_k$ — множество скользящих индексов ($i_j^s \in I_j$), которые при $j \in [1, k]$ образуют множества фиксированных индексов; T обозначает множество компонентов $t_{i_1 i_2, \ldots, i_k}$ тензора, построенного на множестве скользящих индексов. θ-операция выполняется на основе обобщенного θ-условия над множеством фиксированных индексов и компонентов тензора. Если тензор на заданном множестве скользящих индексов не построен, то θ-условие определяется только над множеством фиксированных индексов. В результате выполнения θ-операции строится тензор θ_1 на множестве $I_1, I_2, \ldots, I_k$ скользящих индексов, компоненты которого определяются следующим образом:

$$\theta_1(I_1, I_2, \ldots I_k) = \begin{cases} 1, \text{ если } (\forall i_j^s \in I_j): \theta(t_{i_1, i_2, \ldots, i_k}, i_1^s, i_2^s, \ldots, i_k^s); \\ 0, \text{ если } (\forall i_j^s \in I_j): \neg\theta(t_{i_1, i_2, \ldots, i_k}, i_1^s, i_2^s, \ldots, i_k^s). \end{cases} \tag{12.127}$$

θ-операция второго рода отличается от θ-операций первого рода тем, что она применяется только к уже построенным тензорам. θ-условие может задаваться как на всем множестве фиксированных индексов и компонентов, так и на любом его подмножестве. Компоненты тензора результата определяются следующим образом:

$$\theta_2(I_1, I_2, \ldots, I_k) = \begin{cases} t_{i_1, i_2, \ldots, i_k}, \text{ если } (\forall i_j^s \in I_j): \theta(t_{i_1, i_2, \ldots, i_k}, i_1^s, i_2^s, \ldots, i_k^s); \\ 0, \text{ если } (\forall i_j^s \in I_j): \neg\theta(t_{i_1, i_2, \ldots, i_k}, i_1^s, i_2^s, \ldots, i_k^s). \end{cases} \tag{12.128}$$

Обе эти операции являются аналогами выборки.

ψ-операция является своеобразным аналогом конкатенации каждого кортежа отношения с самим собой и соответствует раздвоению списка скользящих, а следовательно, и фиксированных индексов при обозначении компонентов тензора.

Если представить тензор $T_{\text{ЗФД}}$ перечислением его ненулевых компонентов в виде

$$T_{\text{ЗФД}}: \quad \begin{array}{cccc} 1 & З1 & Ф1 & Д1 \\ 1 & З1 & Ф2 & Д2, \\ 1 & З2 & Ф3 & Д2 \end{array} \tag{12.129}$$

то результатом выполнения ψ-операции будет тензор

$$T_{\text{ЗФД ЗФД}} = \psi(T_{\text{ЗФД}}): \quad \begin{array}{ccccccc} 1 & З1 & Ф1 & Д1 & З1 & Ф1 & Д1 \\ 1 & З1 & Ф2 & Д2 & З1 & Ф2 & Д2. \\ 1 & З2 & Ф3 & Д2 & З2 & Ф3 & Д2 \end{array} \tag{12.130}$$

Эта операция вводится и применяется потому, что при умножении двух тензоров с совпадающими индексами по этим индексам производится свертка, и эти индексы пропадают. Для сохранения списка индексов в тензоре результатов используется ψ-операция.

В носителе расширенной тензорной алгебры считаются определенными полностью единичные векторы для каждого скользящего индекса. Примерами таких векторов являются

$$U^{Ф} = \begin{array}{|c|c|c|} \multicolumn{1}{c}{Ф1} & \multicolumn{1}{c}{Ф2} & \multicolumn{1}{c}{Ф3} \\ \hline 1 & 1 & 1 \\ \hline \end{array}; \qquad U^{Д} = \begin{array}{|c|c|} \multicolumn{1}{c}{Д1} & \multicolumn{1}{c}{Д2} \\ \hline 1 & 1 \\ \hline \end{array}. \tag{12.131}$$

Расширенная тензорная алгебра определяется следующим образом:

$$TA=\langle T,\ 0\rangle, \tag{12.132}$$

где $T=<T_n,\ U^k>$; T_n — множество тензоров, каждый из которых соответствует отношению и построен по определенному способу; U^k — множество полностью единичных векторов для каждого скользящего индекса; $0=<+,\ \times,\ ^{-1},\ \theta_1,\ \theta_2,\ \psi>$ — множество операций.

Установим соответствие для четырех теоретико-множественных операций. Пусть каждому отношению, участвующему в выполнении реляционной операции, соответствует тензор, построенный определенным выше образом. Отношению R_A соответствует тензор T_A, а отношению R_B — тензор T_B.

1. Объединение:

$$R_A\cup R_B=T_A+T_B. \tag{12.133}$$

Данная операция выполняется только при совпадении списков скользящих индексов. Все фиксированные индексы T_B, которые отсутствуют в T_A, добавляются к тензору результата. При выполнении сложения может возникнуть ситуация, когда значение компонента будет равно 2. Такой результат можно соответствующим образом проинтерпретировать, однако последующее выполнение других теоретико-множественных операций с использованием такого результата может привести к недоразумениям. Это является следствием разного характера операций. Избежать такой ситуации можно, применив к результату ψ-операцию первого рода с условием неравенства компонента тензора нулю. Результатом будет тензор, который заново можно использовать при выполнении теоретико-множественных операций.

2. Пересечение:

$$R_A\cap R_B=T_A\times\psi(T_B). \tag{12.134}$$

Пояснить применение ψ-операции можно на следующем примере. Пусть требуется найти пересечение двух множеств элементов одного домена. Этим множествам (унарным отношениям R_A и R_B) будут соответствовать два вектора, например:

$$T_A=\begin{array}{|c|c|c|c|}\multicolumn{1}{c}{A}&\multicolumn{1}{c}{B}&\multicolumn{1}{c}{C}&\multicolumn{1}{c}{D}\\\hline 1&1&1&0\\\hline\end{array}\qquad\text{и}\quad T_B=\begin{array}{|c|c|c|c|}\multicolumn{1}{c}{A}&\multicolumn{1}{c}{B}&\multicolumn{1}{c}{C}&\multicolumn{1}{c}{D}\\\hline 0&1&0&1\\\hline\end{array} \tag{12.135}$$

Результат выполнения ψ-операции будет выглядеть следующим образом:

$$\psi(T_B)=\begin{array}{c|c|c|c|c|}\multicolumn{1}{c}{}&\multicolumn{1}{c}{A}&\multicolumn{1}{c}{B}&\multicolumn{1}{c}{C}&\multicolumn{1}{c}{D}\\\hline A&0&0&0&0\\\hline B&0&1&0&0\\\hline C&0&0&0&0\\\hline D&0&0&0&1\\\hline\end{array}\ . \tag{12.136}$$

Окончательно после умножения получим

$$T_A\psi(T_B)=\begin{array}{|c|c|c|c|}\multicolumn{1}{c}{A}&\multicolumn{1}{c}{B}&\multicolumn{1}{c}{C}&\multicolumn{1}{c}{D}\\\hline 0&1&0&0\\\hline\end{array}\ , \tag{12.137}$$

что соответствует элементу B домена.

3. Разность:

$$R_A/R_B=T_A-(T_A\psi(T_B)). \tag{12.138}$$

Данная операция определяется через операцию пересечения.

4. Декартово расширенное произведение:

$$R_A\otimes R_B=T_A\times T_B. \tag{12.139}$$

Специальные операции реляционной алгебры представляются следующим образом.

5. Проекция. Примем следующие обозначения:

D — список атрибутов отношения R_D, для соответствующего T_D — это множество скользящих индексов;

A — список атрибутов, на которые производится проецирование, для T_D — это подмножество скользящих индексов;

$\bar{A}$ — список атрибутов, являющийся дополнением A до D, а каждому атрибуту $a_i\in\bar{A}$ соответствует скользящий индекс:

$$U^{\bar{A}}=U^{a_1}\times U^{a_2}\times\ .\ .\ .\ \times U^{a_l},\qquad\forall a_i\in\bar{A},\qquad i\in[1,\ l]$$

— произведение полностью единичных векторов для каждого скользящего индекса из дополнения.

Операция проекции будет представлена следующим образом:

$$R_D[A]=T_D\times U^{\bar{A}}. \tag{12.140}$$

Рассмотрим пример проектирования $T_{ЗФД}$ на индексы З и Ф. Дополнением списка проецируемых индексов до полного списка является индекс Д. На основании (12.131) получаем

$$T_{ЗФД} \times U^{Д} = \begin{array}{c|c|c|c|} & Ф1 & Ф2 & Ф3 \\ \hline З1 & 1 & 1 & 0 \\ \hline З2 & 0 & 0 & 1 \\ \hline \end{array}. \qquad (12.141)$$

Умножение этого тензора на $U^З$ даст в результате проекцию его на индекс Ф.

6. Ограничение:

$$R_D[A\theta B] \equiv \theta_1(T_D,\ A,\ B). \qquad (12.142)$$

θ-операция позволяет осуществить выборку таких компонентов из тензора T_D, для которых множества A и B фиксированных индексов удовлетворяют θ-условию.

7. Соединение:

$$R_D[A\theta B]\,R_s \equiv T_D \times T_s \times \theta_1(A,\ B). \qquad (12.143)$$

В случае эквисоединения, т. е. когда на фиксированные индексы (значения атрибутов) накладывается условие равенства, в результате выполнения операции θ_1 $(A,\ B)$ будет построена унитарная матрица, а формула (12.143) примет вид

$$R_D\,[A=B]\,R_s \equiv T_D \times T_s. \qquad (12.144)$$

8. Деление. Деление является производной операцией и выражается через остальные операции. Следовательно, эта операция также покрывается множеством введенных операций.

Таким образом, расширенная тензорная алгебра является реляционно полной. Однако специфика ее заключается в том, что наряду с традиционно выполняемыми реляционной алгеброй структурными преобразованиями отношений тензорная алгебра работает также и с количественными характеристиками, что делает предлагаемый формализм более мощным.

Часть 5

ГЕОМЕТРИЧЕСКИЕ РАЗДЕЛЫ

Глава 13

ЭЛЕМЕНТЫ АЛГЕБРАИЧЕСКОЙ ТОПОЛОГИИ*

В данной главе рассматриваются некоторые аспекты взаимосвязи топологического и алгебраического описания многомерных объектов.

В § 13.1 рассматриваются простейшие понятия из теории цепных и симплициальных комплексов [92—98].

В § 13.2 строятся цепные и коцепные комплексы конкретной топологической сети [94].

В § 13.3 кратко излагается алгебраический аппарат, позволяющий строить многомерные алгебры на топологических сетях [95, 96].

Понятие топологической сети используется в качестве «перекидного мостика» от алгебраической топологии к описанию реальных объектов. С помощью топологических сетей с привязанными к ним цепным и коцепным комплексами, тензорными алгебрами можно моделировать сложные электрические и энергетические сети, трубопроводы, арочные конструкции и банки данных.

13.1. ВВЕДЕНИЕ В ТЕОРИЮ ЦЕПНЫХ И СИМПЛИЦИАЛЬНЫХ КОМПЛЕКСОВ

Напомним (см. гл. 3), что множество X элементов произвольной природы называется *группой*, если в этом множестве определена бинарная операция, сопоставляющая каждую пару элементов $a,\ b \in X$ с элементом $c=ab$ из X таким образом, что выполняются три аксиомы:

1) операция ассоциативна, т. е. $(ab)\,c=a\,(bc)$;

2) существует единичный элемент, т. е. существует элемент e, такой, что $ea=ae=a$;

3) операция обратима, т. е. для любого a существует в X такой элемент a', что $aa'=a'a=e$.

Элемент a' называется *обратным* для a и обычно обозначается a^{-1}.

Подмножество X' элементов X называется *подгруппой* этой группы, если для любой пары элементов $x_1,\ x_2 \in X'$ элемент $x_1x_2^{-1} \in X'$. Подгруппа X' группы X, рассматриваемая как самостоятельное множество с бинаторной операцией из группы X, представляет собой группу.

В абелевых группах бинарная операция, обозначаемая знаком $+$, коммутативна, т. е.

$$a+b=b+a,$$

единичный элемент группы обозначается 0 и называется нулем, обратным для a является элемент $-a$.

Если в абелевой группе X можно выбрать конечную совокупность элементов $M=\{m_1, \ldots, m_k\}$, такую, что любой $x \in X$ является их линейной комбинацией, т. е.

$$x=c_1m_1+\ .\ .\ .\ +c_km_k,$$

где $c_1, \ldots, c_k$ — целые числа, то такую абелеву группу называют абелевой группой с конечным числом образующих.

Система образующих M называется *свободной*, если соотношение

$$c_1m_1+\ .\ .\ .\ +c_km_k=0$$

удовлетворяется тогда и только тогда, когда все c_i — нули, т. е. образующие m_i линейно независимы над кольцом целых чисел z.

Абелева группа называется свободной абелевой группой ранга n, если она имеет n линейно независимых образующих, составляющих систему элементов, называемую *базой* этой группы.

В дальнейшем рассматриваются только абелевы группы с конечным числом образующих.

Рассмотрим некоторую абелеву группу G, в которой действует оператор d, обладающий свойством $ddx=0$ для любого x из G. Подгруппу B группы G, образованную элементами вида $y=dx$, будем называть *подгруппой границ*. Подгруппу $Q \subset G$, для каждого элемента z которой справедливо $dz=0$, будем называть *подгруппой циклов*. Можно отметить, что подгруппа B есть образ оператора d, а подгруппа Q есть ядро оператора d, т. е. $B=\operatorname{im} d$, $Q=\ker d$. Подгруппа границ принадлежит подгруппе циклов, так как для любого $x=dy$, принадлежащего B, справедливо $dx=ddy=0$. Значит, если $x \in B$, то $x \in Q$. Пусть имеется последовательность, образованная из нескольких групп и операторов:

$$N_1 \xrightarrow{P_1} N_2 \xrightarrow{P_2}\ .\ .\ .\ N_{i-1} \xrightarrow{P_{i-1}} N_i \xrightarrow{P_i} N_{i+1} \xrightarrow{P_{i+1}}. \tag{13.1}$$

Если в каждой из последовательностей

$$N_{i-1} \xrightarrow{P_{i-1}} N_i \xrightarrow{P_i} N_{i+1}$$

$\operatorname{im} P_{i-1}=\ker P_i$, то последовательность (13.1) называется *точной*.

Цепным комплексом K называется абелева группа G с оператором, допускающая разложение в прямую сумму $G=\Sigma N_i$ с подгруппой N_i, $i=0,1$, причем $ddx=0$ для $\forall x \in G$, $d(N_i) \subseteq N_{i-1}$ и $d(N_0)=0$. Элементы группы N_i называются *цепями размерности* i. Группам цепей соответствуют подгруппы циклов и границ размерности i:

$$Q_i(k)=Q \cap N_i(k); \qquad B_i(k)=B \cap N_i(k).$$

Пусть M — произвольное множество и F — множество конечных подмножеств множества M. Множество F называется *симплициальным комплексом* с вершинами в множестве M, если из соотношений $\beta \in F$ и $\alpha \in \beta$ следует $\{\alpha\} \in F$. Элементы множества F называются *гранями*. Если грань α состоит из $q+1$ элементов множества M, то говорят, что α имеет размерность q, а эти $q+1$ элементов, которые образуют грань α, называются *вершинами грани* α.

Сингулярным симплексом $k_q=(a_0, \ldots, a_q)$, $a_i \in M$ размерности q называется упорядоченное множество, состоящее из $q+1$ элементов a_i, которые расположены на одной грани α комплекса F. Сингулярные симплексы комплекса F порождают свободную абелеву группу $N(F)$, которая может быть разложена в прямую сумму групп $N_q(F)$, где $N_q(F)$ — подгруппа группы $N(F)$, порожденная сингулярными симплексами размерности q. Элементы группы $N(F)$ называются *сингулярными цепями*.

Граница сингулярного симплекса $k_q=(a_0, \ldots, a_q)$ определяется формулой

$$dk_q=\Sigma\,(-1)^i(a_0,\ .\ .\ .\ .\ ,\hat{a}_i,\ .\ .\ .\ .\ ,a_q),$$

где $\hat{a}_i$ означает отсутствие вершины a_i в последовательности.

Действие оператора d можно распространить и на произвольные линейные комбинации симплексов, тогда граница комбинации симплексов $k_q+m_p+n_r$ будет равна:

$$d(\alpha k_q+\beta m_p+\gamma n_r)=\alpha dk_q+\beta dm_p+\gamma dn_r.$$

Для симплекса S_0 нулевой размерности $dS_0=0$. Теперь покажем, что симплициальный комплекс является и цепным комплексом. Для этого надо доказать, что $dd_n=0$ для любого $n \in N$.

Пусть $n=(a_0, \ldots, a_i, \ldots, a_p)$ — симплекс размерности p. Для каждого индекса $i\,(0 \leqslant i \leqslant p)$ обозначим $n^i=(a_0, \ldots, \hat{a}_i, \ldots, a_p)$ — симплекс размерности $p-1$; $n^{ij}=(a_0, \ldots, \hat{a}_i, \ldots, \hat{a}_j, \ldots, a_p)$ — симплекс размерности $p-2$, считая $i<j$. Тогда полагаем

$$dn=\sum_{i=0}^{p}(-1)^i n^i;$$

$$dn=\sum_{j=0}^{p-1}(-1)^j(n^i)^j;$$

$$ddn=\sum_{i=0}^{p}(-1)^i dn^i=\sum_{i=0}^{p}(-1)^i\sum_{j=0}^{p-1}(-1)^j(n^i)^j.$$

Заметим, что при $j<i$

$$(n^i)^j=n^{ji};$$

а для $j\geqslant i$

$$(n^i)^j=n^{i,j+1},$$

так как на j-м месте в последовательности находится a_{j+1}. Тогда

$$ddn=\sum_{i=0}^{p}(-1)^i\sum_{j=0}^{i-1}(-1)^j n^{ji}+\sum_{i=0}^{p}(-1)^i\sum_{j=1}^{p-1}(-1)^j n^{i,j+1},$$

но, поскольку в этом выражении для любых двух индексов $r<t$, коэффициент при n^{rt} равен $(-1)^{r+t}+(-1)^{r+t-1}=0$. Значит, $ddn=0$.

Стандартным симплексом размерности q называется множество Δ^q точек $x=(x_0,\ldots,x_q)$ евклидова пространства E^{q+1}, удовлетворяющее условиям

$$0\leqslant x_i\leqslant 1;\qquad \sum_{i=0}^{q}x_i=1. \tag{13.2}$$

Подмножество Δ^q евклидова пространства выпукло, так как если x и $y\in\Delta^q$, то все точки отрезка xy принадлежат Δ^q. Действительно, любую точку $z=(z_0,\ldots,z_q)$, принадлежащую отрезку xy, можно записать так:

$$z_i=x_i+t(y_i-x_i)\qquad(0\leqslant t\leqslant 1).$$

Тогда

$$z_i=(1-t)x_i+ty_i\geqslant 0$$

и

$$\sum_{i=0}^{q}z_i=(1-t)\sum_{i=0}^{q}x_i+t\sum_{i=0}^{q}y_i=1,$$

так как x и y удовлетворяют условиям (13.2).

Последнее свойство позволяет ввести в Δ^q линейную структуру: будем говорить, что точки $x^1,\ldots,x^r$ из Δ^q линейно зависимы, если существуют r действительных чисел $\lambda_1,\ldots,\lambda_r$, таких, что

$$\lambda_1^2+\ldots+\lambda_r^2\neq 0;$$
$$\lambda_1+\ldots+\lambda_r=0$$

и

$$\lambda_1x^1+\ldots+\lambda_rx^r=0.$$

Точки $x^1,\ldots,x^p$ называются линейно независимыми, если они не являются линейно зависимыми. Точки A^α из $\Delta^q(\alpha=0,\ldots,q)$, имеющие координаты

$$x_i^\alpha=\begin{cases}0, & \text{если}\quad i\neq\alpha;\\ 1, & \text{если}\quad i=\alpha,\end{cases}$$

линейно независимы, потому что соотношение

$$\sum_{\alpha=0}^{q}\lambda_\alpha A^\alpha=0$$

влечет за собой $\lambda_i=0$ $(i=0,\ldots,q)$. Точки A^α называются вершинами симплексов Δ^q. Любая точка $x\in\Delta^q$ линейно зависит от точек A^α, так как

$$x=x_0A^0+x_1A^1+\ldots+x_qA^q,$$

где x_i — координаты точки x, для которых $\sum_{i=0}^{q}x_i=1$. Отсюда следует, что в Δ^q существует $q+1$ линейно независимых точек, но не существует $q+2$ линейно независимых точек.

Чтобы наделить множество Δ^q структурой метрического пространства, определим расстояние между точками x и y формулой

$$m(x,y)=[(x_0-y_0)^2+\ldots+(x_q-y_q)^2]^{1/2}. \tag{13.3}$$

Расстоянием между точками симплекса будем считать расстояние между этими же точками, как и в евклидовом пространстве с метрикой (13.3).

Примеры симплексов Δ^0, Δ^1 и Δ^2 приведены на рис. 13.1.

Границей симплекса Δ^q называется множество его точек, для которых выполняется соотношение (13.2), причем, по крайней мере, одна из координат тождественно равна нулю. Внутренность симплекса образована точками симплекса, все координаты которых положительны. Обозначим через $\Delta^q_{i_1,\ldots,i_p}$ $(0<i_1<\ldots<i_p<q)$ множество точек $x=(x_0,\ldots,x_q)$ из Δ^q, для которых

$$x_{i_1}=\ldots=x_{i_p}=0.$$

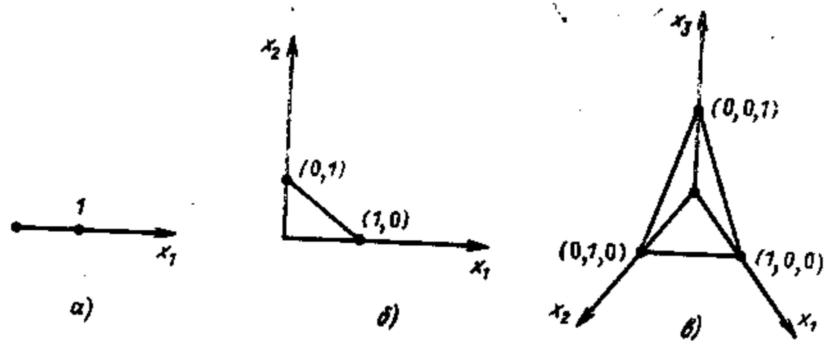


Рис. 13.1. Пояснение к понятию симплекса:
а — 0-мерный симплекс; б — 1-мерный симплекс; в — 2-мерный симплекс

Множества $\Delta^q_{i_1, \ldots, i_p}$ называются *гранями симплекса* размерности $q-p$. Они принадлежат границе симплекса Δ^q, и их число равно c^p_{q+1} (число сочетаний из $q+1$ элементов по P).

Каждая грань $\Delta^q_{i_1, \ldots, i_p}$ содержит $q-p+1$ вершин A^α, где $\alpha \neq i_1, \ldots, i_p$. Грани $\Delta_i{}^q$ размерности $q-1$ содержат каждая q вершин. Например, грань $\Delta_0{}^q$ содержит вершины $A^1, \ldots, A^q$ без A^0.

Линейный симплекс является обобщением стандартного симплекса. Если $p^0, \ldots, p^q$ — линейно независимые точки в евклидовом пространстве E^N, то будем называть линейным симплексом, порожденным этими точками, множество точек X из E^N вида

$$x = \lambda_0 p^0 + \ . \ . \ . + \lambda_p p^q;$$

$$0 \leqslant \lambda_i \leqslant 1, \qquad \sum_{i=0}^{q} \lambda_i = 1.$$

Два линейных симплекса могут представлять собой одно и то же множество точек из E^N, например линейные симплексы $(p^0, \ . \ . \ . \ , p^q)$ и $(p^{i_0}, \ . \ . \ . \ , p^{i_q})$, где $i_0, \ . \ . \ . \ , i_q$ — перестановка чисел $0, \ldots, q$. Эти симплексы считаются различными, т. е. линейный симплекс — это множество точек с определенным порядком следования его вершин.

Можно установить однозначное отображение линейного симплекса $(p^0, \ldots, p^q)$ в стандартный симплекс Δ^q, когда координаты x_i приравниваются λ_i.

Грани симплекса Δ^q являются примерами линейных симплексов. Например, грань $\Delta^q_{i_1, \ldots, i_p}$ может быть представлена как $(A^{j_0}, \ . \ . \ . \ , A^{j_{q-p}})$, где $0 \leqslant j_0, \ldots, j_{q-p} \leqslant q$ и $j_0, \ldots, j_{q-p}$ отличны от $i_1, \ldots, i_p$. Например, $\Delta_0{}^q = (A^1, \ldots, A^q)$.

13.2. ПОСТРОЕНИЕ ЦЕПНОГО КОМПЛЕКСА НА КОНКРЕТНОЙ ТОПОЛОГИЧЕСКОЙ СЕТИ

Объект, изображенный на рис. 13.2, будем называть *топологической сетью*, элементы топологической сети — *гранями*. Запишем множества граней данной сети:

$M = \{a_1, a_2, a_3, a_4\}$ — множество нуль-мерных граней, или 0-граней (узлов);

$$M_1 = \{b_1, b_2, b_3, b_4, b_5, b_6\} = \{a_1a_2, a_2a_3, a_3a_1, a_2a_4, a_1a_4, a_4a_3\}$$

— множество одномерных граней, или 1-граней (ветвей);

$$M_2 = \{c_1, c_2, c_3, c_4\} = \{a_1a_2a_3, a_1a_2a_4, a_2a_3a_4, a_1a_3a_4\}$$

— множество двухмерных граней, или 2-граней;

$$M_3 = \{d_1\} = \{a_1a_2a_3a_4\}$$

— множество трехмерных граней или 3-граней.

Каждая грань характеризуется размерностью и ориентацией. Ориентация грани задается порядком узлов в ее записи. Считаем, что в записи множеств M_0, M_1, M_2, M_3 все грани даны в положительной ориентации. Изменение порядка узлов в записи грани вызывает изменение ее ориентации, например:

$$a_2a_1 = -a_1a_2;$$

$$a_1a_2a_3 = -a_2a_1a_3 = a_2a_3a_1.$$

Для каждой пары граней соседних размерностей можно определить коэффициент инцидентности. Если $(n-1)$-я грань x является подгранью грани y, т. е. имеет единственный узел, не принадлежащий x, то коэффициент инцидентности равен:

$$y : x = a_{i_1} \ . \ . \ . \ a_{i_{m-1}} a_{i_m} a_{i_{m+1}} \ . \ . \ . \ a_{i_n} : a_{i_1} \ . \ . \ . \ a_{i_{m-1}} \widehat{a}_{i_m} a_{i_{m+1}} \ . \ . \ . \ a_{i_n} = (-1)^{m-1};$$

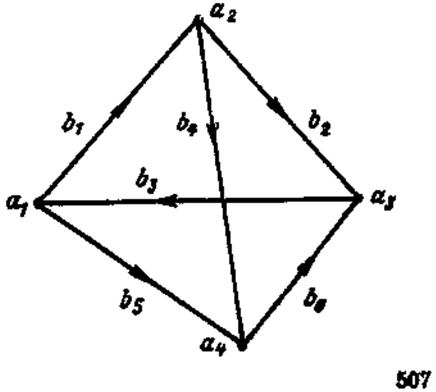


Рис. 13.2. Трехмерная топологическая сеть

здесь $\hat{a}_{i_m}$ означает отсутствие узла a_{i_m} в записи грани x. Если x не является подгранью грани y, то

$$y : x = 0.$$

Приведем примеры расчета коэффициентов инцидентности для данной сети:

$$b_2 : a_3 = a_2a_3 : a_3 = a_{i_1}a_{i_2} : \hat{a}_{i_1}a_{i_2} = (-1)^{1-1} = 1;$$

$$c_1 : b_4 = a_1a_2a_3 : a_3a_4 = 0;$$

$$d_1 : c_4 = a_1a_2a_3a_4 : a_1a_4a_3 = a_1a_2a_3a_4 : \quad -\left(a_1\hat{a}_2a_3a_4\right) = 1.$$

Коэффициенты инцидентности являются элементами матриц инцидентности, которые для данной сети имеют вид:

матрица инцидентности множеств M_0 и M_1

$$\begin{array}{c|cccccc|} & b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ \hline a_1 & -1 & 0 & 1 & 0 & -1 & 0 \\ a_2 & 1 & -1 & 0 & -1 & 0 & 0 \\ a_3 & 0 & 1 & -1 & 0 & 0 & 1 \\ a_4 & 0 & 0 & 0 & 1 & 1 & -1 \\ \hline \end{array} \tag{I}$$

матрица инцидентности множеств M_1 и M_2

$$\begin{array}{c|cccc|} & c_1 & c_2 & c_3 & c_4 \\ \hline b_1 & 1 & 1 & 0 & 0 \\ b_2 & 1 & 0 & 1 & 0 \\ b_3 & 1 & 0 & 0 & -1 \\ b_4 & 0 & 1 & -1 & 0 \\ b_5 & 0 & -1 & 0 & -1 \\ b_6 & 0 & 0 & -1 & 1 \\ \hline \end{array} \tag{II}$$

матрица инцидентности множеств M_2 и M_3

$$\begin{array}{c|c|} & d_1 \\ \hline c_1 & -1 \\ c_2 & 1 \\ c_3 & 1 \\ c_4 & -1 \\ \hline \end{array} \tag{III}$$

Абелева группа X называется векторным пространством над полем действительных чисел R, если выполнены требования

1) $ax \in X$;
2) $a(x+y) = ax + ay$;
3) $(a+b)x = ax + bx$;
4) $(ab)x = a(bx)$;
5) $1 \cdot x = x$

для любых a, $b \in R$; x, $y \in X$, где 1 — единичный элемент поля R. Элементы поля R будем называть *коэффициентами* или *скалярами*, а элементы группы X — *векторами*.

Векторное пространство X разлагается в прямую сумму подпространств X' и X'', т. е. $X = X' \oplus X''$, если любой элемент $x \in X$ может быть единственным способом представлен в виде суммы $x = x' + x''$, где $x' \in X'$ и $x'' \in X''$.

Данной топологической сети могут быть поставлены в соответствие свободные абелевы группы 0-, 1-, 2-, 3-мерных цепей со свободными системами порождающих M_i, $i = 0, 1, 2, 3$. Тогда матрица инцидентности множеств M_1 и M_2 и матрица инцидентности множеств M_2 и M_3 имеют вид II и III соответственно.

Данной топологической сети могут быть поставлены в соответствие вещественные векторные пространства 0-, 1-, 2-, 3-мерных цепей. Действительно, пусть M_i — множество i-мерных цепей, $i = 0, 1, 2, 3$. Рассмотрим декартово произведение $R \times M_i$, где R — поле вещественных чисел. Пусть N_i $(i = 0, 1, 2, 3)$ — абелева группа, порожденная множеством $R \times M_i$. Определяя в N_i умножение на вещественные числа слева по правилу

$$P\left(\sum_{j=1}^{k}(r^j, m_i^j)\right) = \sum_{j=1}^{k}(pr^j, m_i^j),$$

превращаем N_i, $i = 0, 1, 2, 3$ в вещественное векторное пространство. Пары (r, m_i) в дальнейшем будем записывать в виде

rm_i. Тогда элементы векторных пространств N_0, N_1, N_2, N_3 можно записать в виде

$$\left.\begin{aligned} n_0 &= n_0^1 a_1 + n_0^2 a_2 + n_0^3 a_3 + n_0^4 a_4 = n_0^i a_i; \\ n_1 &= n_1^1 b_1 + n_1^2 b_2 + n_1^3 b_3 + n_1^4 b_4 = n_1^j b_j; \\ n_2 &= n_2^1 c_1 + n_2^2 c_2 + n_2^3 c_3 + n_2^4 c_4 = n_2^k c_k; \\ n_3 &= n_3^1 d_1, \end{aligned}\right\} \qquad (13.4)$$

где коэффициенты $n_0{}^i$, $n_1{}^j$, $n_2{}^k$, $n_3{}^1$ являются координатами векторов n_0, n_1, n_2, n_3 в векторных пространствах N_0, N_1, N_2, N_3 соответственно.

Рассмотрим линейное отображение

$$q^0(a_i) = (0, \ . \ . \ . \ , 0, 1, 0, \ . \ . \ . \ , 0),$$

где единица стоит на i-м месте. Тогда для нашей сети имеем

$$\left.\begin{aligned} q^0(a_1) &= q^0_{a_1} = (1, \ 0, \ 0, \ 0); \\ q^0(a_2) &= q^0_{a_2} = (0, \ 1, \ 0, \ 0); \\ q^0(a_3) &= q^0_{a_3} = (0, \ 0, \ 1, \ 0); \\ q^0(a_4) &= q^0_{a_4} = (0, \ 0, \ 0, \ 1). \end{aligned}\right\} \qquad (13.5)$$

Линейно независимые векторы $q^0_{a_i}$ образуют базис четырехмерного векторного пространства N_0. Соотношения (13.5) устанавливают взаимно однозначную связь элементов базы M_0 и базисных векторов пространства N_0. Для произвольной цепи $n_0 \in N_0$ можно получить

$$q^0(n_0) = q^0\left(n_0^1 a_2 + \ . \ . \ . \ + n_0^4 a_4\right) = n_0^1 q^0(a_1) + \ . \ . \ . \ + n_0^4 q^0(a_4) = \\ = n_0^i q^0(a_i) = n_0^i q^0_{a_i},$$

или в матричном виде

$$n_0^a \ \begin{array}{|c|c|c|c|}\hline n_0^1 & n_0^2 & n_0^3 & n_0^4 \\ \hline\end{array}$$

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Таким образом, взаимно однозначное отображение q^0, переводящее цепи в векторы, позволяет перейти к матричным соотношениям.

Аналогично можно построить координатные векторные пространства N_1—N_3, соответствующие данной сети.

Можно задать взаимно однозначное отображение q^1 из множества M_1 в пространство N_1 в виде

$$\begin{aligned} q^1(b_1) &= q^1_{b_1} = (1, \ 0, \ 0, \ 0, \ 0, \ 0); \\ q^1(b_2) &= q^1_{b_2} = (0, \ 1, \ 0, \ 0, \ 0, \ 0); \\ &. \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \ . \\ q^1(b_6) &= q^1_{b_6} = (0, \ 0, \ 0, \ 0, \ 0, \ 1). \end{aligned}$$

Тогда для произвольной 1-й цепи n_1 из (13.4) можно получить

$$q^1(n_1) = q^1\left(n_1^1 b_1 + \ . \ . \ . \ + n_1^6 b_6\right) = n_1^1 q^1(b_1) + \ . \ . \ . \ + \\ + n_1^6 q^1(b_6) = n_1^j q^1(b_j).$$

Для пространства N_2 и множества M_2 можно ввести взаимно однозначное отображение q^2, для которого

$$\begin{aligned} q^2(c_1) &= q^2_{c_1} = (1, \ 0, \ 0, \ 0); \\ q^2(c_4) &= q^2_{c_4} = (0, \ 0, \ 0, \ 1). \end{aligned}$$

Тогда

$$q^2(n_2) = q^2\left(n_2^1 c_1 + \ . \ . \ . \ + n_2^4 c_4\right) = \\ = n_2^1 q^2_{c_1} + \ . \ . \ . \ + n_2^4 q^2_{c_4} = n_2^j q^2_{c_j}.$$

Можно задать q^3, действующее из M_3 в N_3, для которого $q^3(d_1) = q^3_{d_1} = (1)$. Тогда $q^3(n_3) = n_3{}^1 q^3_{d_1}$. Каждой цепи соответствует единственный вектор, так как отображения q_i взаимно однозначны. Поэтому в дальнейшем эти понятия не будем различать и будем считать, что $q^i(n_i) = n_i$.

В общем случае можно сказать, что отображения переводят базы векторных пространств цепей в базисные векторы координатных векторных пространств. Выше мы рассмотрели только базы M_0, M_1, M_2, M_3, хотя на их месте могут быть и другие базы.

Объединяя базисные векторы в матрицы, получаем матрицы базисов. Например, в матрицах базисов $q_a{}^0$, $q_b{}^1$, $q_c{}^2$, $q_d{}^3$ базисные вектор-столбцы записываются в строки. Обычно при индексировании матриц перед индексом столбцов ставится точка, поэтому, записывая базисные векторы в столбцы или в строки, можно получить

$$n_0' = \left(\overset{q_a^0}{\begin{array}{|c|c|c|c|}\hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}} \cdot \overset{n_0}{\begin{array}{|c|}\hline n_0^1 \\ \hline n_0^2 \\ \hline n_0^3 \\ \hline n_0^4 \\ \hline \end{array}} \right) =$$

$$= \overset{n_0}{\begin{array}{|c|c|c|c|}\hline n_0^1 & n_0^2 & n_0^3 & n_0^4 \\ \hline \end{array}} \cdot \overset{q_a^0}{\begin{array}{|c|c|c|c|}\hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}},$$

где $'$ — операция транспонирования матриц; аналогично и для n_1—n_3

$$n_1' = \left(\overset{q_b^1}{\begin{array}{|c|c|c|c|c|c|}\hline 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ \hline \end{array}} \cdot \overset{n_1}{\begin{array}{|c|}\hline n_1^1 \\ \hline n_1^2 \\ \hline n_1^3 \\ \hline n_1^4 \\ \hline n_1^5 \\ \hline n_1^6 \\ \hline \end{array}} \right)' =$$

$$\overset{n_1}{\begin{array}{|c|c|c|c|c|c|}\hline n_1^1 & n_1^2 & n_1^3 & n_1^4 & n_1^5 & n_1^6 \\ \hline \end{array}} \cdot \overset{q_b^1}{\begin{array}{|c|c|c|c|c|c|}\hline 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ \hline \end{array}};$$

$$n_2' = \left(\overset{q_c^2}{\begin{array}{|c|c|c|c|}\hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}} \cdot \overset{n_2}{\begin{array}{|c|}\hline n_2^1 \\ \hline n_2^2 \\ \hline n_2^3 \\ \hline n_2^4 \\ \hline \end{array}} \right)' =$$

$$= \overset{n_2}{\begin{array}{|c|c|c|c|}\hline n_2^1 & n_2^2 & n_2^3 & n_2^4 \\ \hline \end{array}} \cdot \overset{q_c^2}{\begin{array}{|c|c|c|c|}\hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array}};$$

$$n_3' = \overset{q_d^3}{\begin{array}{|c|}\hline 1 \\ \hline \end{array}} \cdot \overset{n_3}{\begin{array}{|c|}\hline n_3' \\ \hline \end{array}}.$$

Возьмем другой базис q_f^0 в пространстве N_0. Каждый вектор базиса q_f^0 может быть представлен линейной комбинацией векторов базиса q_α^0, что можно записать в матричном виде:

$$\begin{array}{|c|c|c|c|}\hline q^0_{f_1} & q^0_{f_2} & q^0_{f_3} & q^0_{f_4} \\ \hline \end{array} = \begin{array}{|c|c|c|c|}\hline q^0_{a_1} & q^0_{a_2} & q^0_{a_3} & q^0_{a_4} \\ \hline \end{array} \times$$

$$\times \begin{array}{|c|c|c|c|}\hline co^1_{.1} & co^1_{.2} & co^1_{.3} & co^1_{.4} \\ \hline co^2_{.1} & co^2_{.2} & co^2_{.3} & co^2_{.4} \\ \hline co^3_{.1} & co^3_{.2} & co^3_{.3} & co^3_{.4} \\ \hline co^4_{.1} & co^4_{.2} & co^4_{.3} & co^4_{.4} \\ \hline \end{array} \qquad (13.6)$$

Здесь $co_{.f}{}^0$ называется матрицей перехода от базиса $q_{.a}{}^0$ к базису $q_{.f}{}^0$. Тогда из соотношения

$$n_0 = n_0^a q^0_{.a} = n_0^f q^0_{.f}$$

при подстановке в него (13.6) получим

$$n_0^a = n_0^f co^a_{.f}. \qquad (13.7)$$

Соотношение (13.7) называется преобразованием координат вектора относительно базиса $q_{\cdot f}{}^0$ в координаты относительно базиса $q_{.a}{}^0$. Преобразование, обратное (13.7), можно записать так:

$$n_0^f = AO^f_{.a} n_0^a,$$

где $AO_{.a}{}^f$ — элементы матрицы, обратной матрице $[co_{.f}{}^a]$.

Аналогичные соотношения можно записать и для преобразований в пространствах N_1—N_3.

Если $n_1 = n_1{}^b q_{\cdot b}{}^1 = n_1{}^r q_{\cdot r}{}^1$, то $n_1{}^b = c1_{\cdot r}{}^b n_1{}^r$ и $n_1{}^r = a1_{\cdot b}{}^r n_1{}^b$, где $A1_{\cdot b}{}^r$ — элементы матрицы, обратной матрице $[c1_{\cdot r}{}^b]$. Если $n_2 = n_2{}^c c_{1\cdot c}{}^2 = n_{.2}{}^m q_{.m}{}^2$, то $n_2{}^c = n_2{}^m c2_{\cdot m}{}^c$ и $n_2{}^m = n_2{}^c A2_{.c}{}^m$, где $A2_{.c}{}^m$ — элементы матрицы, обратной матрице $[c2_{.m}{}^c]$.

Линейной формой называется определенная на векторном пространстве N_0 функция F со значениями $f(n_0)$ в поле R, являющаяся линейной в следующем смысле:

$$f(n_0 + m_0) = f(n_0) + f(m_0), \qquad (13.8)$$

где n_0, $m_0 \in N_0$; $\alpha \in R$.

Линейной форме $f(n_0)$ можно поставить в соответствие объект $u_0 \in u_0$, который будем называть нуль-мерным ковектором, или 0-ковектором. Здесь $u_0 = N_0{}^*$.

Запишем ковектор u_0 в базисах p_0^a и $p_0{}^f$ пространства u^0:

$$u^* = u^0_{.a} p^a_a = u^0_{.f} p^f_0.$$

Если базисы связаны преобразованием

$$p_0^f = A^f_{.a} P^a_0,$$

то можно получить связь координат

$$u^0_{.a} = u^0_{.f} A^f_{.a}. \qquad (13.9)$$

Зададим скалярное произведение соотношением

$$\langle u^0,\ n_0\rangle = \langle u^0_{.a} P^a_0,\ q^0_{.a} n^a_0\rangle = u^0_{.a} \langle P^a_0,\ q^0_{.a}\rangle n^a_0.$$

Очевидно, что оно обладает свойством (13.8) по каждому своему аргументу, т. е. билинейно. Если базис $P_0{}^a$ является двойственным для $q_{\cdot a}{}^0$, т. е.

$$\langle P^f_0,\ q^0_{.j}\rangle = \delta^f_j,$$

то

$$\langle p^a_0,\ q^0_{.a}\rangle = \delta^a_{.a},$$

и тогда матрица $P_0{}^a$ является единичной:

$$\begin{array}{|c|c|c|c|}\hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 1 \\ \hline \end{array} = [P^a_0].$$

Если в скалярном произведении вектор и ковектор берутся в двойственных базисах, то

$$\langle u^0,\ n_0\rangle = u^0_{.a} \delta^a_{.a} n^a_0 = u^0_1 n^1_0 + \ .\ .\ .\ + u^0_4 n^4_0.$$

Исходя из различий в преобразовании координат при переходе к новому базису, вектор называют также *один раз контравариантным тензором*, или (1,0)-тензором, а ковектор — *один раз ковариантным тензором*, или (0,1)-тензором.

Назовем ковектор u^0 двойственным вектору n_0, если

$$u^0 = zn_0, \qquad (13.10)$$

где z — (0,2)-тензор. Запишем (13.10) в базисах $p_0{}^a$ и $p_a{}^0$:

$$u^0_a = z_{ab} n^b_0.$$

После подстановки (13.7) и (13.9) в (13.10) получим

$$u^0_f = c^{\cdot a}_f z_{ab} c^b_{.f} n^f_0.$$

Каждому вектору n_0 взаимно однозначно соответствует двойственный ему ковектор u^0, причем если $n_0 \neq 0$, то u^0 тоже не равен нулю.

Аналогично можно определить векторное пространство u^1 одномерных ковекторов, двойственное пространству N_1, векторное пространство u^2 двухмерных ковекторов, двойственное пространству N_2, векторное пространство u^3 трехмерных ковекторов, двойственное пространству N_3. Рассмотрим последовательности линейных отображений

$$u^3 \xleftarrow{F3} u^2 \xleftarrow{F2} u^1 \xleftarrow{F1} u^0; \tag{13.11}$$

$$N_3 \xrightarrow{E3} N_2 \xrightarrow{E2} N_1 \xrightarrow{E1} N_0. \tag{13.12}$$

Операторы $F1$ и $E1$ называются двойственными, если

$$\langle u^0,\ n,\ E1\rangle = \langle F1,\ u^0,\ n_1\rangle, \tag{13.13}$$

где $u^0 \in U^0$, $n_1 \in N_1$.

Левая часть (13.13) равна:

$$\langle u^0_{.a} p^a_0,\ (q^1_{.b} n^b_1)\, E1\rangle = u^0_{.a} \langle p^a_0,\ q^0_{.a}\rangle\, E1^a_{.b} n^b_1 = u^0_{.a} E1^a_{.b} n^b_1. \tag{13.14}$$

Правая часть (13.13) равна:

$$\langle F1\,(u^0_{.a} p^a_0),\ q^1_{.b} X1^b\rangle = u^0_{.a} F1^a_{.b} \langle P^b_1,\ q^1_{.b}\rangle\, n^b_1 = u^0_{.a} F1^a_{.b} n^b_1. \tag{13.15}$$

Приравняв (13.14) и (13.15), получим, что

$$E1^a_{.b} = F1^a_{.b}.$$

Аналогично можно определить операторы $F2$ и $F3$, двойственные $E2$ и $E3$, т. е.

$$\langle u^1,\ n_2 E2\rangle = \langle F2 u^1,\ n_2\rangle;$$
$$\langle u^2,\ n_3 E3\rangle = \langle F3 u^2,\ n_3\rangle.$$

Для этих операторов $F2_{.c}{}^b = E2_{.c}{}^b$ и $F3_{.d}{}^c = E3_{.d}{}^c$. Ядром линейного оператора $E1$ называется множество всех элементов $n_1 \in N_1$, для которых

$$n_1 E1 = 0.$$

Образом линейного оператора $E1$ называется множество тех элементов $n_0 \in N_0$, для которых

$$n_0 = n_1 E1,$$

где n_1 пробегает все множество N_1. Если n_0 разложить по векторам базиса $q_{.a}{}^0$, а $n_1 - n_0 q_{.b}{}^1$, то получим

$$n^a_0 q^0_{.a} = (q^1_{.b} n^b_1)\, E1 = (q^1_{.b} E1)\, n^b_1.$$

Так как векторы $q_{.b}{}^1 E1 \leqslant E0$, то можно записать

$$q^1_{.b} E1 = q^0_{.a} E1^a_{.b}.$$

Пусть оператор $E1$ относительно базисов $q_{.a}{}^0$ и $q_{.b}{}^1$ выражается матрицей инцидентности I, тогда

$$\begin{vmatrix} q^1_{.1} & q^1_{.2} & q^1_{.3} & q^1_{.4} & q^1_{.5} & q^1_{.6} \end{vmatrix} = \begin{vmatrix} q^0_{.1} & q^0_{.2} & q^0_{.3} & q^0_{.4} \end{vmatrix} \times$$

$$\times \begin{vmatrix} -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{vmatrix}$$

и преобразование координат примет вид

$$\begin{array}{c} n_0 \\ \begin{vmatrix} n^1_0 \\ n^2_0 \\ n^3_0 \\ n^4_0 \end{vmatrix} \end{array} = \begin{array}{c} E1^a_{.b} \\ \begin{vmatrix} -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{vmatrix} \end{array} \times$$

$$\times \begin{array}{c} n^b_1 \\ \begin{vmatrix} n^1_1 \\ n^2_1 \\ n^3_1 \\ n^4_1 \\ n^5_1 \\ n^6_1 \end{vmatrix} \end{array} \tag{13.16}$$

Аналогично на основе матрицы инцидентности II можно построить матрицу $E2_{.c}^{b}$ для линейного оператора $E2$ и на основе матрицы инцидентности III — матрицу $E3_{.\alpha}^{c}$ для линейного оператора $E3$. Таким образом,

$$\overset{n_1^b}{\begin{array}{|c|}\hline n_1^1\\\hline n_1^2\\\hline n_1^3\\\hline n_1^4\\\hline n_1^5\\\hline n_1^6\\\hline\end{array}} = \overset{E2_{.c}^{b}}{\begin{array}{|c|c|c|c|}\hline 1 & 1 & 0 & 0\\\hline 1 & 0 & 1 & 0\\\hline 1 & 0 & 0 & -1\\\hline 0 & 1 & -1 & 0\\\hline 0 & -1 & 0 & -1\\\hline 0 & 0 & -1 & 1\\\hline\end{array}} \quad \overset{n_2^c}{\begin{array}{|c|}\hline n_2^1\\\hline n_2^2\\\hline n_2^3\\\hline n_2^4\\\hline\end{array}}$$

$$\overset{n_2^c}{\begin{array}{|c|}\hline n_2^1\\\hline n_2^2\\\hline n_2^3\\\hline n_2^4\\\hline\end{array}} = \overset{E3_{.\alpha}^{c}}{\begin{array}{|c|}\hline -1\\\hline 1\\\hline 1\\\hline -1\\\hline\end{array}} \quad \overset{n_1^\alpha}{\begin{array}{|c|}\hline n_3^1\\\hline\end{array}} \qquad (13.17)$$

Как известно, по виду матрицы линейного оператора можно определить базисные векторы подпространств ядра и образа этого оператора. Ядро оператора $E2$ образует подпространство пространства N_2, натянутое на вектор $q_{.4}^{2} \begin{array}{|c|c|c|c|}\hline -1 & 1 & 1 & 1\\\hline\end{array}$, ортогональный к строкам матрицы $E2_{.c}^{b}$. Но, исходя из (13.17), можно сказать, что образ оператора $E3$ является одномерным подпространством пространства N_2 и натянут на вектор $q_{.4}^{2}$. Значит, $\ker E2 = \operatorname{Im} E3$.

Образом оператора $E2$ является подпространство пространства N_1, натянутое на линейно независимые столбцы матрицы $E2_{.c}^{b}$, например это могут быть первые три столбца. Из (13.16) видно, что ядро оператора $E1$ является трехмерным подпространством из N_1. Путем непосредственной подстановки можно легко убедиться, что выбранные три столбца из $E2_{.c}^{b}$ являются базисом для ядра оператора $E1$, т. е. $\ker E1 = \operatorname{Im} E2$.

Таким образом, получаем, что последовательность (13.11) точная.

Последовательность (13.12) также является точной, так как матрицы операторов $F1$ — $F3$ в соответствующих базисах получаются транспонированием матриц операторов $E1$ — $E3$.

Выделим в пространстве U^0 совокупность элементов u^0 со свойством

$$\langle u^0,\ x_0\rangle = 0$$

для любого $x_0 \in A_0 \in N_0$. Такие элементы u^0 называются *аннулирующими подпространство* A_0. Множество всех элементов пространства u^0, аннулирующих подпространство A_0 пространства N_0, есть подпространство B_0 пространства u^0 и называется *аннулятором подпространства* A_0. Верно также, что A_0 является аннулятором B^0. Обозначим: $A_0 = \operatorname{im} E1$, $A_1 = \ker E1$, $B^0 = \ker F1$, $B_1 = \operatorname{im} F1$. Представим пространства N_0, N_l, u^0, u^1 в виде сумм подпространств $N_0 = A_0 \oplus D_0$, $N_1 = A_1 \oplus D_1$, $u^0 = B^0 \oplus G^0$, $u^1 = B^1 \oplus G^1$. Тогда для подпоследовательностей, изображенных на рис. 13.3, справедливо следующее:

B^1 является аннулятором для A_1, и, наоборот A_1 — аннулятор для B^1;

B^1 и D_1 двойственны;

G^1 и A_1 двойственны;

B^0 и A_0 аннулируют друг друга;

G^0 и A_0 двойственны;

B^0 и D_0 двойственны;

B^1 и A_0 двойственны.

Аналогичные соотношения можно получить и при рассмотрении подпоследовательностей, представленных на рис. 13.4. Здесь

$$N_3 = A_3 \oplus D_3; \qquad u^3 = B^3 \oplus G^3;$$

$$N_2 = A_2 \oplus D_2; \qquad u^2 = B^2 \oplus G^2;$$

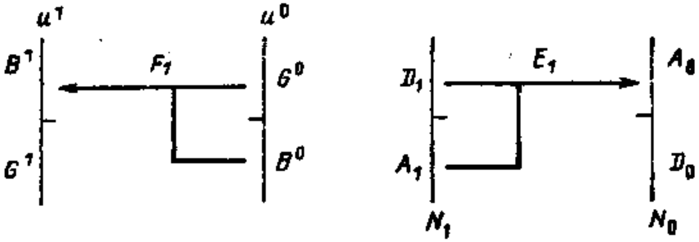


Рис. 13.3. Пояснение к процессу аннулирования подпространств (случай I)

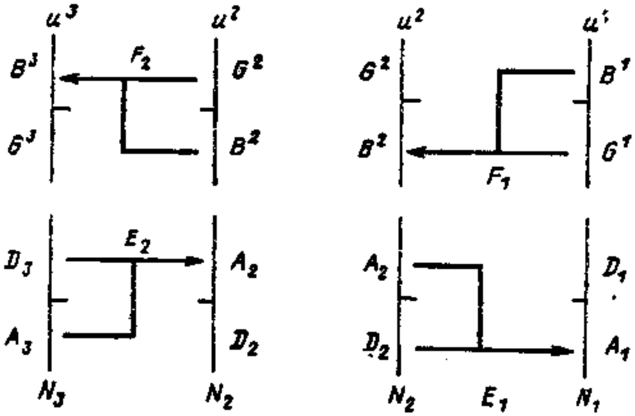


Рис. 13.4. Пояснение к процессу аннулирования подпространств (случай II)

$$N_1 = A_1 \oplus D_1; \qquad u^1 = B^1 \oplus G^1;$$
$$u^3 \leftarrow u^2; \qquad u^2 \leftarrow u^1;$$
$$N_3 \rightarrow N_2; \qquad N_2 \rightarrow N_1;$$

B^2 и A_2 аннулируют друг друга;
B^2 и D_2 двойственны;
B^2 и A_2 двойственны;
B^3 и A_3 аннулируют друг друга;
B^3 и D_3 двойственны;
B^3 и A_3 двойственны.

Представленные в этом параграфе соотношения для цепей и коцепей размерностей 0, 1, 2, 3 можно обобщить и дальше на размерности больше трех.

13.3. ВНЕШНЯЯ АЛГЕБРА ВЕКТОРНОГО ПРОСТРАНСТВА

Внешней (альтернативной или кососимметрической) p-формой называется форма φ, удовлетворяющая условию

$$\varphi(v_1, \ldots, v_\gamma, v_{\gamma+1}, \ldots, v_p) = -\varphi(v_1, \ldots, v_{\gamma+1}, v_\gamma, \ldots, v_p).$$

Для каждого p внешние p-формы образуют действительное векторное пространство $\Lambda^p V^*$, при этом $\Lambda^0 V^* = R$ и $\Lambda^1 V^* = V^*$. Для $\gamma \neq \mu$ имеем

$$\varphi(v_1, \ldots, v_\gamma, \ldots, v_\mu, \ldots, v_p) = -\varphi(v_1, \ldots, v_{\gamma-1}, v_\mu, \ldots, v_{\mu-1}, v_\gamma, \ldots, v_p).$$

Если $v_\gamma = v_\mu$ при $\gamma \neq \mu$, то

$$\varphi(v_1, \ldots, v_\gamma, \ldots, v_\mu, \ldots, v_p) = 0.$$

Для любых натуральных чисел i_1 и i_2 положим

$$\delta(i_1, i_2) = \begin{cases} 1, & \text{если } i_1 < i_2; \\ 0, & \text{если } i_1 = i_2; \\ -1, & \text{если } i_1 > i_2. \end{cases}$$

Функцию $\delta(i_1, \ldots, i_p)$, удовлетворяющую соотношению

$$\delta(i_1, \ldots, i_p) = \Pi\delta(i_\gamma, i_\mu),$$
$$\gamma, \mu = 1, \ldots, p; \ \gamma < \mu,$$

назовем обобщенным символом Кронекера. Кроме того, положим $\delta(i) = 1$. Таким образом, обобщенный символ Кронекера принимает только значения -1, 0, $+1$, причем он равен нулю, если $i_\gamma = i_\mu$ при $\gamma \neq \mu$. Для обобщенного символа Кронекера справедливо тождество

$$\delta(i_1, \ldots, i_\gamma, i_{\gamma+1}, \ldots, i_p) = -\delta(i_1, \ldots, i_{\gamma+1}, i_\gamma, \ldots, i_p).$$

Тогда, если числа $i_1, \ldots, i_p$ записать в их естественном порядке по возрастанию, например $j_1 < j_2 < \ldots < j_p$, то

$$\delta(i_1, \ldots, i_p) = (-1)^a \delta(j_1, \ldots, j_p),$$

где a — число всех необходимых транспозиций пар чисел для перехода от $i_1, \ldots, i_p$ к $j_1, \ldots, j_p$. Так как

$$\delta(j_1, \ldots, j_p) = 1,$$

то

$$\delta(i_1, \ldots, i_p) = (-1)^a.$$

Число $(-1)^a$ называется *четностью перестановки.* Совокупность из p объектов можно упорядочить $p!$ различными способами. Тогда p-набору векторов $(v_1, \ldots, v_p) \in V^p$ соответствуют $p!$ кортежей $(v_{i_1}, \ldots, v_{i_p})$, у которых $v_\gamma \neq v_\mu$ при $\gamma \neq \mu$.

Под альтернативной формой φ понимается некоторая форма $[\varphi]$, определяемая по правилу

$$[\varphi](v_1, \ldots, v_p) = \frac{1}{p!} \sum_{1 < i_1, \ldots, i_p < p} \delta(i_1, \ldots, i_p) \times$$
$$\times \varphi(v_{i_1}, \ldots, v_{i_p});$$
$$i_\gamma \neq i_\mu \quad \text{при} \quad \gamma \neq \mu.$$

Если $p=0$ или $p=1$, то полагаем $[\varphi]=\varphi$. Например, при $p=2$

$$[\varphi](v_1, v_2)=\frac{1}{2}(\varphi(v_1, v_2)-\varphi(v_2, v_1));$$

при $p=3$

$$[\varphi](v_1, v_2, v_3)=\frac{1}{6}(\varphi(v_1, v_2, v_3)+\varphi(v_2, v_3, v_1)+\varphi(v_3, v_1, v_2)-$$

$$-\varphi(v_2, v_1, v_3)-\varphi(v_3, v_2, v_1)-\varphi(v_1, v_3, v_2)).$$

Если форма φ является внешней, то при перестановке двух i одновременно изменяется знак как у δ, так и у φ^{μ}, поэтому значение произведения

$$\delta(i_1, \ldots, i_p)\varphi(v_{i_1}, \ldots, v_{i_p})$$

остается неизменным при транспозиции двух i_μ и для внешней формы $[\varphi]$ выполняется равенство

$$[\varphi](v_1, \ldots, v_\gamma, \ldots, v_\mu, \ldots, v_p)=$$

$$=[\varphi](v_1, \ldots, v_\mu, \ldots, v_\gamma, \ldots, v_p),$$

т. е. альтернативная внешняя форма не изменяется при перестановке двух v_i.

Кроме того, можно получить

$$[[\varphi]]=[\varphi].$$

Действительно, $[[\varphi]](v_1, \ldots, v_p)=\frac{1}{p!}(p!\,[\varphi](v_1, \ldots, v_p))=[\varphi](v_1, \ldots, v_p)$;

$$[\varphi+\psi]=[\varphi]+[\psi];$$

$$[\alpha+\varphi]=\alpha+[\varphi],$$

где $\alpha\in R$.

Назовем произведением форм $\varphi(v_1, \ldots, v_p)$ и $\psi(u_1, \ldots, u_q)$ $(p+q)$-линейное отображение $\varphi\psi: V^{p+q}\to R$, задаваемое формулой

$$(\varphi\psi)(v_1, \ldots, v_p, u_1, \ldots, u_q)=$$

$$=\varphi(v_1, \ldots v_p)\psi(u_1, \ldots, u_q),$$

где v_i и u_j принадлежат V $(i=1, \ldots, p;\ j=1, \ldots, q)$.

Произведения форм удовлетворяют следующим тождествам:

$$(\varphi\psi)\chi=\varphi(\psi\chi);$$

$$(\varphi+\psi)\chi=\varphi\chi+\psi\chi;$$

$$\varphi(\psi+\chi)=\varphi\psi+\varphi\chi;$$

$$a(\varphi\psi)=(a\varphi)\psi=\varphi(a\psi),$$

где $a\in R$.

Связь произведения и альтернирования определяется формулой

$$[\varphi[\psi]]=[[\varphi]\psi]=[[\varphi][\psi]]=[\varphi\psi].$$

Внешним произведением внешней p-формы φ на внешнюю q-форму ψ называется внешняя $(p+q)$-форма, определяемая по правилу

$$\varphi\wedge\psi=\frac{(p+q)!}{p!q!}[\varphi\psi].$$

Внешнее произведение — это отображение декартова произведения $\Lambda^pV^*\times\Lambda^qV^*$ в пространство $\Lambda^{p+q}V^*$, удовлетворяющее следующим свойствам:

$$(\varphi+\psi)\wedge\chi=\varphi\wedge\chi+\psi\wedge\chi,\quad \text{где}\quad \varphi,\ \psi\in\Lambda^qV^*,\ \chi\in\Lambda^pV^*;$$

$$\chi\wedge(\varphi+\psi)=\chi\wedge\varphi+\chi\wedge\psi),\quad \text{где}\quad \varphi,\ \psi\in\Lambda^qV^*,\ \chi\in\Lambda^pV^*;$$

$$\varphi\wedge(\psi\wedge\chi)=(\varphi\wedge\psi)\wedge\chi,\quad \text{где}\quad \varphi\in\Lambda^pV^*,\ \psi\in\Lambda^qV^*,\ \chi\in\Lambda^rV^*;$$

$$a(\varphi\wedge\psi)=(a\varphi)\wedge\psi=\varphi\wedge(a\psi),$$

где $a\in R$; $\varphi\in\Lambda^qV^*$; $\psi\in\Lambda^pV^*$.

Векторное пространство ΛV^* вместе с операцией внешнего произведения образует алгебру над полем R, которая называется *внешней алгеброй пространства* V^*.

13.4. ТЕНЗОРНЫЕ АЛГЕБРЫ НА ТОПОЛОГИЧЕСКИХ СЕТЯХ

Можно построить тензорные алгебры для рассматриваемой цепи: $C(N_0)=R+N_0+N_0\otimes N_0+\ldots+N_0\otimes N_0\otimes N_0\otimes N_0+\ldots$ — тензорная алгебра на пространстве нуль-мерных цепей; $C(N_1)=R+N_1+N_1\otimes N_1+\ldots+N_1\otimes N_1\otimes N_1\otimes N_1\otimes N_1\otimes N_1+\ldots$ — тензорная алгебра на пространстве одномерных цепей; $C(N_2)=R+N_2+N_2\otimes N_2+\ldots N_2\otimes N_2\otimes N_2\otimes N_2+\ldots$ — тензорная алгебра на пространстве двухмерных цепей; $C(N_3)=R+N_3+\ldots$ — тензорная алгебра на пространстве трехмерных цепей;

$T^*(u^0)=R+u^0+u^0\otimes u^0+\ldots+u^0\otimes u^0\otimes u^0\otimes u^0\ \ldots$ — тензорная алгебра нуль-мерных коцепей; $T^*(u^1)=R+u^1+u^1\otimes u^1+\ldots+u^1\otimes u^1\otimes u^1\otimes u^1\otimes u^1\otimes u^1+\ldots$ — тензорная алгебра одномерных коцепей; $T^*(u^2)=R+u^2+u^2\otimes u^2+\ldots+u^1\otimes u^2\otimes u^2\otimes u^2+\ldots$ — тензорная алгебра двухмерных коцепей; $T^*(u^3)=R+u^3+\ldots$ — тензорная алгебра трехмерных коцепей.

В паре двойственных пространств $C(N_0)$ и $T^*(u^0)$ можно задать операцию спаривания:

$$\langle x, y\rangle=\begin{cases} 0, & \text{если}\quad x\in T^p(u^0)\ \text{и}\ y\in C^q(N_0);\\ x\otimes y, & \text{если}\quad x\in T^p(u^0)\ \text{и}\ y\in C^p(N_0);\end{cases}$$

Здесь $x \otimes y$ — операция полной свертки двух тензоров по всем индексам. В координатной форме это можно записать так:

$$\langle x, y \rangle = \sum_{i_1, \ldots, i_p} x_{i_1, \ldots, i_p} y^{i_1, \ldots, i_p}. \tag{13.18}$$

Для четырехмерных пространств N_0 и u^0 значения p могут быть от 0 до 4.

Аналогичные операции спаривания могут быть определены и для остальных пар: $C(N_1)$ и $T^*(u^1)$, $C(N_2)$ и $T^*(u^2)$, $C(N_3)$ и $T^*(u^3)$.

Линейные отображения (13.12) индуцируют линейные отображения алгебр контравариантных векторов

$$C^p(N_3) \xrightarrow{C^p(E_3)} C^p(N_2) \xrightarrow{C^p(E_2)} C^p(N_1) \xrightarrow{C^p(E_1)} C^p(N_0). \tag{13.19}$$

Связь (13.11) с (13.12) является ковариантным функтором. Аналогично ковариантный функтор

$$u^3 \xleftarrow{F_3} u^2 \xleftarrow{F_2} u_1 \xleftarrow{F_1} u^0; \tag{13.20}$$

$$T^p(u^3) \xleftarrow{T^p(E_3)} T^p(u^2) \xleftarrow{T^p(E_2)} T^p(u^1) \xleftarrow{T^p(E_1)} T^p(u^0) \tag{13.21}$$

связывает отображения пространств коцепей с отображениями между алгебрами на этих пространствах. Операция спаривания позволяет ввести отношение двойственности между элементами $C^p(N_i)$ и $T^p(u^i)$. Соотношения (13.19) и (13.21) образуют контравариантный функтор с отношениями сопряженности между операторами:

$$\langle y^*, xC^p(E_1) \rangle = \langle T^p(E_1) y^*, x \rangle,$$

где $y_{i_1, \ldots, i_p}$ и $x^{i_1, \ldots, i_p}$ — компоненты тензоров $y^* \in T^p(u^0)$ и $x \in C^p(N_1)$;

$$\langle u^*, vC^p(E_2) \rangle = \langle T^p(F_2) u^*, v \rangle,$$

где $u_{j_1, \ldots, j_p}, v^{j_1, \ldots, j_p}$ — компоненты тензоров $u^* \in T^p(u^1)$ и $v \in C^p(N_2)$; аналогичное соотношение можно получить для сопряженных операторов $C^p(E_3)$ и $T^p(F_3)$.

Задав значения p от 1 до 6, получим взаимосвязи многомерных тензоров на данной топологической сети.

Глава 14

АФФИННАЯ, ПРОЕКТИВНАЯ И ЕВКЛИДОВА ГЕОМЕТРИИ

14.1. ВВЕДЕНИЕ В ГЕОМЕТРИЮ

В последнее время появился очень удачный термин «машинная геометрия», который именует новое научно-техническое направление, связанное с геометрическими исследованиями, геометрическим моделированием с помощью ЭВМ. Пространства размерности более трех и некоторые виды пространств размерностей 3 и 2 не могут быть интерпретированы как физические пространства. Поэтому очень удобной средой для их исследования является компьютер, с помощью которого можно осуществлять эксперименты в пространствах такого вида, невозможные в физическом мире. В частности, можно осуществить переход от трехмерного пространства в четырехмерное, вкладывая в него все объекты трехмерного мира, манипулируя ими в четырехмерном мире и затем совершая обратный переход в трехмерное пространство.

Такие средства позволяют создавать совершенно новые методы конструирования, применять такие приемы, которые неосуществимы обычными методами. В связи с этим нас интересуют в геометриях прежде всего преобразования одних пространств и геометрий в другие, одних геометрических объектов в другие в одном или другом пространстве. Этот «динамический» (в отличие от «статического») подход к геометриям определил стиль их изложения, основанный на тензорных и дифференциальных формах (дифференциальная геометрия) с использованием теории инвариантов.

Вторая особенность — алгебраический стиль изложения и широкая физическая интерпретация, приводящая к геометризации физики,— явление (геометрическое или кибернетическое), которое можно наблюдать и исследовать только на компьютере, так как с физической точки зрения безразмерные пространства и геометрии (так же как метрические с числом измерений более трех) как физические пространства не существуют.

Третья особенность рассмотрения геометрий — это их алгебраический и дискретный характер (см. «Алгебраическая геометрия», «Дискретная геометрия» в [97—100]). Строго говоря, нас, как кибернетиков, главным образом интересуют дискретные пространства и геометрии, так же как топологии и топологические пространства. Общность тензорного подхода к рассмотрению геометрий заключается в том, что он позволяет рассматривать дискретные пространства наряду с гладкими многообразиями и римановыми пространствами.

В данной главе кратко рассматриваются четыре геометрии: аффинная, проективная, евклидова, риманова, причем три последние геометрии излагаются с использованием понятий дифференциальной формы, тензора, методов дифференциальной геометрии.

Приведенный выше перечень и заданная в нем последовательность геометрий характерны и удобны для алгебраического подхода, где нет требования непрерывности пространства (оно может быть дискретным). При алгебраическом подходе, возможно, удастся с единых позиций рассмотреть геометрические и топологические пространства. Для метода, использующего дифференциальную геометрию, более удобны следующие последовательности:

евклидова геометрия, аффинная геометрия, проективная геометрия;

риманова геометрия, геометрия аффинной связности, геометрия проективной связности.

Дадим грубое поверхностное определение этих геометрий.

Аффинная геометрия — это теория тех свойств фигур и тел, которые остаются неизменными при преобразованиях, в результате которых сохраняется коллинеарность точек и переводятся параллельные прямые в параллельные прямые (т. е. при аффинных преобразованиях).

Это позволяет рассматривать отдельные геометрические объекты (шары, многогранники и т. д.) в процессе их аффинных преобразований.

Проективное пространство характеризуется отсутствием параллельных прямых. Истоками *проективной геометрии* явились графика и архитектура. В конце XIX века исследования по проективной геометрии и по основаниям геометрии объединились, включив в единую схему, в рамках проективной геометрии, геометрии Евклида, Лобачевского и Римана. Проективное пространство не является метрическим, но иногда может быть метризовано введением в него проективной метрики.

Аналитическая геометрия описывает геометрические объекты с помощью алгебраических уравнений, алгебраическая геометрия устанавливает связи между системами алгебраических уравнений и геометрическими объектами (переход от системы алгебраических уравнений к геометрическому объекту).

Когда рассматривается пространство, состоящее из полиэдров (симплексное пространство), осуществляется переход к топологическому пространству (топологии).

Обобщением аффинного (проективного) пространства является пространство аффинной (проективной) связности. Геометрия пространства аффинной связности может выступать как геометрия путей и быть безразмерной. При определенных условиях пространство аффинной связности может быть превращено в риманово пространство, частным случаем которого является евклидово пространство (когда тензор кривизны тождественно равен нулю).

14.2. ОСОБЕННОСТИ АЛГЕБРАИЧЕСКОГО ИЗЛОЖЕНИЯ ГЕОМЕТРИИ

Таким образом, если в аналитической геометрии получают аналитическое, алгебраическое выражение для различных геометрических объектов, то при алгебраическом подходе (в алгебраической геометрии, геометрической алгебре) стараются установить соответствие геометрических объектов заданным алгебраическим конструкциям (выражениям) [96].

Если X — произвольное множество, G — группа, то действием группы G на множестве X называется гомоморфизм $\varphi : G \to G_x$. Говорят, что G действует в X (посредством φ). Через G_x обозначается группа всех биекций $f' : x \to X$ или группа (относительно композиции) его (X) перестановок. Если A и B — два подмножества E, то пользуемся обозначением

$$A \mid B = \{x \in E : x \in A \quad \text{и} \quad x \in B\}.$$

Если $f : E \to F$ — некоторое отображение и $A \subset E$, то через $f|A$ обозначается сужение f на A. Через I_{dx} обозначается тождественное отображение множества X на себя.

Через N обозначается множество всех неотрицательных целых чисел, через Z — множество всех целых чисел, через R — поле вещественных чисел. Через C и H обозначим соответственно поле комплексных чисел и тело кватернионов. Символом R_+ (соответственно R_-) обозначается множество неотрицательных (неположительных) вещественных чисел, а $R_+{}^*$ (соответственно $R_-{}^*$) — то же множество без нуля. Если K — некоторое поле, то через K^* обозначается это же поле без нуля.

Композиция отображений обозначается знаком 0.

Для записи действия φ группы G на множестве X используют обозначения

$$\varphi(g)(x) = g(x) \quad \text{для} \quad g \in G, \quad x \in X \quad \text{и} \quad (G, X, \varphi) \quad \text{или} \quad (G, X).$$

В частности, $\forall g \in G : x \to g(x)$ есть биекция $\forall g,\ h \in G$, $\forall x \in \in X : g(h(x)) = (gh)(x)$.

Например, $l(x) = x$ для всякого x и единицы l группы G; кроме того, $\varphi(g^{-1}) = (\varphi(g))^{-1}$.

Пример 14.1. $G \subseteq G_x : G$ — подгруппа G_x (может выделяться некоторыми условиями).

Пример 14.2. Пусть $A = \{1, \ldots, n\}$, $G_A = G_n$ (симметричная группа), A_n — ее подгруппа, образованная четными перестановками A (знакопеременная группа) и $G = G_n$. Тогда G действует в A, но она действует также и в $X=\mathcal{P}_{h}, p=\{p\subset A : \# p=p\}$, т. е. в множестве всех P-элементных подмножеств из A $(0<p<n)$. Через $\neq x$ обозначается мощность (кардинальное число) множества X.

Пример 14.3. Для заданного $g\in G_n$ положим $G=\{g^k : K\in z\}\subset G_n$. Тогда G действует в $A=X$.

Пример 14.4. Если X — векторное пространство, то в нем действует линейная группа

$$G=GL(X)=\{f : X\to X;\ f : \text{— линейная биекция}\}.$$

Через H_{om} (...) и J_{som} (...) обозначаем соответственно множество гомоморфизмов и изоморфизмов некоторого множества, наделенного какой-нибудь алгебраической структурой, в другое множество, наделенное той же структурой. В случае векторных пространств множество линейных отображений из E в H будет обозначаться не $H_{om}(E;\ F)$, а $L(E',\ g)$ и будет рассматриваться линейная группа векторного пространства — $J_{som}(E;\ E)=$ $=GL(E)$.

Пример 14.5. Пусть E — евклидово векторное пространство и $O(E)=$ $=\{f\in GL(E) : f \text{ — изометрия}\}$. Тогда $G=O(E)$ естественным образом действует в $X=E$. Но, кроме того, $O(E)$ индуцирует действия групп в грассманианах $X=G_{E,\,p}=\{V\subset E : V \text{ — векторное подпространство из } E \text{ размерности } p\}$, а также в множестве $X=$ {ортонормированные базисы в пространстве E}.

Пример 14.6. Пусть $X=G$ — группа; тогда $G=G$ действует в себе самой несколькими способами, причем каждый из них имеет важное значение:

$\varphi(g)(h)=gh$ (левые сдвиги);

$\varphi(g)(h)=hg$ (правые сдвиги);

$\varphi(g)(h)=ghg^{-1}$ (внутренние автоморфизмы).

Пример 14.7. Пусть X — аффинная плоскость (см. гл. 2), $GA(x)=G$ — группа аффинных биекций X (аффинная группа плоскости X) и $\Gamma(X)$ — множество конических сечений X, определяемых как некоторые подмножества X или с помощью их уравнений. Тогда $GA(X)$ действует в $\Gamma(X)$.

Пример 14.8. Пусть X — произвольное метрическое пространство с расстоянием d. Тогда в нем действует его группа изометрий:

$$I_s(X)=I(X);$$

$$G=I_1(X)=\{f\in G_x : \forall x,\ y\in X : d\,(f(x) : (f(y))=d\,(x,\ y)\}.$$

Пример 14.9. Пусть $G=R$, $X=S^3=\{X\in R^4 : \overline{\|X\|=1}\}\subset R^4$. Отождествим R^4 с C^2 и определим действие $(R,\ S^3,\ \varphi)$ соотношением $\varphi(t)(z,\ z')=(e^{it}z,$ $e^{it}z')$.

Эффективность. Действие $(G,\ X,\ \Pi)$ называется эффективным, если Π инъективно (т. е. если один и только один элемент $e\in G$ приводит к тождественному преобразованию на X). Это всегда выполняется, если $G\subset G_X$ (см. пример 14.1). Если действие G неэффективно, то при необходимости действие $(G/K_{er}\varphi,$ $X,\ \varphi)$ будет эффективным. В ранее приведенных примерах только случаи 14.6 и 14.9 не являются эффективными априори.

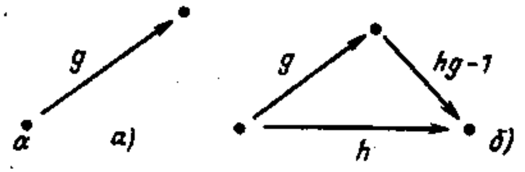


Рис. 14.1. Пояснение к понятию транзитивности

Для примера 14.6 как левые, так и правые сдвиги эффективны, а внутренние автоморфизмы, напротив, эффективны в том и только в том случае, если центр Z_G группы G сводится к $\{e\}$. В примере 14.9, как будет показано далее,

$$K_{er}\varphi=2\pi z.$$

Транзитивность. Действие $(G,\ X,\ \varphi)$ называется транзитивным, если

$$\forall x,\quad \forall y\in X : \exists g\in G\,|\,g(x)=y.$$

Практически достаточно проверить, что

$$\forall a\in X\,|\,\forall x\in X : \exists g\in G\,|\,g(a)=x.$$

Действительно (рис. 14.1), если $g(a)=x$, $h(a)=y$, то $y=$ $=(hg^{-1})(x)$.

Действие в примере 14.2 транзитивно, в примере 14.3 транзитивно в том и только в том случае, если перестановка g циклическая, в примере 14.4 нетранзитивно, в примере 14.5 нетранзитивно в E, но транзитивно во всех $G_{E,\,p}$, в примере 14.6 сдвиги транзитивны, а внутренние автоморфизмы нет (если G коммутативна, то каждая точка остается неподвижной), в примере 14.7 нетранзитивно (парабола — это не эллипс), в примере 14.8 нетранзитивно. Действие $(G_1,\ X)$ называется просто транзитивным, если для всех $x, y\in X$ существует единственное $g\in G$, такое, что $g(x)=y$.

Пример 14.10. Если абелева группа G действует эффективно и транзитивно, то она действует просто транзитивно. Действительно, если $\exists\ x,\ g,$ $h(g(x))=h(x)$, то $g(y)=h(y)\ \forall\ y\in X$, так как $g(y)=g(k(x))=k(\overline{g}(x))=$ $=k(g(x))=h(k(x))=h(y)$.

Левые сдвиги в примере 14.6 просто транзитивны.

Пример 14.11. Действие ортогональной группы $O(E)$ на $G_{E,\,p}$ не является просто транзитивным при $O<p<\dim E$ и, напротив, является таковым на множестве ортонормированных базисов. $(G,\ X)$ называется P-транзитивным $(p\in N)$, если действие G транзитивно на множестве наборов из P различных точек, принадлежащих X.

Стабилизаторы. Однородные пространства. Рассмотрим случаи, когда условие просто транзитивности нарушено. Прежде всего, заметим, что если $y=g(x)=h(x)$, то $(h^{-1}g)(x)=x$. Пусть задано действие (G, X).

Стабилизатором (или *группой изотропии*) элемента $x\in X$ называется подгруппа $G_x=\{g\in G: g(x)=X\}$.

Пример 14.12. Если в примере 14.2 положить $X=A$ и $X=1$, то возникнет естественный изоморфизм $G_1\cong G_{n-1}$.

Если $X=\Phi_{n,p}$, то $G_{\{1,\ldots,p\}}\cong G_p x G_{n-p}$. Эти изоморфизмы позволяют найти число перестановок и число сочетаний: $\# G_n=n!$, $\#\Phi_{n,p}=\binom{n}{p}=G_n{}^p$.

В примере 14.6 для внутренних автоморфизмов имеем

$$G_g=\{h\in G;\quad hg=gh\},$$

т. е. G_g — это коммутатор элемента g.

В примере 14.5 для $G_{E,p}$ получается изоморфизм $G_V\cong\cong O(p)xO(n-p)$ (где $n=\dim E$), а $O(p)=O(R^p)$.
Всегда справедливо соотношение

$$G_g(X)=gG_xg^{-1},$$

т. е., иначе говоря, $Gg(x)$ и G_X — *сопряжение* (посредством внутреннего автоморфизма) подгруппы группы G, в частности они всегда изоморфны. Множество X называется *однородным пространством* (относительно действия группы G), если существует транзитивное действие (G, X, φ). Пусть X однородно относительно (G, X, φ). Для фиксированного $x\in X$ определим отображение $\theta: G\to X$ соотношением $\theta(g)=g(x)$.

Так как равенство $g(x)=h(x)$ эквивалентно тому, что $h^{-1}g\in\in GX$, это означает, что θ переносится на фактор G/G_x группы G, порожденный отношением эквивалентности $\mathcal{R}$:

$$g\mathcal{R}h\Leftrightarrow h^{-1}g\in Gx,$$

в соответствии с $\mathcal{R}$-схемой рис. 14.2.

Заметим, что G/G_x — вообще говоря, не группа. Построенное $\theta: G/G_x\to X$ биективно. Отсюда получаем следующее теоретико-множественное соотношение:

$$X_{\text{множ}}\cong G/G_x. \qquad (14.1)$$

Это соотношение позволяет свести изучение множества X к алгебраической задаче — исследованию пары (G, G_x).

Следствие. Если G конечно, то и X конечно, причем $\neq X=(\neq G)/(\neq G_x)$ (для любого $x\in X$). Если G и X — топологические пространства, а φ — непрерывное отображение из $G\times X$ в X, то топологическое пространство X, вообще говоря, не гомеоморфное пространству G/G_x, снабжено фактор-топологией.

$$\begin{tikzcd}
G \arrow[rr, "\theta"] \arrow[dr, "p"'] & & X \\
& G/G_x \arrow[ur, "\underline{\theta}"'] &
\end{tikzcd}$$

Рис. 14.2. Пояснение к свойству однородности пространства

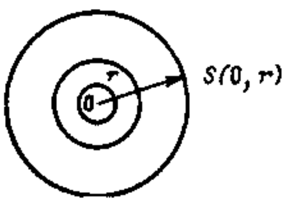


Рис. 14.3. Пояснение к примеру 14.13

Орбиты, формула числа классов. Рассмотрим случаи, когда свойство транзитивности не выполняется.

Определение 14.1. Пусть задано (G, X); орбитой $O(x)$ элемента $x\in X$ называется множество

$$O(x)=\{g(x): g\in G\}.$$

Вводя на X отношение эквивалентности $xRy\Longleftrightarrow\exists g\in G/g(x)=y$, видим, что его классы эквивалентности — это орбиты и что, следовательно, орбиты определяют разбиение пространства X и порождают пространство орбит X/G. Из (14.1) получим, что

$$\forall x\in X: O(x)\underset{\text{множ}}{\cong} G/G_x. \qquad (14.2)$$

Следствие. Если G конечна, то для любого X конечна и орбита $O(x)$, причем

$$\# O(x)=(\# G)/(\# G_x).$$

Пример 14.13. Для случая в примере 14.3 орбиты называются *циклами перестановки* g. Исследование циклов позволяет дать полную классификацию сопряженных элементов в G_n. Для примера 14.5 орбиты представляют собой сферы с центром в начале координат: $S(o, r)$, $r\in R_t$ (рис. 14.3).

Множество M орбит в примере 14.9 будет множеством всех окружностей [так как $S^1=R(2\pi z)$]. Кроме того, они попарно сцеплены и соответствующее фактор-пространство гомеоморфно сфере S^2. Это разбиение S^2 на окружности называется расслоением $X_{\text{он}}$ [84]. Если задано действие (G,X), то нахождение орбит и параметризацию множества X/G можно часто трактовать как задачу *классификации.*

Следствие (формула числа классов). Пусть X и G конечны и G действует в X. Если множество $A\subset X$ пересекается с каждой орбитой заданного действия в одной и только одной точке, то

$$\# X=\sum_{x\in A}\# G/\# Gx.$$

Такое множество A называется *сечением проекции* $p: X\to X/G$, так как имеет место отображение $s: X/G\to X$ со свойствами

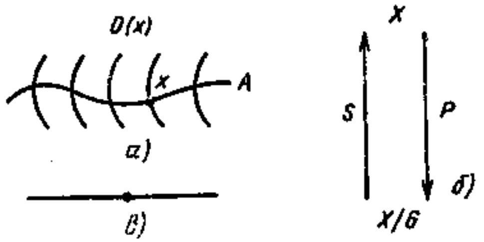


Рис. 14.4. Пояснение к параметризации орбит

$p \bigcirc s = Id x/G$ и $A = s(x|G)$. Можно считать, что сечение A параметризует множество орбит (рис. 14.4).

Р-группы. Определение. Группа G называется p-группой, если она конечна и $\# G = p^m$, где p — некоторое простое число, а $m \in \in N^*$.

Теорема 14.1. *У всякой р-группы есть нетривиальный центр.*

Действительно, пусть G действует в G с помощью внутренних автоморфизмов (см. пример 14.6). Обозначим через z_G центр группы G и рассмотрим действие, индуцированное группой G в $X = G/Z$. Параметризуя X/G с помощью некоторого сечения A, получаем

$$\# X = \sum_{x \in A} \# G / \# G_x,$$

откуда

$$\# G = p^m = \# Z_G + \sum_{x \in A} \# G / \# G_x.$$

Так как кардинальное число $\neq G_x$ делит p^m, оно совпадает с некоторой степенью p и $\# G / \# G_x = p^{m(x)}$, $m(x) \geqslant 1$. Но тогда p делит $\# G$ и $\# X$, а следовательно, и $\# ZG$.

Группа замощений. Имеется в виду, прежде всего, задача о покрытии (без пробелов) плоскости плоскими фигурами, составленными из повторяющихся мотивов. Оказывается, существует конечное число способов без повторения, которое равно 17 (ни больше, ни меньше). Задача замощений является основой для комбинаторной (дискретной, алгебраической) топологии. Допустим, мастер-паркетчик заполняет данный участок плоскости стандартными плитками. Вначале для простоты предположим, что паркетчик размещает плитки, не переворачивая их, т. е. узоры на плитках нанесены на одной стороне. Пусть E — евклидова плоскость, а $p \subset E$ — часть этой плоскости, соответствующая стандартной плитке. Необходимо, чтобы P и ее сдвиги заполняли всю плоскость без пробелов и наложений (см. аксиому $GP2$).

Аксиомы группы замощений. Пусть E — евклидова плоскость, P — связный компакт в E с непустым множеством внутренних точек и G — подгруппа группы $I_s^+(E)$ прямых изометрий (движений) плоскости E. Потребуем, чтобы выполнялись следующие два условия:

$$GP1: \bigcup_{g \in G} g(P) = E;$$

$$GP2: \text{ Из } g\left(\mathring{P}\right) \cap h\left(\mathring{P}\right) \neq \varnothing \text{ следует, что } g_*(P) = h(P).$$

Можно показать [99], что с точностью до сопряжения в линейной группе плоскости E существуют только пять таких групп G. Эти группы соответствуют случаю, когда паркетные пометки являются примитивными в том смысле, когда они удовлетворяют более сильной аксиоме, чем $GP2$, а именно из $g(\mathring{P}) \cap h(\mathring{P}) = \varnothing$ следует $g = h$.

14.3. АФФИННАЯ ГЕОМЕТРИЯ

Как уже указывалось, аффинная геометрия изучает такие свойства геометрических фигур и тел, которые остаются при преобразованиях, в результате которых сохраняется коллинеарность точек и переводятся параллельные прямые в параллельные. Поэтому элементарные свойства аффинных пространств и аффинные подпространства представляют более-менее замаскированные свойства из линейной алгебры. Большая часть доказательств осуществляется автоматически: для получения результата в аффинном пространстве необходимо с помощью параллельных переносов свести его к векторному случаю, для чего требуется «векторизировать» рассматриваемое аффинное пространство, фиксируя одну из его точек. Все рассматриваемые в данном параграфе пространства являются векторными пространствами над коммутативным телом, т. е. над полем. При одновременном рассмотрении двух векторных или аффинных пространств предполагается, что они построены над одним и тем же полем, за исключением случая, рассмотренного в доказательстве основной теоремы аффинной геометрии. Аффинное пространство над полем K — это тройка $(X, \vec{X}, \Phi)$, где $\vec{X}$ — векторное пространство над K, действующее на X как аддитивная группа эффективно и транзитивно.

Положим $\Phi(\vec{\xi}, x) = X + \vec{\xi}$, $\Phi(\vec{\xi})$ — это *параллельный перенос* на вектор $\vec{\xi}$, будем его обозначать через $\overset{\leftrightarrow}{\xi} X$. Иногда говорят, что $\vec{X}$ — это векторное пространство, присоединенное к аффинному пространству X (рис. 14.5). В соответствии

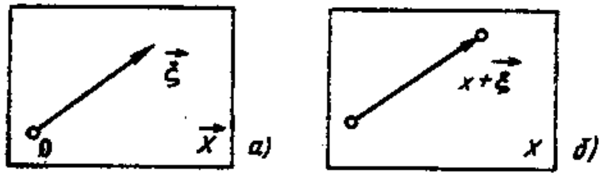


Рис. 14.5. Векторное (а) и аффинное (б) пространства

с примером 14.10 аддитивная группа $\vec{X}$ действует на X просто транзитивно; следовательно, существует такое отображение $\theta : XxX \to \vec{X}$, обозначаемое также через $(x, y) \mapsto \vec{xy}$, что $y = \Phi(\theta(xy), x)$ для всех $x, y \in X$. Можно сказать, что $\vec{xy}$ — свободный вектор, соответствующий закрепленному вектору $(x, \vec{xy})$ или паре точек (x, y). В дальнейшем будет использоваться запись $\vec{xy} = y - x$. Тот факт, что $\vec{X}$ действует в X, выражается с помощью соотношения

$$\left(x+\vec{\xi}\right)+\vec{h}=x+\left(\vec{\xi}+\vec{h}\right),$$

поэтому будет использоваться обозначение $x+\vec{\xi}+\vec{h}$.

Отображение θ, в частности, обладает следующими свойствами:

$$\forall x \in X : \theta x : y \mapsto \theta(x, y) \ \left(\text{биекция } x \to \vec{X}\right);$$

$$\forall x, y, z \in X : \theta(x, y) + \theta(y, z) = \theta(x, z) \text{ (соотношение Шаля)}. \tag{14.3}$$

Эти два свойства легко вытекают из того, что

$$\theta_x^{-1}\left(\vec{\xi}\right) = x + \vec{\xi}$$

— эквивалентное определение. Пусть X — непустое множество; $\vec{X}$ — векторное пространство над K и θ: $XxX \to \vec{X}$ — такое, что при всех x отображение θ_x из (14.3) биективно и $\theta(x, y) + \theta(y, z) = \theta(x, z)$ для всех $\vec{x}$. Тогда X — аффинное пространство, где действие группы определяется по формуле $\Phi(\vec{\xi})(x) = \theta_x^{-1}(\vec{\xi})$, верно и обратное. Действительно,

$$\theta(x, x) = 0; \qquad \theta(y, x) = -\theta(x, y);$$

$$\Phi\left(-\vec{\xi}\right) \circ \Phi\left(\vec{\xi}\right) = Id_x \ \text{ и } \ \Phi\left(\vec{\eta}\right) \circ \Phi\left(\vec{\xi}\right) = \Phi\left(\vec{n}+\vec{\xi}\right).$$

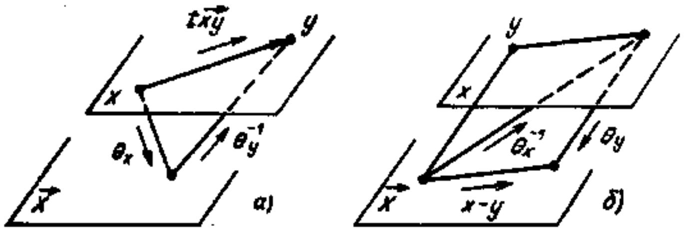


Рис. 14.6. Пояснение к понятию размерности аффинного пространства

Определение. Назовем размерностью аффинного пространства X размерность пространства $\vec{X}$, так что $\dim X = \dim \vec{X}$. Если эта размерность равна 0, то X — точка, если она равна 1, то X называют *аффинной прямой*, а если размерность равна 2, то X — *аффинная плоскость* (рис. 14.6).

Эквивалентное определение, введенное выше, полностью оправдывается последующим материалом. В частности, имеется в виду существование биекций $\theta_x : X \to \vec{X}$. Кроме того, они удовлетворяют соотношениям

$$\theta_y^{-1} \circ \theta_x = \Phi\left(\vec{xy}\right); \qquad \theta_y \circ \theta_x^{-1} : z \mapsto z + $$
$$+ \vec{yx}; \qquad \forall x, y \in X.$$

Определение. Для $a \in \vec{X}$ обозначим через x_a векторное пространство, образованное элементами пространства X и наделенное такой структурой векторного пространства, что $\theta_a : x \to \vec{x}$ порождает естественный изоморфизм между $\vec{X}$ и $\vec{X}_a$. Поэтому говорят, что x_a — векторизация пространства X в точке a.

Аффинные реперы

Пример 14.14. Возьмем $X = \vec{X}$ и положим

$$\Phi\left(\vec{\xi}\right)\left(\vec{h}\right) = \vec{\xi} + \vec{h},\ \vec{\xi},\ \vec{h} \in \vec{X}.$$

Таким образом, на каждом векторном пространстве можно ввести естественную аффинную структуру.

Пример 14.15. Если $(X, \vec{X}, \Phi)$ и $(X', \vec{X}', \Phi')$ — два аффинных пространства, то аффинными пространствами будет и

$$(x \times x', \vec{X} \times \vec{X}', \Phi \times \Phi'),$$

где

$$(\Phi \times \Phi')\left(\vec{\xi}, \vec{\xi}'\right)(x, x') = \left(\Phi\left(\vec{\xi}\right)(x), \Phi'\left(\vec{\xi}'\right)(x')\right).$$

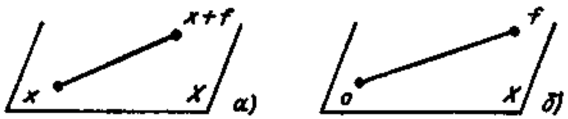


Рис. 14.7. Пояснение к примеру 14.16

Пример 14.16. Пусть E — векторное пространство, $\mathscr{F}$ — его векторное подпространство и X — класс эквивалентности по отношению эквивалентности, связанному с $\mathscr{F}$. Тогда $(X, \mathscr{F}, \Phi)$ — аффинное пространство, где $\Phi(f)(x)=x+f$, $x\in X$, $f\in\mathscr{F}$. В этом примере (рис. 14.7) хорошо иллюстрируется преимущество аффинного пространства, которое заключается в том, что в отличие от ситуации в примере 14.14 множество X не обладает «выделенной» точкой, какой в случае $X=\vec{X}$ была точка 0. Действительно, в случае, когда $\mathscr{F}$ — гиперплоскость в X, задать точку в X — это значит задать некоторое дополнение к $\mathscr{F}$ в E; но, вообще говоря, никакого «естественного дополнения не существует».

В частном случае X может представлять собою множество решений системы линейных алгебраических уравнений $\sum\limits_{j} a_{ij}x_j=b_i$ $(i=1, \ldots, k)$, которое на самом деле можно интерпретировать как прообраз $f^{-1}(B)'$ при подходящем линейном отображении и, следовательно, как класс эквивалентности, связанный с подпространством $\mathscr{F}=f^{-1}(0)$ (ядром f).

Пример 14.17. Пусть $(X, \vec{X}, \Phi)$ — аффинное пространство и $\vec{S}\in\vec{X}$ — векторное подпространство. Это позволяет определить на X отношение эквивалентности

$$R(\vec{S}): xRx' \Leftrightarrow \overrightarrow{xx'} \in S.$$

Тогда $(X/R, \vec{X}/\vec{S}, \Phi)$ — естественным образом превращается в аффинное пространство, где Φ становится отображением фактор-множеств (рис. 14.8).

Пример 14.18. Пусть E — векторное пространство; H — одна из его гиперплоскостей (рис. 14.9); положим $E_H=\{W:$ подпространство в E дополнительно к $H\}$.

Каждое W — это векторная прямая, и, используя обозначения примера 14.5, можно написать $E_H=G_{E,1}|G_{H,1}$.

Отсюда можно утверждать следующее.

Предложение. На E_H существует естественная структура аффинного пространства, для которого присоединенным векторным пространством служит $\vec{E}_H=L(E/H;\ H)$ [для двух векторных пространств через $L(\ldots;\ \ldots)$ обозначается множество всех линейных отображений из первого пространства во второе].

Заметим, что $\dim E/H=1$, откуда

$$\dim\vec{E}_H=\dim H=\dim E-1.$$

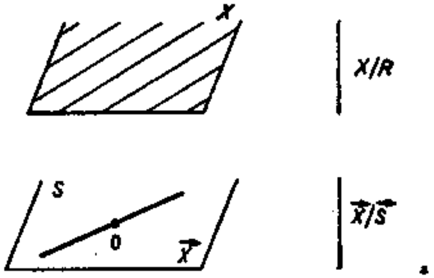


Рис. 14.8. Пояснение к примеру 14.17

Для определения $\theta: E_H\times E_H\rightarrow L(E/H;\ H)$ положим W, $W'\in E_H$, $L\in E/H$. Если $p: E\rightarrow E/H$ — каноническая проекция, то $p^{-1}(\alpha)$ — аффинная гиперплоскость в E (см. пример 14.16). Известно, что сужение p на произвольное дополнение к H представляет собой изоморфизм. Следовательно, можно положить

$$\theta(w,\ w')(\alpha)=(p/w')^{-1}(\alpha)-(p/w)^{-1}(\alpha)\in H.$$

Непосредственно проверяется, что 0 удовлетворяет определениям (см. рис. 14.5). Заметим, что на рис. 14.9

$$w=(p/w')^{-1}(\alpha); \qquad w(p/w)^{-1}(\alpha).$$

Утверждение. Пусть E — векторное пространство, $\mathscr{T}$ — гиперплоскость в нем и X — класс эквивалентности, связанный с $\mathscr{T}$. Тогда между E_H и X существует естественный изоморфизм в смысле аффинных пространств, задаваемых отображением

$$E_H \ni w\rightarrow w\cap X\in X.$$

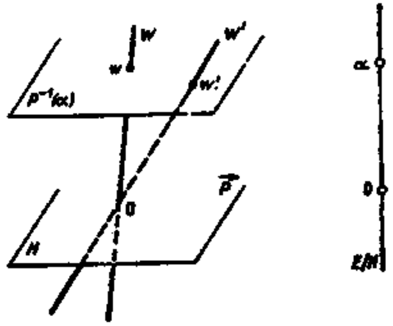


Рис. 14.9. Пояснение к примеру 14.18

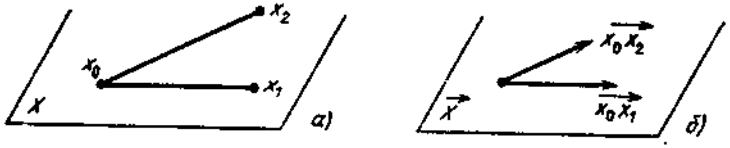


Рис. 14.10. Пояснение к понятию векторов в аффинном пространстве

Для вычислений в аффинном пространстве вводится понятие, которое обобщает понятие базиса векторных пространств в виде реперов, причем об аффинных реперах будем говорить только в случае конечномерных пространств.

Определение. *Аффинный репер* в аффинном пространстве X — это множество из $d+1$ его точек $\{x_i\}_i=0, 1, \ldots, d$, таких, что $\{\overrightarrow{x_0x_1}\}_i=1, \ldots, d$ — базис в X (и, следовательно, $d=\dim\vec{X}$).

Координатами точки $x\in X$ в данном репере называются числа $\{\lambda_i\}$, $i=1, \ldots, d$, такие, что $\overrightarrow{x\cdot x}=\sum\limits_i\lambda_i\overrightarrow{x\bigcirc x_i}$. Таким образом, они совпадают с координатами вектора $\overrightarrow{x\bigcirc x}$ в базисе $\{\overrightarrow{x\bigcirc x_i}\}_i=1, \ldots, d$ пространства $\vec{X}$ (рис. 14.10). Это равносильно тому, что $\{x_i\}$, $i=1, \ldots, d$ — базис векторизации X_{x_0}, а также тому, что $\{x_0, x_1, \ldots, x_d\}$ — симплекс в X.

Морфизм аффинных пространств. Пусть $(X, \vec{X}, \theta)$, $(X', \vec{X}', \theta')$ — два аффинных пространства (над одним и тем же полем) и $f: X\to X'$ — некоторое отображение (априори в смысле теории множеств). Существуют следующие условия эквивалентности.

Утверждение. Для $f: X\to X'$ следующие условия эквивалентны:

1) $\exists a\in X/f\in L(X_a;\quad X'f(a))$;

2) $\forall a\in X: f\in L(X_a;\quad X'f(a))$;

3) $\exists a\in X/\theta f(a)\bigcirc f\bigcirc\theta_a^{-1}\in\alpha(\vec{X};\ \vec{X}')$;

4) $\forall a\in X:\theta f(a)\bigcirc f\bigcirc\theta_a^{-1}\in\alpha(\vec{X};\ \vec{X}')$.

Кроме того, $\theta f(a)\bigcirc f\bigcirc\theta_a{}^{-1}$ зависит только от f; обозначим это отображение через $\vec{f}$ или $\alpha(f)$. Тогда

$$\vec{f}\bigcirc\theta=\theta\bigcirc(f\times f).$$

Отображение, удовлетворяющее одному из этих эквивалентных условий, называется *морфизмом* (аффинных пространств), или *аффинным отображением* из X в X'. Множество всех таких отображений обозначим $\mathrm{A}(X;\ X')$. Говорят, что

$$\begin{tikzcd} X_a \arrow[r, "f"] & X'_{f(a)} \end{tikzcd}$$

а)

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d] & X' \arrow[d] \\ \vec{X} \arrow[r, "\vec{f}"] & \vec{X}' \end{tikzcd}$$

б)

$$\begin{tikzcd} X\times X \arrow[r, "f\times f"] \arrow[d, "\theta"'] & X'\times X' \arrow[d, "\theta'"] \\ \vec{X} \arrow[r, "\vec{f}"] & \vec{X}' \end{tikzcd}$$

в)

Рис. 14.11. Пояснение к понятию морфизмов аффинных пространств

f — *изоморфизм* (аффинных пространств), если f, кроме того, биективно, и *автоморфизм*, если f биективно и $X=X'$ (рис. 14.11).

Утверждение 14.3. Если f — аффинное отображение, то для всех x, $y\in X$

$$\overrightarrow{f(x)f(y)}=f\,(\overrightarrow{xy}),\quad f(y)=f(x)+\vec{f}\,(\overrightarrow{xy}).$$

Дело в том, что на эвристическом уровне f состоит из параллельного переноса и линейного отображения.

Пример 14.19. Если $X=X'=R$, то получим хорошо известные отображения

$$x\to ax+b;\ a,\ b\in R.$$

Пример 14.20. Всякое постоянное отображение f является аффинным, и в этом случае $f=0$. Наоборот, если f — аффинное отображение и $\vec{f}=0$, то f постоянно.

Пример 14.21. Если $f=I_{d_X}$, то $\vec{f}=I_{d_{\vec{X}}}$ (и f аффинно).

Пример 14.22. Если $f\in\mathrm{A}(X;\ X')$; $g\in\mathrm{A}(X';\ X'')$, то $g\bigcirc f\in\mathrm{A}(X;\ X'')$ и $\overrightarrow{g\bigcirc f}=\vec{g}\bigcirc\vec{f}$. В частности, $G\mathrm{A}(X)=\mathrm{A}(X;\ X)\cap G_*$ представляет собой группу относительно композиции, которая называется *аффинной группой* пространства X. Кроме того, существует гомоморфизм

$$L: G\mathrm{A}(X)\in f\to\vec{f}=L(f)\in GL(\vec{X}),$$

где $GL(\vec{X})=L(\vec{X};\ \vec{X})\cap G_*$ — линейная группа пространства $\vec{X}$.

Ядро L, т. е. $L^{-1}(I_{d_{\vec{X}}})$, совпадает, естественно, с X множеством параллельных переносов в пространстве X. Это множество в дальнейшем будет обозначаться как

$$\mathrm{Ker}\,L=T(X);$$

это нормальная подгруппа группы $G\mathrm{A}(X)$.

Пример 14.23. Аффинная группа на аффинных реперах просто транзитивна. Это следует, например, из того, что если $\{x_i\}$, $i=0, \ldots, d$, $\{x_i'\}$,

$i=0, \ldots, d$ — аффинные реперы в X и X', то существует единственное $f\in A(X, X')$, такое, что $f(x_i)=x_i'$, $\forall\, i=0, \ldots, d$.

Пример 14.24. Возможно еще дополнительное утверждение, что X — это однородное пространство для группы $GA(X)$ [более обширной, чем группа $\vec{X}=T(X)$, которая определяла X из определения аффинного пространства].

Обозначим через $GA_a(X)$ стабилизатор точки $a\in X$ в $GA(X)$.

Тогда согласно (14.1)

$$X \cong GA(X)/G_{A_a}(X)$$

(в теоретико-множественном смысле) и сужение $L: GA_a(X)\to\to GL(\vec{X})$ оказывается изоморфизмом групп. Чтобы уточнить, как именно $GA(X)$ получается из $T(X)$ и $GA_a(X)$, введем следующее определение.

Определение. Группа G есть полупрямое произведение своих подгрупп H и K, если $G=HK=\{Lk: h\in H,\ k\in K\}$, $H\cap K=\{l\}$ и H нормальна. Отсюда следует, что всякое представление элемента в виде $g=hk$ единственно, что $G=KH$ также с единственным представлением элементов, но при этом, если $g=hk=k'h$, то, вообще говоря, $k\neq k'$.

Утверждение. Для $a\in X$ существует полупрямое произведение

$$GA(X)=T(X)\,GA_a(X).$$

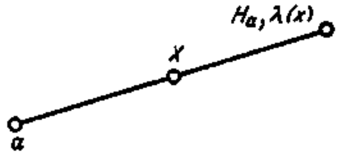


Рис. 14.12. Пояснение к понятию полупрямого произведения

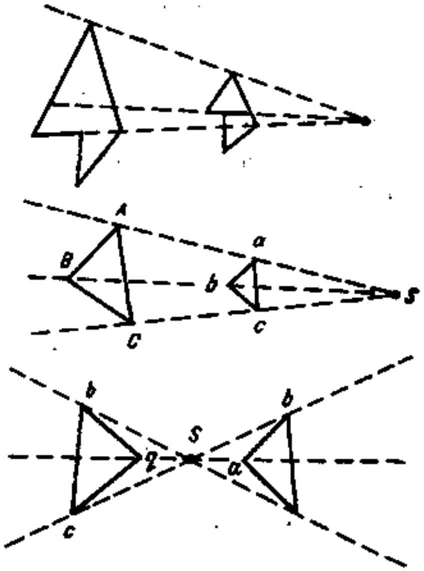


Рис. 14.13. Пояснение к понятию гомотетии

Искомое отображение $GA(X)$ в $T(X)\times GA_a(X)$ имеет, очевидно, вид $f\mapsto(\underset{f(\bar a)_a}{t}\rightarrow,\quad \underset{f(\bar a)_a}{t}\rightarrow\bigcirc f)$. Для $a\in X$, $\lambda\in K^*=K\setminus\bigcirc$ назовем гомотетией с центром a и коэффициентом λ отображение

$$H_{a,\lambda}x\mapsto a+\lambda_{\overline{ax}}$$

[аффинное отображение (рис. 14.12)]; $\vec{H}_{a,\lambda}=\lambda Id_{\vec{x}}$ (используется приведенное выше утверждение). Примеры таких отображений приведены на рис. 14.13. Здесь существенным образом используется коммутативность K.

Утверждение. Пусть отображение $f\in GA(X)$ такое, что $\vec{f}=\lambda Id_{\vec{x}}$ и $\lambda\in K\setminus 1$. Тогда существует единственная точка a, такая, что $f=H_{a,\lambda}$.

Для доказательства этого утверждения целесообразно использовать векторизацию. Для этого в данном случае возьмем произвольное $b\in X$ и произведем вычисление в векторном пространстве X_b. Ищем такое a, чтобы $f(a)=a$; это приводит к соотношению

$$a=f(b)+\lambda(a-b),$$

откуда находим единственное решение

$$a=\frac{1}{1-\lambda}(f(b)-\lambda b).$$

Нетрудно убедиться в справедливости формулы

$$H_{a,\lambda}\bigcirc H_{a,\mu}=H_{a,\lambda\mu},\qquad \forall a\in X,\qquad \forall\lambda,\ \mu\in K. \quad (14.4)$$

14.4. ГЕОМЕТРИЯ ПРОСТРАНСТВ АФФИННОЙ СВЯЗНОСТИ

Дифференциальное исчисление в аффинных пространствах при алгебраическом подходе можно ввести следующим образом. Если X, X' — аффинные пространства, U — открытое множество в X и $f: U\to X'$, то определенные понятия «f дифференцируемого класса C^p» и т. д. можно ввести, если позаботиться, чтобы производная f' отображения f принадлежала $L(\vec{X}, \vec{X}')$. Дифференциальное исчисление может быть введено двумя способами:

с помощью того обстоятельства, что разные векторизации X и X' отличаются только на параллельные переносы и что эти параллельные переносы принадлежат классу C^∞;

с помощью того обстоятельства, что определение дифференцируемости сводится к изучению разности

$$f(x)-f(a)-g(x-a),\qquad g\in L(\vec{X}, \vec{X}'),$$

которая представляет собой некоторое выражение в X, так как $x-a=\overrightarrow{ax}\in\vec{X}$ и $f(x)-f(a)=f(a)f(x)\in X'$. В продолжение евклидовой геометрии (см. гл. 10) приведем дифференциальную квадратичную форму, которая в некоторой специальной системе координат имеет вид

$$ds^2=(dy^1)^2+(dy^2)^2+\ .\ .\ .\ +(dy^n)^2. \tag{14.5}$$

Теорию этой дифференциальной формы будем рассматривать как евклидову геометрию n измерений с фиксированной единицей длины. Каждую координатную систему, в которой квадратичная форма имеет вид, представленный в (14.5), будем называть ортогональной системой координат (рис. 14.14). Совокупность всех таких координатных систем обладает тем свойством, что относительно преобразований от одной такой координатной системы к другой квадратичная форма ds инвариантна. Каждую координатную систему, в которой коэффициенты ds^2 постоянны, будем называть декартовой системой координат [см. (10.2)]. Квадратичная форма инвариантна относительно совокупности и этих координатных систем. Эта совокупность включает в себя ортогональные системы координат в качестве подмножества; именно для ортогональных систем постоянные значения g задаются формулой

$$g_{ij}=\delta^i_j. \tag{14.6}$$

Все компоненты основной аффинной связности дифференциальной формы

$$\Gamma^i_{jk}=\frac{1}{2}\,g^{ia}\left(\frac{\partial g_{ak}}{\partial x^j}+\frac{\partial g_{ja}}{dx^k}-\frac{\partial g_{jk}}{\partial x^a}\right) \tag{14.7}$$

обращаются в нуль в каждой декартовой координатной системе вследствие того, что g^{ia} постоянны. Наоборот, если все Γ равны нулю, то из равенства

$$g_{ij,\,k}=0 \tag{14.8}$$

следует, что

$$g_{ij}=\mathrm{const}. \tag{14.9}$$

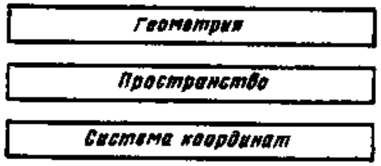


Рис. 14.14. Пояснение к основным геометрическим понятиям

Дадим пояснение к этим формулам. Мы рассматривали три системы координат x, $\bar{x}$, $\tilde{x}$ или три точки x, $\bar{x}$, $\tilde{x}$ в одной системе координат. Введем квадратичную форму как однородную квадратичную форму, компоненты которой имеют вид

$$g_{ij}dx^idx^j=g_{11}dx^1dx^1+g_{12}dx^1dx^2+\ .\ .\ .\ +g_{1n}dx^1dx^n+\ .\ .\ .$$
$$.\ .\ .\ +g_{n1}dx^ndx^1+g_{n2}dx^ndx^2+\ .\ .\ .\ +g_{nn}dx^ndx^n. \tag{14.10}$$

Коэффициенты этой формы зависят от x. Не нарушая общности, будем считать

$$g_{ij}=g_{ji}. \tag{14.11}$$

Из определения следует, что

$$\bar{g}_{ij}d\bar{x}^id\bar{x}^j=g_{ab}dx^adx^b=g_{ab}\frac{\partial x^a}{\partial\bar{x}^i}\frac{\partial x^b}{\partial\bar{x}^j}d\bar{x}^id\bar{x}^j,$$

и поэтому

$$g_{ij}=g_{ab}\frac{\partial x^a}{\partial\bar{x}^i}\frac{\partial x^b}{\partial\bar{x}^j}. \tag{14.12}$$

Таким образом, в каждой системе координат квадратичная форма определяет систему n^2 функций $g_{ij}(x)$; эти системы функций в двух разных координатных системах связаны законом преобразования (14.12). Следовательно, они являются компонентами ковариантного тензора второй валентности. Из (14.12) следует, что соотношение (14.11) сохраняется в любой системе координат, следовательно, тензор g_{ij} есть симметрический тензор. Наоборот, из того, что g_{ij} есть симметрический ковариантный тензор второй валентности, следует, что $g_{ij}dx^idy^j$ — симметрическая билинейная форма. Это означает, что теория симметрических билинейных дифференциальных форм полностью эквивалентна теории симметрических ковариантных тензоров. Рассмотрим инварианты, компонентами которых являются функции от функций g_{ij} и их первых производных, что приводит нас к понятию основной аффинной связности. Дифференцируя формулу преобразования для g_{ij} (14.12) и их первых производных, получаем

$$\frac{\partial g_{ij}}{\partial\bar{x}^k}=\frac{\partial g_{ab}}{\partial x^c}\frac{\partial x^a}{\partial\bar{x}^i}\frac{\partial x^b}{\partial\bar{x}^j}\frac{\partial x^c}{\partial\bar{x}^k}+g_{ab}\frac{\partial^2x^a}{\partial\bar{x}^i\partial\bar{x}^k}\frac{\partial x^a}{\partial\bar{x}^j}+g_{ab}\times$$
$$\times\frac{\partial x^a}{\partial\bar{x}^i}\frac{\partial^2x^b}{\partial\bar{x}^j\partial\bar{x}^k}. \tag{14.13}$$

Из рассмотрения этих соотношений следует, что нет такого тензора, компонентами которого были бы рациональные функции от первых производных $\bar{g}_{ij}$ или от самих функций g_{ij} и их первых производных. Действительно, закон преобразования такого тензора мог бы быть получен с помощью исключения вторых производных из формулы (14.13) и соотношений

$$\bar{g}_{ij} = g_{ab} u_i^a u_j^b, \tag{14.14}$$

где

$$u_j^i = \frac{\partial x^i}{\partial \bar{x}^j}. \tag{14.15}$$

Но соотношение (14.13) может быть записано в виде системы соотношений, каждое из которых дает явное выражение второй производной через первые. Для получения этой системы сложим соотношения, полученные из (14.13) циклической перестановкой нижних индексов i, j, k, вычтем из этой суммы (14.13) и разделим пополам. В результате получим

$$\frac{1}{2}\left(\frac{\partial \bar{g}_{ik}}{\partial \bar{x}^j} + \frac{\partial \bar{g}_{jk}}{\partial \bar{x}^i} - \frac{\partial \bar{g}_{ij}}{\partial \bar{x}^k}\right) =$$
$$= \frac{1}{2}\left(\frac{\partial g_{ac}}{\partial x^b} + \frac{\partial g_{ab}}{\partial x^a} - \frac{\partial g_{ab}}{\partial x^c}\right) \times$$
$$\times \frac{\partial x^a}{\partial \bar{x}^j} \frac{\partial x^b}{\partial \bar{x}^j} \frac{\partial x^c}{\partial \bar{x}^k} + g_{cb} \frac{\partial x^a}{\partial \bar{x}^k} \frac{\partial^2 x^b}{\partial \bar{x}^i \partial \bar{x}^j} \tag{14.16}$$

Умножая обе части этого равенства на соответствующие части равенства

$$g^{bk} \frac{\partial x^a}{\partial \bar{x}^l} = g^{dc} \frac{\partial \bar{x}^k}{\partial x^c},$$

которое эквивалентно (14.12), получаем

$$\frac{1}{2} g^{lk}\left(\frac{\partial \bar{g}_{ik}}{\partial \bar{x}^j} + \frac{\partial \bar{g}_{ik}}{\partial \bar{x}^i} - \frac{\partial \bar{g}_{ij}}{\partial \bar{x}^k}\right) \times$$
$$\times \frac{\partial x^{\alpha}}{\partial \bar{x}^l} = \frac{1}{2} g^{dc}\left(\frac{\partial g_{ac}}{\partial x^b} + \frac{\partial g_{bc}}{\partial x^a} - \right.$$
$$\left. - \frac{\partial g_{ab}}{\partial x^c}\right) \frac{\partial x^a}{\partial \bar{x}^i} \frac{\partial x^b}{\partial \bar{x}^j} + \frac{\partial^2 x^d}{\partial \bar{x}^i \partial \bar{x}^j}, \tag{14.17}$$

что дает явное выражение каждой второй производной $\dfrac{\partial^2 x^d}{\partial \bar{x}^i \partial \bar{x}^j}$ через величины $u_j{}^i$, g_{ij} и первые производные от метрических тензоров g, $\bar{g}$. Наоборот, формулы (14.13) могут быть получены из (14.17), т. е. эти две системы формул эквивалентны. Очевидно, что из (14.14) и (14.15), не содержащих вторые производные, из (14.17), дающей одно и только одно явное выражение для каждой из вторых производных, невозможно исключить вторые производные. Таким образом, теорема доказана. Комбинации из g функций, появляющиеся в (14.16), называются *трехиндексными символами Кристоффеля первого рода*. Для их записи примем следующие обозначения:

$$\frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right) = \Gamma_{k,\,ij} = [k,\ ij]. \tag{14.18}$$

Комбинации из функций, появляющиеся в (14.17), называются *трехиндексными символами Кристоффеля второго рода*. Для них приняты следующие обозначения:

$$\frac{1}{2} g^{ka}\left(\frac{\partial g^{ia}}{\partial x^j} + \frac{\partial g^{aj}}{\partial x^i} - \frac{\partial g^{ij}}{\partial x^a}\right) = \Gamma_{ij}^k = \left\{ \begin{matrix} k \\ ij \end{matrix} \right\}. \tag{14.19}$$

Теперь уравнения (15.17) могут быть записаны в виде

$$\bar{\Gamma}_{pq}^r = \Gamma_{ij}^k \frac{\partial x^i}{\partial \bar{x}^p} \frac{\partial x^j}{\partial \bar{x}^q} \frac{\partial x^r}{\partial x^k} + \frac{\partial^2 x^k}{\partial \bar{x}^p \partial \bar{x}^q} \frac{\partial \bar{x}^r}{\partial \bar{x}^i}. \tag{14.20}$$

Согласно общему определению инвариантов функции $\Gamma_{jk}{}^i$ представляют собой компоненты инварианта, так как они единственным образом определены в каждой координатной системе и в каких-либо двух системах координат их связывает определенный закон преобразования, выраженный формулой (14.20). В связи с тем геометрическим значением, которое имеет этот инвариант, он носит название *основной аффинной связности квадратичной дифференциальной формы* $g_{ij}dx^i dx^j$. Меняя местами k и l в (14.18) и складывая полученное с (14.18), получаем следующую формулу:

$$\frac{\partial g^{ik}}{\partial x^j} = [k,\ ij] + [i,\ kj] = g_{ka}\Gamma_{ij}^a + g_{ia}\Gamma_{kj}^a. \tag{14.21}$$

Формулы преобразования (14.20) обладают свойством транзитивности. Это означает, что если при переходе от координат $\bar{x}$ к координатам $\tilde{x}$ символы Кристоффеля преобразуются по формулам

$$\tilde{\Gamma}_{ij}^l = \bar{\Gamma}_{ab}^d \frac{\partial x^a}{\partial \tilde{x}^i} \frac{\partial \tilde{x}^b}{\partial \tilde{x}^j} \frac{\partial \tilde{x}^l}{\partial \tilde{x}^d} + \frac{\partial^2 \tilde{x}^d}{\partial \tilde{x}^i \partial \tilde{x}^j} \frac{\partial \tilde{x}^l}{\partial \bar{x}^d}, \tag{14.22}$$

то, подставляя (14.20) в эти формулы, получаем

$$\tilde{\Gamma}_{ij}^l = \Gamma_{ab}^d \frac{\partial x^a}{\partial x^i} \frac{\partial x^b}{\partial x^j} \frac{\partial x^l}{\partial x^d} + \frac{\partial^2 x^d}{\partial \tilde{x}^i \partial \tilde{x}^j} \frac{\partial x^l}{\partial x^d}, \tag{14.23}$$

т. е. при переходе от x к $\tilde{x}$ формулы преобразования Г имеют тот же вид, что и при переходе от x к x и от $\bar{x}$ к $\tilde{x}$.

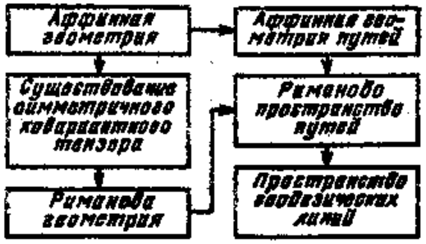


Рис. 14.15. Связь аффинной и римановой геометрий

Теория произвольной аффинной связности называется *геометрией пространств аффинной связности*. Таким образом, теория ковариантного дифференцирования и теория тензора кривизны (см. гл. 10) относятся к геометрии пространств аффинной связности. Объект, полученный после внесения аффинной связности в пространство некоторого числа измерений, называется *пространством аффинной связности*. Когда компонентами аффинной связности являются символы Кристоффеля, т. е. когда основная аффинная связность обладает тем свойством, что существует симметричный ковариантный тензор, удовлетворяющий соотношению

$$g_{ij,k}=0, \qquad (14.24)$$

геометрия путей сводится к римановой геометрии (рис. 14.15). Условия, которым должны удовлетворять для этого компоненты связности, могут быть записаны в виде условий, накладываемых на тензор кривизны и его ковариантные производные. Если существует система координат y, в которой все компоненты аффинной связности обращаются в нуль, то аффинное пространство называется *плоским*. Такое пространство можно рассматривать как эвклидово, так как в этом случае равенство (14.21) в системе координат y принимает вид $\partial g_{ij}/\partial y^k=0$ и, следовательно, компоненты аффинной связности являются символами Кристоффеля некоторой квадратичной формы с постоянными в этой системе координат коэффициентами. Можно показать, что пространство аффинной связности является плоским тогда и только тогда, когда тензор кривизны обращается в нуль.

Геометрия путей. Основным геометрическим образом геометрии аффинной связности является система кривых, определяемая системой дифференциальных уравнений

$$\frac{d^2x^i}{dt^2}+\Gamma^i_{jk}\frac{\partial x^j}{dt}\frac{dx^k}{dt}=0, \qquad i=1,\ 2,\ \ldots,\ n, \qquad (14.25)$$

коэффициенты которых представляют собой компоненты аффинной связности. В случае плоского пространства эти образы — прямые линии. В общем случае они обладают следующим свойством: любые две точки могут быть соединены одной и только одной кривой системы. Поэтому эти кривые называются *путями*, так как они служат для установления пути между двумя достаточно близкими точками в пространстве аффинной связности. В случае риманова пространства пути являются геодезическими линиями, т. е. экстремальными кривыми для интеграла длины

$$\int\sqrt{g_{ij}dx^idx^j}\ .$$

Рассмотрим геометрию путей в целях изучения координатных систем, относительно которых форма (14.25) инвариантна.

Введем преобразование координат

$$\bar{x}^i=f^i(x).$$

Тогда дифференциальные уравнения (14.25) примут вид

$$\frac{d^2\bar{x}^i}{dt^2}+\bar{\Gamma}^i_{jk}\frac{d\bar{x}^j}{dt}\frac{d\bar{x}^k}{dt}=0, \quad i=1,\ \ldots,\ n, \qquad (14.26)$$

где $\bar{\Gamma}$ являются компонентами аффинной связности в системе координат $\bar{x}$, т. е. дифференциальные уравнения (14.25) при преобразовании координат не изменяют своей формы. Более того, если

$$x^i=\varphi^i(t) \qquad (14.27)$$

являются решениями (14.25), то

$$\bar{x}^i=f^i[\varphi^1(t),\ \ldots,\ \varphi^n(t)]$$

будут решениями преобразованных уравнений (11.26) и определяют ту же кривую, что и (14.27). Следовательно, система кривых, заданных уравнениями (14.25), определяется единственным образом, если указана аффинная связность, и является, таким образом, ее инвариантом. Для получения формального решения уравнений (14.25) продифференцируем их последовательно и приведем к следующей последовательности уравнений, которые должны удовлетворяться любым решением уравнений (11.25):

$$\frac{d^3x^i}{dt^3}+\Gamma^i_{abc}\frac{dx^a}{dt}\frac{dx^b}{dt}\frac{dx^c}{dt}=0; \qquad (14.28)$$

$$\frac{d^4x^i}{dt^4}+\Gamma^i_{abcd}\frac{dx^a}{dt}\frac{dx^b}{dt}\frac{dx^c}{dt}\frac{dx^d}{dt}=0 \qquad (14.29)$$

и т. д.

Функции Γ^i_{abc} и т. д. определяются равенствами

$$\Gamma_{abc}=\frac{1}{3}p\left(\frac{\partial\Gamma^i_{ab}}{\partial x^c}-\Gamma^i_{jb}\Gamma^j_{ac}-\Gamma^i_{aj}\Gamma^j_{bc}\right); \tag{14.30}$$

$$\Gamma^i_{abcd}=\frac{1}{4}p\left(\frac{\partial\Gamma^i_{ac}}{\partial x^{dl}}-\Gamma^i_{jbc}\Gamma^j_{ad}-\Gamma^i_{ajc}\Gamma^j_{bd}-\Gamma^i_{abj}\Gamma^j_{cd}\right) \tag{14.31}$$

и т. д., где p обозначает сумму членов, полученную циклической перестановкой свободных индексов в выражении, стоящем внутри скобок. Таким образом, функции Γ^i_{abc} и т. д. симметричны относительно нижних индексов. Если наложим на решения уравнений (14.25) требования, чтобы

$$x^i=g^i \quad \text{и} \quad \frac{\partial x^i}{dt}=a^i, \tag{14.32}$$

когда $t=0$, то из (14.25) и (14.28) и т. д. следует, что решение должно иметь вид

$$x^i=g^i+a^it-\frac{1}{2}(\Gamma^i_{jk})_g a^ja^kt^2-$$

$$-\frac{1}{3!}(\Gamma^i_{jkl})a^ja^ka^lt^3, \tag{14.33}$$

где нижний индекс g показывает, что значения функции, стоящей внутри скобок, должны быть взяты в точке g. Сходимость рядов (14.33) можно доказать с помощью мажорирующих функций. Действительно, это следует из общей теоремы существования решения для системы дифференциальных уравнений вида

$$\frac{dx^i}{dt}=f^i(x,\ t). \tag{14.34}$$

Если учесть, что уравнения (14.26) могут быть переписаны в виде

$$\frac{dx^i}{dt}=z^i; \qquad \frac{dz^i}{dt}=-\Gamma^i_{ab}z^az^b, \tag{14.35}$$

т. е. в виде системы дифференциальных уравнений первого порядка относительно двух переменных $x^1, \ldots, x^n, \ldots, z^1, \ldots, z^n$, и если выполняется преобразование координат

$$x^i-g^i=y^i-\frac{1}{2}(\Gamma^i_{ab})gy^ay^b-$$

$$-\frac{1}{3!}(\Gamma^i_{abc})_g y^ay^by^c, \tag{14.36}$$

то очевидно, что кривые, заданные уравнениями (14.33), определяются в новой системе уравнениями

$$y^i=a^it \tag{14.37}$$

в системе координат $\bar{y}$. Эти выражения имеют вид (14.35) только в случае, если

$$b^i_{jk}=0; \qquad b^i_{jkl}=0. \tag{14.38}$$

Следовательно, преобразование одной аффинной нормальной координатной системы в другую должно иметь вид (14.37). С каждой исходной системой координат x и точкой g можно связать специальную аффинную нормальную систему координат, удовлетворяющую следующим начальным условиям:

$$y=0 \quad \text{и} \quad \frac{dy^i}{dx_j}=\delta^i_j, \quad \text{когда} \quad x=q. \tag{14.39}$$

Если $\bar{y}$ — аффинная нормальная система координат, определенная аналогичным образом относительно другой произвольной системы координат $\bar{x}$, то

$$\frac{\partial\bar{y}^i}{\partial y_j}=\frac{\partial\bar{y}^i}{\partial\bar{x}^a}\frac{\partial\bar{x}^a}{\partial x^b}\frac{\partial\bar{x}^b}{\partial x^j}. \tag{14.40}$$

Беря значение этого выражения в начале координат и используя тот факт, что любые две аффинные нормальные системы, имеющие общее начало, связаны линейным однородным преобразованием с постоянными коэффициентами, придем к выводу, что в нашем случае преобразование одной аффинной нормальной системы в другую осуществляется по формулам

$$\bar{y}^i=\left(\frac{\partial\bar{x}^j}{\partial x_j}\right)y^i. \tag{14.41}$$

Обозначим компоненты аффинной связности в аффинной нормальной координатной системе через $\overset{*}{\Gamma}{}^i_{jk}$. Дифференциальное уравнение путей

$$\frac{d^2y^i}{dt^2}+\overset{*}{\Gamma}{}^i_{jk}\frac{dy^j}{dt}\frac{dy^k}{dt}=0 \tag{14.42}$$

удовлетворяется выражениями

$$y^i=a^it, \tag{14.43}$$

где через a обозначены произвольные постоянные. Подставляя (14.43) в (14.42), получаем

$$\overset{*}{\Gamma}{}^i_{jk}a^ja^k=0, \tag{14.44}$$

что должно иметь место для всех y, принадлежащих пути (14.43). Умножив (14.44) на t^2, можно получить соотношение

$$\Gamma^{*}_{jk}y^jy^k = 0, \tag{14.45}$$

тождественно выполняющееся вдоль пути для всех точек пространства. Разлагая левую часть (14.45) в степенной ряд и подставляя в него выражения из (14.43), получаем

$$0 = \Gamma^{*i}_{jk}y^jy^k = (\Gamma^{*i}_{jk})_0\, a^ja^kt^2 + \left(\frac{\partial\Gamma^{*i}_{jk}}{\partial y^l}\right) a^ja^ka^lt^3 + \dots$$

Отсюда непосредственно следует, что

$$\left.\begin{aligned} &(\Gamma^{*i}_{jk})_0 = 0;\\ &\left(\frac{\partial\Gamma^{*i}_{jk}}{\partial y^l} + \frac{\partial\Gamma^{*i}_{jl}}{\partial y^k} + \frac{\partial\Gamma^{*i}_{kl}}{\partial y^j}\right)_0 = 0 \end{aligned}\right\} \tag{14.46}$$

и вообще

$$S\left(\frac{\partial^m\Gamma^{*i}_{jk}}{\partial x^l \dots \partial x^q}\right)_0 = 0, \tag{14.47}$$

где $S(\quad)$ обозначает сумму всех $(m+2)(m+1)/2$ членов, полученных из выражения, стоящего внутри скобок, заменой пары нижних индексов jk любой парой индексов совокупности j, kl, …, g. Дифференциальные уравнения (14.45) характеризуют полностью аффинные нормальные координаты. Это значит, что если подставим в уравнение (14.45) вместо Γ^{*i}_{jk} их выражения, определяемые законом преобразования компонентов аффинной связности, то получим систему дифференциальных уравнений в частных производных, решения которых определяют преобразования произвольной системы координат к аффинной нормальной системе координат. Точно так же, как выше из (14.25) было выведено (14.45), можно из (14.28), (14.29) и т. д. вывести, что в нормальной системе координат выполняются тождества

$$\Gamma^{*i}_{jkl}y^jy^ky^l = 0; \tag{14.48}$$

$$\Gamma^{*i}_{jklm}y^jy^ky^ly^m = 0. \tag{14.49}$$

Аффинные расширения. Ковариантное дифференцирование (см. гл. 12) можно рассматривать как процесс обыкновенного дифференцирования в аффинной нормальной системе координат с последующим вычислением полученного результата в начале координат. Это допускает следующее непосредственное обобщение. Пусть дана произвольная аффинная связность. Тогда k-м аффинным расширением какого-либо инварианта

$$T^{ij\dots l}_{ab\dots c}$$

будем называть инвариант

$$T^{ij\dots l}_{ab\dots cd\dots f}, \tag{14.50}$$

компоненты которого в какой-либо системе координат и в точке g задаются формулой

$$(T^{ij\dots l}_{ab\dots cd\dots f})_g = \left(\frac{\partial^kT^{*ij\dots l}_{ab\dots c}}{\partial y^d \dots \partial y^f}\right)_0, \tag{14.51}$$

где y — координаты точки в аффинной нормальной координатной системе, определенной аффинной связностью и координатной системой x в точке g, а выражения

$$T^{*ij\dots l}_{ab\dots c}$$

суть компоненты T в координатной системе y. Ниже индексы g и 0 обозначают, что значения взяты соответственно в точках $y=0$. Первое аффинное расширение тензора есть, очевидно, ковариантная производная; k-е аффинное расширение тензора есть также тензор. Действительно, рассмотрим преобразование системы координат x в систему координат $\bar{x}$; оно влечет за собой преобразование (14.41) аффинной системы координат y с началом в точке g в систему координат $\bar{y}$, определяемую системой x в этой точке. Компоненты T в двух нормальных системах координат связаны соотношениями

$$T^{*\pi\dots\sigma}_{\alpha\dots\gamma} = T^{*i\dots l}_{a\dots c}\left(\frac{\partial x^a}{\partial\bar{x}^\alpha}\right)_g \dots \left(\frac{\partial x^c}{\partial\bar{x}^\gamma}\right)_g \times$$
$$\times\left(\frac{\partial\bar{x}^\pi}{\partial x^i}\right)_g \dots \left(\frac{\partial\bar{x}^\sigma}{\partial x^l}\right)_g.$$

После k-кратного дифференцирования по $\bar{y}$, если учесть, что

$$\frac{\partial^k\bar{T}^{*\pi\dots\sigma}_{\alpha\dots\gamma}}{\partial\bar{y}^\delta \dots \partial\bar{y}^\xi} = \frac{\partial^kT^{*i\dots l}_{a\dots c}}{\partial y^d \dots \partial y^f}\left(\frac{\partial x^i}{\partial\bar{x}^\alpha}\right)_g \dots$$
$$\dots \left(\frac{\partial x^l}{\partial\bar{x}^\gamma}\right)_g\left(\frac{\bar{\partial} x^\pi}{\partial x^i}\right)_g \dots$$
$$\dots \left(\frac{\partial\bar{x}^\sigma}{\partial x^l}\right)_g\left(\frac{\partial x^d}{\partial\bar{x}^\delta}\right)_g \dots \left(\frac{\partial x^f}{\partial\bar{x}^\xi}\right)_g, \tag{14.52}$$

получим

$$\frac{\partial y^i}{\partial \bar{y}^j} = \left(\frac{\partial x^i}{\partial \bar{x}^j}\right)_g .$$

Формулы (14.52) показывают, что величины (14.50) образуют собой тензор, имеющий ковариантных индексов на g больше, чем исходный. Из определения ясно, что аффинное расширение симметрично относительно индексов d, l, ..., f в (14.51). Далее, k-е аффинное расширение суммы каких-либо двух тензоров представляет собой сумму k-х аффинных расширений двух тензоров, и k-е аффинное расширение произведения двух тензоров образуется по правилу повторного дифференцирования произведения в обычном анализе.

Аффинные нормальные тензоры. Аффинные расширения аффинной связности в силу их прямой связи с аффинными нормальными координатами сами образуют важную систему тензоров, называемых аффинными нормальными тензорами. Тот факт, что аффинные расширения аффинной связности являются тензорами, следует из того, что закон преобразования Г при переходе из одной нормальной системы координат в другую имеет такой же вид, как и закон преобразования тензоров:

$$\bar{\Gamma}^{*i}_{jk} = \Gamma^{*a}_{bc} \frac{\partial \bar{y}^i}{\partial y^a} \frac{\partial y^b}{\partial \bar{y}^j} \frac{\partial y^c}{\partial \bar{y}^k} .$$

Так как величины $\frac{\partial y^i}{\partial \bar{y}^j}$ постоянны, то, дифференцируя это соотношение, получаем

$$\frac{\partial \bar{\Gamma}^{*i}_{jk}}{\partial \bar{y}^l} = \frac{\partial \Gamma^{*a}_{bc}}{\partial y^d} \frac{\partial \bar{y}^i}{\partial y^a} \frac{\partial y^b}{\partial \bar{y}^j} \frac{\partial y^c}{\partial \bar{y}^k} \frac{\partial y^d}{\partial \bar{y}^l} ,$$

откуда можно сразу заключить, что

$$\bar{\Gamma}^i_{jk,l} = \Gamma^a_{bc,d} \frac{\partial \bar{x}^i}{\partial x^a} \frac{\partial x^b}{\partial \bar{x}^j} \frac{\partial x^c}{\partial \bar{x}^k} \frac{\partial x^d}{\partial \bar{x}^l} .$$

Аналогично получим расширения

$$\Gamma^i_{jk,\,lm}, \quad \Gamma^i_{jk,\,lmp}, \quad \Gamma^i_{jk,\,lmpq} \text{ и т. д.}$$

Из (14.46) следует, что эти тензоры удовлетворяют тождествам

$$\Gamma^i_{jk,\,l} + \Gamma^i_{kl,\,j} + \Gamma^i_{lj,\,k} = 0, \tag{14.53}$$

а в общем случае

$$S(\Gamma^i_{jk,\,l\dots s}) = 0, \tag{14.54}$$

где $S(\ \ )$ имеет тот же смысл, что и в (14.48). Во избежание путаницы с другими системами функций, обозначенных Г с различными распределениями индексов, обозначим нормальные тензоры через A с соответствующими индексами; таким образом, определяем

$$\left.\begin{aligned} A^i_{jkl} &= \Gamma^i_{jk,\,l}; \\ A^i_{jklm} &= \Gamma^i_{jk,lm}. \end{aligned}\right\} \tag{14.55}$$

Степенной ряд, выражающий компоненты аффинной связности в аффинной нормальной системе координат, может быть записан теперь в виде

$$\Gamma^{*i}_{jk} = (A^{*i}_{jka})_0\, y^2 + \frac{1}{2}(A^{*i}_{jkab})_0\, y^a y^b + \\ + \frac{1}{3!}(A^{i*}_{jkabc})_0\, y^a y^b y^c + \dots \tag{14.56}$$

Отсюда следует, что аффинная связность полностью определяется заданием аффинных нормальных тензоров.

Все тождества, которым удовлетворяют аффинные нормальные тензоры, являются следствием тождества (14.54) и соотношений симметрии:

$$A^i_{jkl\dots s} = A^i_{kjl\dots s} n A^i_{jkab\dots c} = A^i_{jkpq\dots s}, \tag{14.57}$$

где $pq\dots s$ означает некоторую перестановку индексов $ab\dots c$. Действительно, если в некоторой точке зададим произвольные значения величин

$$A^i_{jkl}, \quad A^i_{jklm}\dots, \tag{14.58}$$

которые удовлетворяют лишь условиям (14.54) и (14.57) и такие, что ряд (14.56) сходится, то аффинная связность определяется рядом (14.56) и координаты y, в которых она записывается, являются нормальными, так как они удовлетворяют соотношению (14.45). Так как условия сходимости ряда имеют форму неравенств, то величины (14.58) могут быть выбраны так, что они будут удовлетворять лишь таким заданным соотношениям, которые являются следствиями (14.54) и (14.57). Этот факт можно выразить в виде утверждения, что (14.57) и (14.54) образуют полную систему тождеств.

14.5. ГЕОМЕТРИЯ ПРОЕКТИВНОГО ПРОСТРАНСТВА

Аффинная геометрия изучает свойства фигур и тел, сохраняющиеся при аффинных преобразованиях, т. е. при преобразованиях, сохраняющих коллинеарность точек и переводящих параллельные прямые в параллельные (рис. 14.16). Хотелось

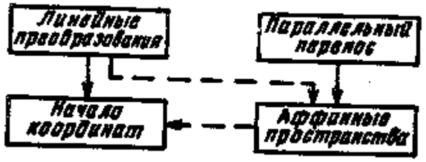


Рис. 14.16. Пояснение к свойствам аффинной геометрии

бы иметь геометрию, изучающую свойства фигур, сохраняющихся при любых преобразованиях и переводящих коллинеарные точки в коллинеарные (но необязательно переводящих параллельные прямые в параллельные). Такая геометрия называется *проективной* и является обобщением аффинной, а соответствующие преобразования называются проективными (рис. 14.17). Имеются средства естественного пополнения аффинного пространства до проективного (рис. 14.18). В частности, Дезарг пополнил вещественную аффинную плоскость бесконечными точками, а именно добавил по одной бесконечной точке к каждому пучку параллельных прямых, и множество всех таких точек объявил бесконечной прямой. В полученном таким образом проективном пространстве все прямые параллельного пучка прямых проходят через такую точку (рис. 14.19). В геометрии представляет интерес множество прямых, проходящих через заданную точку, или (что то же самое) множество одномерных векторных подпространств заданного векторного пространства (рис. 14.20). Такое множество встречается, например, при изучении касательной к кривой в некоторой точке. Одна из моделей проективного пространства основана именно на $G_{E,1}$. Заметим, что наше зрительное восприятие окружающего мира не плоское, а конечностное: вершина конуса совпадает с центром нашего глаза. Кроме того, необходимо связывать два изображения, к примеру, одной и той же плоскости, полученной из двух различных центров. Это имеет отношение к понятию перспективы в аэрофотосъемке. В проективной геометрии широко используются методы линейной алгебры, связывая, например, квадрику (т. е. кривую или поверхность второго порядка) с квадратичной формой. Проективные пространства над полем вещественных или комплексных чисел играют основную роль в дифференциальной геометрии, алгебраической топологии (к примеру, понятие кобордизма) и, конечно, в алгебраической геометрии. Одна из причин интереса к проективным пространствам заключается в том, что они представляют собой самые простые после сферы компактные многообразия. Проективная геометрия, так же как и геометрия Лобачевского, широко используется при геометризации физики. Пусть E — векторное пространство. Назовем проективным пространством, порожденным E или образованным из E, и обозначим через $P(E)$ фактор-множество $(E/R\backslash O/R)$, где отношение эквивалентности R определяется следующим образом: xRy в том и только в том случае, если $y=\lambda x$, $\lambda\in K$. Если кратко, то проективное пространство — это некоторое $P(E)$. Размерность $P(E)$ равна по определению dim $(E-1)$. Каноническая проекция — это отображение p; $E\backslash O\to P(E)$ (рис. 14.21). Далее будет дано более строгое определение размерности, но в данном случае это определение интуитивно принято, так как $P(E)$ как множество прямых имеет размерность на 1 меньшую, чем размерность E.

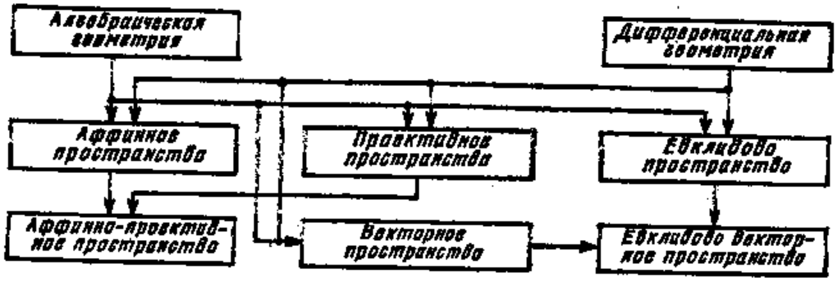


Рис. 14.17. Пояснение к понятиям геометрий и пространств

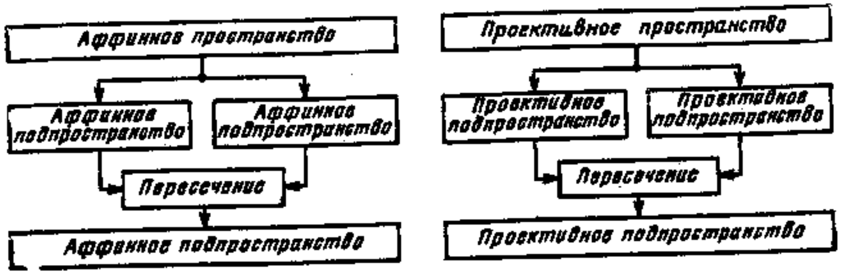


Рис. 14.18. Сравнение аффинного и проективного пространств

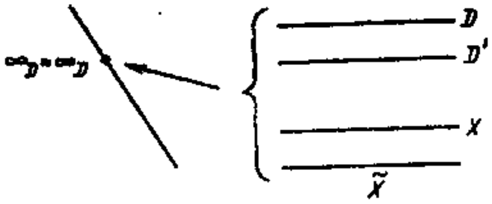


Рис. 14.19. Пояснение к понятию проективного пространства

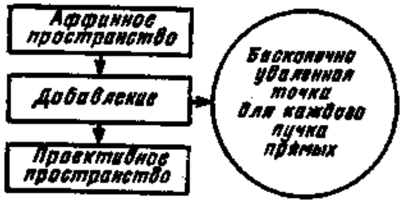


Рис. 14.20. Связь аффинного и проективного пространств

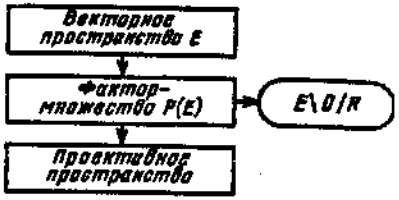


Рис. 14.21. К определению проективного пространства с помощью векторного пространства

Приведенное выше определение проективного пространства тесно привязано к понятию векторного пространства. Далее будут даны другие определения, которые освобождают от этой зависимости. Кроме того, напомним, что, как рассмотрено ранее, проективное пространство получается из аффинного пространства путем процедуры пополнения.

Пример 14.25. Для любого целого $n\geqslant 1$ положим $p^n(k)=p(k^n)$ и будем говорить о стандартном проективном пространстве размерности над полем K.

Пример 14.26. Проективное пространство называется вещественным, если $K=R$, и комплексным, если $K=C$.

Пример 14.27. Если размерность проективного пространства равна 0, то говорят о точке; если его размерность равна 1 (соответственно 2), то говорят, что это проективная прямая (соответственно проективная плоскость).

Пример 14.28. Существует естественная биекция между проективными пространствами $P(E)$ и $G_{E,1}$. При необходимости будем их отождествлять.

Пример 14.29. Обозначив множество гиперплоскостей из E через $H(E)$, получим биекцию (а следовательно, при необходимости и отождествим)

$$H(E)\rightarrow P(E^*),$$

где E^* — пространство, действенное к E.

Пример 14.30. Пусть $\mathscr{P}_k(x)$ — векторное пространство однородных полиномов степени K на аффинном пространстве X и N — ядерное отображение $N:\mathscr{P}_k(X)\rightarrow$подмножества в X, при котором полином $f\in\mathscr{P}_k(X)$ сопоставляется с его ядром $f^{-1}(0)$ с X, т. е. множеством нулей полинома f. Так как $(\lambda f)^{-1}(0)=f^{-1}(0)$ для всех $\lambda\in k^*$, то получаем факторизацию (рис. 14.22). Можно показать, что в некоторых случаях N инъективно. Если V — векторное пространство и если заметить, что $()\check{f}(\lambda_x)=\lambda^k f(x)$ для $f\in\mathscr{P}_k(X)$, то становится ясно, что естественно рассматривать диаграмму на рис. 14.23 для ядерного отображения $\mathscr{P}_k{}^*(v)$ подмножества в V, так как каждое $f^{-1}(0)$ — это на самом деле некоторый конус в V. Пусть K — конечное поле из $K=\#K$ элементов и $P(E)$ есть n-мерное проективное пространство над K. Тогда $\#\,P(E)=(K^{n+1}-1)/(K-1)$, так как $\#\,K=K^{n+1}$ и каждая прямая содержит, не считая начала координат, $K-1$ элементов. Например, каждая проективная прямая на z_2 содержит три точки, каждая плоскость — сеть точек.

$$\begin{tikzcd}
\mathscr{P}_k(X)\setminus 0 \arrow[rd, "N"] \arrow[dd, "p"'] & \\
& \text{Подмножества в } X \\
P(\mathscr{P}_k(X)) \arrow[ru, "N"'] &
\end{tikzcd}$$

Рис. 14.22. Пояснение к факторизации (пример 14.30)

Описание проективных пространств: карты. Пусть $P(E)$ — проективное пространство. Если $H\subset E$ — гиперплоскость, то в соответствии с $E_{\text{н}}G_{E,1}\setminus G_{H,1}=P(E)\setminus P(H)$ — подмножество в $P(E)$ или почти все $P(E)$. Но $E_{\text{н}}\subset P(E)$ является аффинным пространством, т. е. пожертвовав лишь точками из $P(E)$, мы представили точки из $P(E)$ с помощью точек некоторого аффинного пространства, а это оказывается очень полезным при проведении расчетов, выборе реперов и т. д. Биекцию $P(E)\setminus P(H)\rightarrow E_{\text{н}}$ называют картой пространства $P(E)$. Если хотим получить все $P(E)$, то нужно покрыть его областями определения карт. Предположим, что размерность $P(E)$ равна n и $H_i(i=0,\ldots,n)$ — семейство $n+1$ гиперплоскостей из E, такое, что $\bigcap_i H_i=0$. Тогда $\bigcap_i H_i=0$, и, следовательно, $P(E)=\bigcup_i(P(E)\setminus P(H_i))$ действительно является объединением областей определения карт. Таким образом, нам удалось покрыть $P(E)$ аффинными пространствами. Можно сказать, что у нас имеется атлас пространства $P(E)$. Но для того чтобы атлас оказался пригодным, необходимо согласовать карты, т. е. найти отображение, определенное с помощью диаграммы на рис. 14.24, где $i\neq j$ и штриховая стрелка означает, что отображение определено лишь на части E_{H_i}. Другая, более простая идея относительно того, как следует проводить вычисления

$$\begin{tikzcd}
\mathscr{P}_k^*(V)\setminus 0 \arrow[r, "N"] & \text{Подмножества в } N \\
P(\mathscr{P}_k^*(V)) \arrow[ru, "N"] \arrow[r, "N"'] & \text{Подмножества в } P(V)
\end{tikzcd}$$

Рис. 14.23. Пояснение для ядерного отображения (пример 14.30)

$$\begin{tikzcd}
 & P(E)\setminus(P(H_i)\cup PH_j)) \arrow[dl] \arrow[dr] & \\
E_{H_i} \arrow[rr, dashed] & & E_{H_j}
\end{tikzcd}$$

Рис. 14.24. Пояснение к процедуре покрытия проективного пространства $P(E)$ аффинными пространствами

в $P(E)$, состоит в том, чтобы в случае $\dim E<\infty$ выбрать некоторый базис $\{l_i\}_{i=0,1,\dots,n}$ в E. Тогда любой элемент $m=P(E)$ примет вид $m=p(x)=p(x_0, x_1, \dots, x_n)$, если $x=(x_0, x_1, \dots, x_n)$ в рассматриваемом базисе. Говорят, что $(x_0, x_1, \dots, x_n)$ — система однородных координат точки m в рассматриваемом базисе. Слово «однородные» связано с тем, что вместе с этой системой координат имеется целое семейство таких систем, в которых та же самая точка выражается в виде $(\lambda x_0, \lambda x_1, \dots, \lambda x_n)$, где λ пробегает K. Объединим две предыдущие точки зрения, выбрав в E базис $\{l_i\}_{i=0,1,\dots,n}$ и задав гиперплоскости H_i соотношением $H_i\equiv x_i^{-1}(0)$. Точки на $P(E)\setminus P(H)$ — это те точки, для которых в системах однородных координат выполняется условие $x_i\neq0$. С другой стороны, E_{H_i} изоморфно аффинной гиперплоскости (параллельной H_i) $x_i^{-1}(1)=H_i+l_i$ в E. Наделим, наконец x_i^{-1} (1) аффинным репером $\{l_i\}\cup\{l_i+l_j\}_{j\neq i}$. В итоге получим биекцию π_i, $p(E)\setminus p(H_i)\to k^n$, задаваемую соотношением

$$\pi_i : p(E)\setminus p(H_i)\ni p(x_0, x_1, \dots, x_n)\mapsto\left(\frac{x_0}{x_i}, \dots, \frac{x_{i-1}}{x_i},\right.$$
$$\left.\frac{x_{i+1}}{x_i}, \dots, \frac{x_n}{x_i}\right)\in K^n, \tag{14.59}$$

которое представляет собой явное выражение некоторой карты пространства $p(E)$ со значениями на этот раз в стандартном векторном пространстве K^n.

Теперь можно непосредственно вычислить $\pi_i\bigcirc\pi_i^{-1}$ с помощью рис. 14.25. Сначала находим

$$\pi_i^{-1} : (v_1, \dots, v_n)\mapsto p(v_1, \dots, v_{i-1}, 1, v_{i+1}, \dots, v_n), \tag{14.60}$$

далее

$$p(v_1, \dots, v_{i-1}, 1, v_{i+1}, \dots, v_n)=p\left(\frac{v_1}{v_{j-1}}, \dots, \frac{v_{i-1}}{v_{j-1}},\right.$$
$$\left.\frac{1}{v_{j-1}}, \frac{v_{i+1}}{v_{j-1}}, \dots, \frac{v_{i-2}}{v_{j-1}}, 1, \frac{v_j}{v_{j-1}}, \dots, \frac{v_n}{v_{j-1}}\right),$$

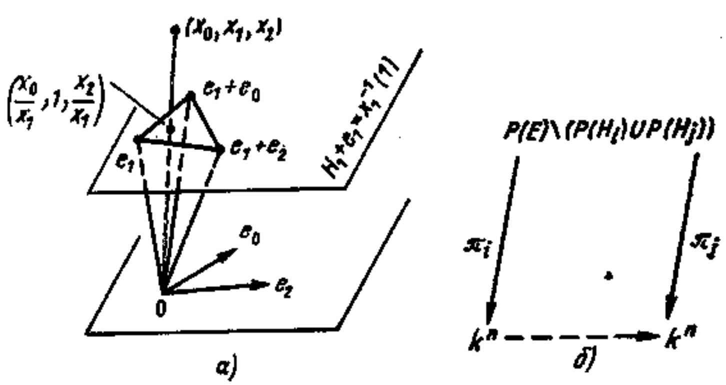


Рис. 14.25. Пояснение к согласованию карт атласа

следовательно,

$$\pi_j\bigcirc\pi_i^{-1};\quad (v_j, \dots, v_n)\mapsto\left(\frac{v_1}{v_{j-1}}, \dots\right.$$
$$\left.\dots, \frac{v_{i-1}}{v_{j-1}}, \frac{1}{v_{j-1}}, \frac{v_{i+1}}{v_{j-1}}, \dots, \frac{v_{j-2}}{v_{j-1}}, \frac{v_j}{v_{j-1}}, \dots, \frac{v_n}{v_{j-1}}\right), \tag{14.61}$$

где на самом деле $\pi_j\bigcirc\pi_i^{-1}$ определено лишь как $K^n\times\times(\pi_i(p(H_j)))=v_{j-1}^{-1}(0)$.

Пример 14.31. $E=k^2$. Это важный случай, так как $p(K^2)=p'(K)$ — самое простое нетривиальное проективное пространство, а именно стандартная проективная прямая над K. Разумеется, в качестве $\{l_0, l_1\}$ берется стандартный канонический базис в K^2. Здесь формула (14.60) сводится к отображению $K^*\ni v\mapsto1/v\in K^*$. Можно утверждать, что $p^1(K)$ получается склейкой двух экземпляров K с помощью отображения $v\mapsto1/v$ на K. В случае $K=C$ получаем одно из определений сферы Римана. Можно видеть, что $p^1(K)$ представляет собой объединение K, вложенное в $p^1(K)$ с помощью $\pi_0^{-1} : v\mapsto p(v, 1)$ и точки $p(1, 0)$. В этом смысле $p^1(K)$ можно рассматривать как пополнение K бесконечно удаленной точкой $p(1, 0)$. Отображения $\pi_j\bigcirc\pi_i^{-1} : v_{j-1}^{-1}(0)\mapsto K^n$, не будучи ни линейными, ни аффинными, все же настолько регулярны, насколько это возможно — они выражаются рациональными функциями. Если $K=R$ или C, то эти функции непрерывны и, более того, принадлежат классу C^∞ [это имеет смысл, поскольку $v_{j-1}^{-1}(0)$ — открытое множество в K^n]. Если $K=C$, то эти функции принадлежат классу C^ω (т. е. комплексно аналитичны).

Таким образом, на вещественных и комплексных конечномерных проективных пространствах можно было бы ввести понятия из топологии, дифференциальной геометрии и комплексного переменного. На языке теории многообразий (см. гл. 10) можно было бы сказать, что эти проективные пространства представляют собой топологические многообразия или гладкие многообразия класса C^∞ или C^ω. Если $K=C$, то комплексные проективные пространства конечной размерности служат естественными объектами изучения алгебраической геометрии.

Топология и алгебраическая топология

Заметим, что здесь рассмотрим только конечномерные проективные пространства, вещественные или комплексные. Если E — конечномерное векторное пространство над R или C, то у него есть каноническая топология.

Определение. Каноническая топология на проективном пространстве $p(E)$ — это фактор-топология, которая получается из топологии пространства $E\backslash 0$ с помощью отношения эквивалентности, определенного по правилу $xRy \Leftrightarrow (\exists\lambda(-K)(\lambda\neq 0$ и $y=\lambda x)$. Пространство $p(E)$ всегда будем наделять этой топологией. Если H — гиперплоскость в E, то биекция $E_H \to p(E)\backslash p(H)$ представляет собой гомоморфизм между $E(H)$ и $p(E)\backslash p(H)$, наделенным индуцированной топологией (рис. 14.26). Если выразить это отображение в координатах, то оно совпадет с отображение $\pi_n{}^{-1}$, заданным формулой (14.60), которое, очевидно, непрерывно, как и

$$\pi_n : p(v_1, \ldots, v_n) \mapsto (v_1, \ldots, v_n),$$

в силу свойств фактор-топологии. Отсюда следует, что топология пространства $p(E)$, определенная выше, совпадает с топологией многообразия. Пространство $p(E)$ отделимо, линейно связно и компактно.

Будем рассуждать геометрически: пусть m, n — две различные точки в $p(E)$; в E существует такая гиперплоскость H, что аффинное пространство отделимо. Алгебраическое доказательство этого факта выглядит следующим образом. Обозначим через $\lambda^2 E$ векторное пространство 2-векторов, принадлежащее внешней алгебре λE пространства E, и введем отображение

$$\alpha : (E\backslash 0)\times(E\backslash 0) \ni (x, y) \mapsto x\wedge y \in \lambda^2 E.$$

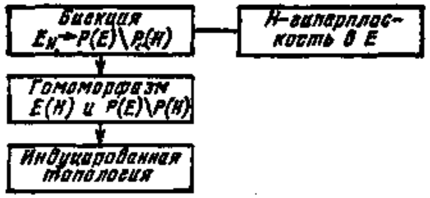


Рис. 14.26. Пояснение к индуцированной топологии

Это отображение непрерывно, так как оно билинейно, и его ядро $\alpha^{-1}(0)$ в точности совпадает с графиком отношения эквивалентности, определяющего $P(E)$. Являясь прообразом замкнутого множества при непрерывном отображении, $\alpha^{-1}(0)$ замкнуто; следовательно, фактор-пространство $p(E)=(E\backslash 0)/R$ отделимо. Можно не знать внешней алгебры и обойтись без нее, если выбрать базис в E и задать отображение α в координатной форме:

$$\alpha((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = ((x_1y_2 - x_2y_1), \ldots, x_{n-1}y_n - -x_ny_{n-1}) \in K^{n(n+1)/2}. \tag{14.62}$$

Для доказательства двух последних утверждений в (14.62) рассмотрим E как векторное пространство над R (пусть $K\mp R$ или C) и наделим E структурой евклидова векторного пространства; пусть тогда $S\{E\}=\{x\in E : \|x\|=1\}$ — единичная сфера в E. Поскольку $p(x)=p(x/\|x\|)\,\forall x\in E\backslash 0$, то получим $P(S(E))= =P(E)$.

Известно, что $S(E)$ компактна, следовательно, $P(E)$ компактно, поскольку оно отделимо. Сфера $S(E)$ линейно связна, если $\dim E\geqslant 2$, следовательно, $P(E)$ тоже линейно связно, правда, априори за исключением случая $\dim E=1$. Но в этом случае $P(E)$ состоит всего из одной точки. Это доказательство показывает, помимо всего прочего, что $P(E)$ гомеоморфно топологическому фактор-пространству сферы $S(E)$ по отношению эквивалентности $xRy : y=\pm x$.

Так как $p(\vec{x})$ отождествляется с прямыми, проходящими через заданную точку вещественного аффинного пространства X, убеждаемся в том, что если C — непрерывная кривая, проходящая через a, то множество прямых в X, соединяющих a с другими точками C, содержит, по крайней мере, один предельный элемент. Это хороший кандидат на роль касательной к C в точке a (рис. 14.27). Заметим, что можно определить топологию на проективном пространстве $P(E)$ и в более общих,

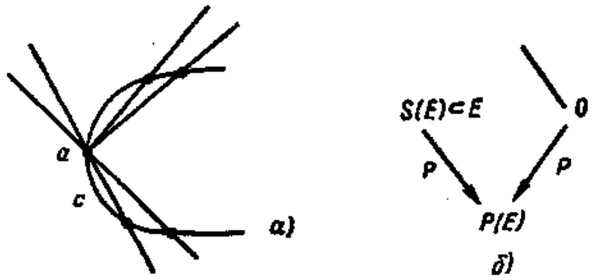


Рис. 14.27. Пояснение к свойствам $P(E)$

$$\begin{tikzcd}
S^1 \arrow[dr, "c"] \arrow[dd, "P"'] & \\
& S^1 \\
P^1(R) \arrow[ur, "\underline{c}"'] &
\end{tikzcd}$$

Рис. 14.28. Пояснение к гомеоморфизму

чем $K=R$ или C, например, если K — локально-компактное поле.

Пространство $p^1(R)$ гомеоморфно сфере S^1, а пространство $p^1(C)$ сфере S^2. Отождествим R^2 с C, и пусть $S^1=S(R^2)=S(c)$, тогда $P^1(R)=P(S^1)$. Построим диаграмму (рис. 14.28). Рассмотрим отображение $C:S^1\ni z\mapsto z^2\in S^1$, которое заключается в том, что комплексное число $z\in S^1$ возводится в квадрат (при этом, если $|z|=1$, данное значение модуля сохраняется и после отображения C).

14.6. ГЕОМЕТРИЯ ЕВКЛИДОВА ПРОСТРАНСТВА

Вначале будут рассмотрены евклидовы векторные пространства.

Определение. Евклидовым векторным пространством E называется конечномерное пространство над R, снабженное билинейной симметрической положительно определенной формой $\varphi : E\times E\to R$ [положительная определенность означает, что $\varphi(x, x)>0$ при всех $x\neq 0$]. Величина $\varphi(x, y)$ обозначается через (x/y) и называется скалярным произведением векторов x и y. Нормой $\|x\|$ вектора x называется число $\sqrt{\varphi(x, x)}=\sqrt{(x/x)}$. Если $(x/y)=0$, то говорят, что x и y ортогональны. Множество векторов $\{l_i\}$, $i=1, \ldots, k\subset E$ называется ортогональной системой, если $(l_i/l_j)=0$ для всех $i\neq j$, ортонормированной системой, если, кроме того, $\|l_i\|=1$ для всех i.

Заметим, что в любом векторном подпространстве F евклидова векторного пространства E естественным образом вводится структура евклидова векторного пространства посредством сужения формы φ на F (рис. 14.29).

Стандартный пример евклидова пространства:

$$E=R^n,\ \varphi((x_1, \ldots, x_n), (y_1, \ldots, y_n))=\sum_{i=1}^{n} x_i y_i.$$

Билинейная форма φ определяет квадратичную форму $\|x\|^2$ на E: $\|x\|^2=(x|x)$. Наоборот, квадратичная форма $(x|x)$ определяет билинейную форму на E: для всех $x, y\in E$ полагаем

$$(x|y)=\frac{1}{2}(\|x+y\|^2-\|x\|^2-\|y\|^2). \tag{14.63}$$

Ортогональная система, состоящая из ненулевых векторов, всегда линейно независима. Пусть $\{l_i\}$ — ортонормированный базис в E. Тогда разложение вектора x по базису $\{l_i\}$ дается формулой $x=\sum\limits_{i=1}^{n} x_i l_i=\Sigma(x|l_i)l_i$. Кроме того,

$$\left(\sum_{i=1}^{n} x_i l_i \middle| \sum_{j=1}^{n} y_j l_j\right)=\sum_{i=1}^{n} x_i y_i.$$

Предложение. Для всех $x, y\in E$ имеем $|(x/y)|\leqslant\|x\|\times\|y\|$. Равенство достигается тогда и только тогда, когда векторы x и y линейно зависимы. Точнее, равенство

$$(x/y)=\pm\|x\|\times\|y\|$$

эквивалентно существованию таких вещественных чисел λ и μ, не равных нулю одновременно, что $\lambda x+\mu y=0$. Далее, для всех $x, y\in E$ выполнено неравенство $\|x+y\|\leqslant\|x\|+\|y\|$; в частности, если положить $d(x, y)=\|x-y\|$, то E превращается в метрическое пространство с канонической топологией.

Предложение (ортонормализация по Шмидту). Пусть $\{b_i\}$, $i=1, \ldots, k$ — некоторый набор линейно независимых векторов в E. Тогда существует ортонормированная система, обладающая следующими свойствами: 1) набор $\{a_i\}$ гомотопен набору $\{b_i\}$ (...); 2) $\forall n=1, \ldots, k$ — векторное пространство, порожденное векторами $\{a_1, \ldots, a_n\}$, совпадает с векторным пространством, порожденным векторами $\{b_1, \ldots, b_n\}$. Любую ортонормированную систему (в том числе пустую) можно дополнить до ортонормированного базиса. Любой базис гомотопен некоторому ортонормированному базису.

Если x — ненулевой вектор, то обозначим через x^n нормированный вектор, т. е. $x^n=\|x\|^{-1}X$. Векторы x и x^n связаны очевидной гомотопией $x(t)=tx^n+(1-t)x$, $t\in[0, 1]$, описывающей отрезок с концами x и x^n.

В случае $k=1$ достаточно положить $a_1=b_1{}^n\ldots$ Рассуждая по индукции, предположим, что векторы a_i при $i=1, \ldots, n-1$ построены. Положим тогда

$$a_n=\sum_{i=1}^{n-1}(a_i/b_n)a_i;$$

$$c_n=b_n-d_n,\ a_n=c_n^n.$$

Прямая проверка показывает, что система $\{a_i\}_{i=1,\ldots,h}$ обладает нужными свойствами. Например, требуемую геометрию строим, дополняя (существенную по предположению индукции) гомотопию от $\{a_i\}_1\leqslant <i\leqslant h-1$ K b_i, $1\leqslant i\leqslant i-1$, гомотопией от bh к $ch: bh(t)=tc_h+(1+t)b_h$ и затем от c_h к $ah: c_h(t)=ta_h+(1-t)c_h$ (рис. 14.30).

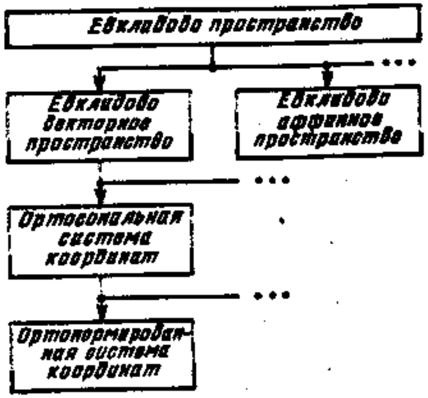


Рис. 14.29. Различные варианты евклидова пространства

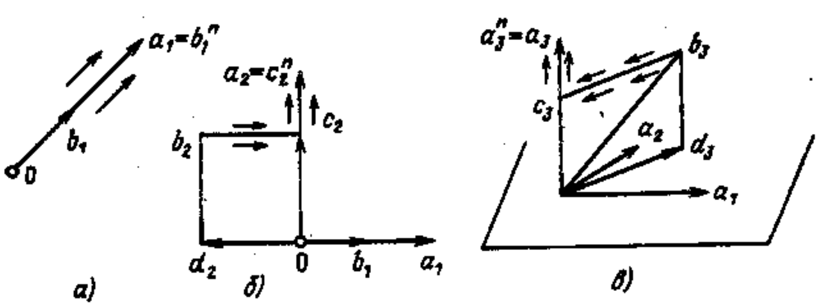


Рис. 14.30. Пояснение к ортонормализации по Шмидту

Предложение. Пусть E и E' — евклидовы пространства одинаковой размерности; $f: E \to E'$ — некоторое отображение. Тогда следующие условия равносильны:

$$(i)\, f \in L(E,\ E^1) \text{ и } \|f(x)\| = \|x\| \forall x \in E;$$

$$(ii) \forall x,\ y \in E: (f(x)/f(y)) = (x/y).$$

Если эти условия выполнены, то биективно.

Такие отображения называются изометриями; множество всех изометрий из E в E' обозначается $O(E;\ E')$.

Импликация $(i) \Rightarrow (ii)$, вытекает из линейности f и равенства (14.63). Для доказательства обратной импликации выберем в E (см. ортонормализацию по Шмидту) ортонормированный базис $\{l_i\}$. Из равенства размерностей и условия (ii) вытекает, что система $\{f(l_i)\}$ образует ортонормированный базис в E', и тогда в силу формулы $x = \Sigma x_i l_i = \Sigma (x/l_i) l_i$ $f(\Sigma x_i l_i) = = \Sigma x_i f(l_i)$ и доказательство линейности f закончено.

Следствие. Все евклидовы векторные пространства данной размерности n изометричны пространству R^n. Поэтому можно говорить просто об «евклидовом n-мерном пространстве».

Двойственность. Лемма. Пусть E — евклидово векторное пространство, E^* — двойственное к нему как векторному пространству. Отображение $\flat: E \ni x \mapsto x^0 = \{y \mapsto (x/y)\} \in E^*$ является изоморфизмом векторных пространств. Обратное отображение обозначается $\sharp = \flat^{-1}: E^* \to E$.

Можно показать, что $\dim E = \dim E^*$ (с помощью двойственного базиса в E^*). Так как $\flat$ линейно, достаточно убедиться в том, что оно имеет нулевое ядро. Это очевидно: если $(x/y) = 0$ для всех y, то $(x/x) = \|x\|^2 = 0$, откуда $x = 0$.

Лемма позволяет ввести в E^* естественную структуру евклидова пространства.

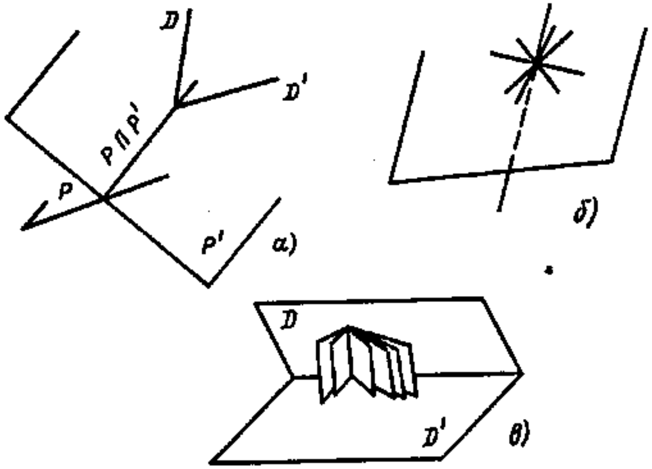


Рис. 14.31. Пояснение к случаю $\dim E = 3$ (классических результатов евклидова пространства)

Определение. Пусть A — подмножество в E. Через $A^\perp$ обозначим множество $\{x \in E: (x|y) = 0 \forall y \in A\}$. Множество $A^\perp$ называется ортогональным дополнением множества A.

Это множество является векторным подпространством в E, и если $<A>$ обозначает векторное подпространство в E, порожденное множеством A, то $<A>^\perp = A^\perp$. Для всякого подпространства A имеем $A \oplus A^\perp = E$, $\dim A + \dim A^\perp = \dim E$, $(A^\perp)^\perp = A$. Если A и B — два подпространства в E, то $(A+B)^\perp = A^\perp \cap B^\perp$, $(A \cap B)^\perp = A^\perp + B^\perp$. Два подмножества A и B называются ортогональными, если $A \subset B^\perp$ или, что равносильно, $B \subset A^\perp$. В этом случае обозначают $A \perp B$.

В случае $\dim E = 3$ на основе аксиоматического метода получаем классические результаты евклидовой геометрии. Например, множество прямых, проходящих через данную точку и перпендикулярных данной прямой, есть плоскость, называемая перпендикулярной плоскостью к данной прямой. Далее, рассмотрим две прямые D и D', проходящие через данную точку и перпендикулярные соответственно двум данным плоскостям P и P'. Тогда плоскость, проходящая через D и D', перпендикулярна прямой $P \cap P'$ (рис. 14.31).

Пусть f — эндоморфизм евклидова векторного пространства. Определим сопряженный эндоморфизм ${}^t f$ формулой ${}^t f = \sharp \circ f^* \circ \circ \flat: E \to E$, где f^* — дуальный к f эндоморфизм, $f^*: E^* \to E^*$. Эндоморфизм ${}^t f$ удовлетворяет соотношению $({}^t f(x)/y) = (x/f(y)) \forall x, y \in E$.

Ортогональная группа: элементарные свойства и план изучения

Предложение. Обозначим множество $O(E; E)$ через $O(E)$ и назовем его ортогональной группой пространства E. Группа $O(R^n)$, обозначается $O(n)$. Эндоморфизм f принадлежит группе $O(E)$ тогда и только тогда, когда в каком-нибудь ортонормированном базисе (а следовательно, и в любом ортонормированном базисе) матрица A этого эндоморфизма удовлетворяет условию ${}^tAA=1$ (где t обозначает транспонирование, 1 — единичную матрицу). В частности, для $f\in O(E)$ выполнено равенство $\det f=\pm 1$; положим $O^+(E)=\{f\in O(E):\det f=1\}$, $O^-(E)=\{f\in\in O(E):\det f=-1\}$. Элементы множества $O^+(E)$ называются вращениями.

Имеем $f\in O(E)\Leftrightarrow {}^tf\cdot f=I_{dE}$. Это предложение вытекает из следующей леммы.

Лемма. Пусть $\{l_i\}$, $i=1, \ldots, n$ — ортонормированный базис в E; A и B — две квадратные матрицы порядка $n(=\dim E)$. Пусть столбцы матриц A и B образованы из координат векторов $x_1, \ldots, x_n$ и соответственно $y_1, \ldots, y_n$. Тогда ${}^tAB==((x_i/y_i))$.

Если $\dim E=n$, тогда группа $O(E)$ изоморфна $O(n)$. Поэтому при изучении ортогональных групп будем ограничиваться группами $O(u)O^+(E)$. Кроме того, имеют место следующие правила умножения:

$$O^-(E)\cdot O^+(E)\subset O^-(E),\quad O^-(E)\cdot O^-(E)\subset O^+(E).$$

Группа $O(E)$ компактна в топологии, индуцированной из $GL(E)$. Действительно, если выбрать в конечномерном векторном пространстве E какой-нибудь базис и обозначить через $A(f)$ матрицу эндоморфизма f в этом базисе, то отображение $f\mapsto(T_r^tA(f))^{1/2}$ задает форму в $GL(E)$, порожденную соответствующей евклидовой структурой. Если E — евклидово пространство и мы выбрали в нем ортонормированный базис, то все элементы $O(E)$ будут иметь одну и ту же норму, равную $\dim E$; стало быть, $O(E)$ — ограниченное множество в $GL(E)$. Легко видеть, что оно замкнуто, так как ${}^tAA=1$. Таким образом, $O(E)$ компактно.

В ограниченности $O(E)$ можно убедиться и другим способом. Введем в пространстве $GL(E)$ (не евклидову) норму, положив по определению $\|f\|=\sup\{\|f(x)\|:x\in E$ и $\|x\|=1\}$. Тогда все элементы $O(E)$ имеют норму, равную 1, и мы получаем требуемое.

Утверждение о компактности группы $O(E)\subset GL(E)$ допускает следующее полезное обращение.

Теорема 14.2. Пусть G — компактная подгруппа в группе $GL(E)$, где E — некоторое конечномерное вещественное векторное пространство (априорно не являющееся евклидовым). Тогда в E существует, по крайней мере, одна евклидова структура φ, такая, что $G\in O(E)$, где $O(E)$ — ортогональная группа для этой евклидовой структуры.

Для доказательства требуется понятие меры Хаара на компактной группе. Пусть dg — мера Хаара на G и φ — какая-нибудь евклидова структура на E. Для $g\in G$ определим евклидову структуру $g^*\varphi$ на E, полагая $(g^*\varphi(x, y)=\varphi(g(x); g(y))$ для всех $x, y\in E$. Искомая евклидова структура на E определяется теперь как усреднение по группе G структур $g^*\varphi$:

$$\overline{\varphi}=\int_{g\in G}g^*\varphi dg,$$

т. е. для всех

$$\overline{\varphi}(x, y)=\int_{g\in G}\varphi(g(x), g(y))\,dg.$$

Очевидно, что $\overline{\varphi}$ — еще одна евклидова структура на E, причем инвариантная относительно действия G (т. е. $g^*\overline{\varphi}=\overline{\varphi}$ для всех $g\in G$). Это вытекает из определения $\overline{\varphi}$ и инвариантности меры dg относительно сдвигов из G. Но инвариантность $\overline{\varphi}$ относительно действия G эквивалентна включению [где $O(E)$ — группа изометрий для евклидовой структуры $\overline{\varphi}$].

В общем случае неверно, что инвариантная относительно G евклидова структура единственна: контрпример доставляет подгруппы на $G=\{I_{dE}\}$. Существует, однако, полезный критерий единственности: инвариантная относительно G евклидова структура единственна (с точностью до скалярного множителя) тогда и только тогда, когда группа G неприводима в $GL(E)$, т. е. когда в E нет других инвариантных относительно действий G пространств, кроме $\{O\}$ и E. Если G конечна, $G=\{g_i:i\in I\}$, то можно определить инвариантное скалярное произведение, положив просто $\overline{\varphi}=\sum_i g_i{}^*\varphi$.

Предложение. Предполагается, что $\dim E\geqslant 2$. Группа $O^+(E)$ транзитивно действует на единичной сфере $S(E)=\{x\in E:\|x\|==1\}$ пространства E, а также на любом грассманиане $G_{E,p}(0\leqslant p\leqslant\dim E)$. Группа $O(E)$ действует просто транзитивно на множестве ортонормированных базисов пространства E, а группа $O(E)$ действует просто транзитивно на множестве гомотопных ортонормированных базисов.

Это предложение позволяет представить стандартную сферу $S^n=S(R^{n+1})$ размерности n как однородное пространство $S^n==O(n+1)/O(n)$, где группа $O(n)$ вложена естественным образом в $O(n+1)$ как подгруппа всех изометрий, оставляющих на месте вектор $(O, O, \ldots, 1)\in R^{n+1}$. Такое же представление имеет место для грассманианов:

$$G_{n,p}=O(n)/(O(p)\times O(n-p)),$$

где $O(p)\times O(n-p)$ есть прямое произведение подгруппы $O(p)\subset O(n)$, которая состоит из всех элементов $f\in O(n)$, оставляющих неподвижными $n-p$ векторов $(l_{p+1}, \ldots, l_n)$ канонического базиса R^n, и аналитической подгруппы $O(n-p)$ для векторов $l_1, \ldots, l_p$.

Предложение. Пусть $E=S\oplus T$ — разложение E в прямую сумму и пусть σ — симметрия относительно подпространства S, параллельного подпространству T. Тогда $\sigma\in O(E)$ в том и только в том случае, если $E=S^\perp\oplus T$ (т. е. $T\in S^\perp$), при этом обозначаем отображение через σ и называем его ортогональной симметрией евклидова пространства E. Если $\dim^\perp=1$, то σ_S называется симметрией относительно гиперплоскости S, если $\dim S^\perp=2$, то σ_S называется переворачиванием. Если $\dim S^\perp$ четна, то σ_S лежит в $O^+(E)$, если $\dim S^\perp$ нечетна, то σ_S лежит в $O^-(E)$. Все симметрии инволютивны, и, наоборот, если $f\in O(E)$ и $f^2=I_{dE}$, то f — симметрия (рис. 14.32).

Можно получить явную формулу для симметрии σ_H относительно гиперплоскости H. Пусть X — ненулевой вектор в $H^\perp$ $(\dim H^\perp=1)$, тогда

$$\sigma_H(y)=y-2\frac{(yx)}{\|x\|^2}x.$$

Если $x, y\in E$ и $\|x\|=\|y\|$, то существует гиперплоскость H, для которой $\sigma_H(x)=y$; кроме того, если $x\neq y$, $x\neq 0$, то H единственна. В самом деле, при $x=y$ в качестве H годится любая гиперплоскость, содержащая X, а при $x\neq y$ искомая гиперплоскость H должна быть ортогональна $x-y$, и гиперплоскость $(x-y)^\perp$ подходит. Заметим, что H есть не что иное, как гиперплоскость, равноудаленная от x и y, т. е. медиатор точек

$$x,\ y: H=\{z\in E: d(x, z)=d(y, z)\}.$$

Теорема 14.3. Любой элемент f группы $O(E)$ представляется в виде композиции не более чем $n=\dim E$ симметрий относительно гиперплоскости (рис. 14.33).

Доказательство дается по индукции. С помощью предыдущих рассуждений можно подобрать симметрию σ относительно гиперплоскости так, чтобы $\sigma f(x)=x$ для некоторого $x\neq 0$. Тогда гиперплоскость $\{x\}$ инвариантна относительно отображения σf и имеет меньшую размерность.

Следствие. Если $\dim E=1$, то $O(E)=1\ dE\cup(-I_{dE})$.

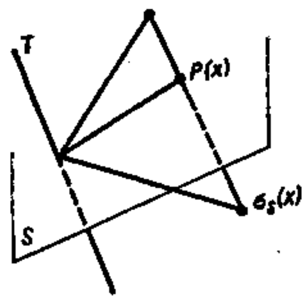


Рис. 14.32. Пояснение к симметрии евклидова пространства

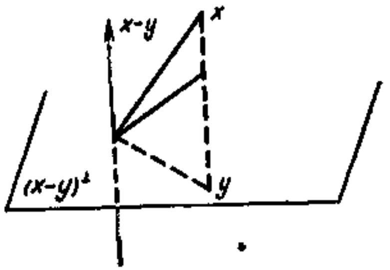


Рис. 14.33. Пояснение к симметрии относительно гиперплоскости

Если $\dim E=2$, то любой элемент из $O(E)$ является симметрией относительно некоторой прямой, а любой элемент $O^+(E)$ является произведением двух симметрий относительно прямых.

План изучения группы $O(E)$. Строение данного элемента группы $O(E)$. Вначале необходимо найти минимальное число симметрий относительно гиперплоскости (соответственно переворачиваний), требуемых для разложения $f\in O(E)$ [соответственно $f\in O^+(E)$]. Далее следует доказать, что любое $f\in O(E)$ разлагается в прямое произведение изометрий одно- и двухмерных подпространств, что мотивирует подробное изучение группы $O(2)$ (так как одномерный случай тривиален), при этом существенную роль играет коммутативность группы $O^+(2)$.

Структура группы $O(E)$ в целом. Необходимо показать, что группа $O^+(2)$ абелева, а также вычислить центры групп $O(E)$ и $O^+(E)$, а затем доказать, что группа $O^+(n)$ проста для $n=3$ и $n\geqslant 5$. Исключительный случай $n=4$ труден и требует привлечения кватернионов.

Топология группы $O(E)$. Необходимо показать, что $O^+(E)$ линейно связна, и, таким образом, $O^+(E)$ и $O^-(E)$ являются связными компонентами $O(E)$. Далее необходимо показать, что фундаментальная группа $\pi_1(O(n))$ совпадает с z при $n=2$ и с z_2 при $n\geqslant 3$. Эти факты имеют основополагающее значение для математики. Равенство $\pi_1(O(2))=z$ является краеугольным камнем теории комплексного переменного и теории кривых на плоскости, а равенство $\pi_1(O(n))=z_2$ (при $n\geqslant 3$) оказывается источником многих неожиданных математических построений.

Определение. Евклидовым аффинным пространством называется аффинное пространство $(x, \vec{x})$, в котором $\vec{x}$ снабжено евклидовой структурой.

СПИСОК ЛИТЕРАТУРЫ

1. **Винер Н.** Кибернетика или управление и связь в животных и машине: Пер. с англ. М.: Советское радио, 1967.

2. **Винер Н.** Кибернетика и общество: Пер. с англ. М.: Изд-во иностр. лит., 1959.

3. **Эшби Росс У.** Введение в кибернетику: Пер. с англ. М.: Изд-во иностр. лит., 1959.

4. **Бир Ст.** Кибернетика и управление производством: Пер. с англ. М.: Физматгиз, 1963.

5. Клаус Г. Кибернетика и общество: Пер. с англ. М.: Прогресс, 1967.

6. **Акофф Р., Сасиени М.** Основы исследований операций: Пер. с англ. М.: Мир, 1971.

7. **Глушков В. М.** Введение в АСУ. Киев: Техника, 1972.

8. **Колмогоров А. Н.** Три подхода к определению понятия информации. Проблемы передачи информации. М.: Наука. 1965. Вып. 1.

9. **Андронов А. А., Витт А. А., Хайкин С. Э.** Теория колебаний.— 2-е изд. М.: Физматгиз, 1958.

10. **Кузин Л. Т.** О системах искусственного интеллекта//Материалы семинара «Искусственный интеллект. Итоги и перспективы». М.: МДНТП им. Ф. Э. Дзержинского, 1974. С. 3—15.

11. **Кузин Л. Т.** Основные проблемы проектирования больших систем. Некоторые вопросы кибернетики. М.: МИФИ, 1970. Вып. 1.

12. **Экономическая** семиотика: Сб. научных трудов. М.: Наука, 1970.

13. **Системные** исследования: Ежегодник 1971. М.: Наука, 1971.

14. **Кузин Л. Т.** Расчет и проектирование дискретных систем управления. М.: Машгиз, 1961.

15. **Семиотические** методы управления в больших системах: Материалы семинара «Семиотические методы управления в больших системах». М.: МДНТП им. Ф. Э. Дзержинского, 1971.

16. **Шалютин С. М.** Искусственный интеллект. М.: Мысль, 1985.

17. **Ляпунов А. А., Яблонский С. В.** Теоретические проблемы кибернетики. М.: Физматгиз, 1963. Вып. 9.

18. **Кузин Л. Т.** Основы кибернетики. М.: МИФИ, 1970.

19. **Техническая** кибернетика. Теория автоматического регулирования/Под ред. В. В. Солодовникова. М.: Машиностроение, 1967.

20. **Глушков В. М.** Введение в кибернетику. Киев: Изд-во АН СССР, 1964.

21. **Шеннон К.** Работы по теории информации и кибернетике: Пер. с англ. М.: Изд-во иностр. лит., 1963.

22. **Кузин Л. Т.** Состояние и перспективы развития научно-технического направления «Искусственный интеллект»//Материалы семинара «Искусственный интеллект. Итоги и перспективы». М.: МДНТП им. Ф. Э. Дзержинского, 1985.

23. **Ершов А. П.** Автоматизация работы служащих//Микропроцессорные средства и системы. М.: 1984. № 2. С. 6—15.

24. **Громов Г. Р.** Персональные вычисления — новый этап информационной технологии//Микропроцессорные средства и системы. М.: 1984. № 1. С. 30—37.

25. **Основы информатики**/А. И. Михайлов и др. М.: Наука, 1968.

26. **Брауэр Ф., Гроэ Г.** Информатика: Пер. с англ. М.: Мир, 1976.

27. **Martin J.** Application Development Without Programmers. 1982. P. 350.

28. **Поспелов Г. С.** Искусственный интеллект — основа новой информационной технологии. М.: Наука, 1988.

29. **Кузин Л. Т.** Основы кибернетики. Т. 1. Математические основы кибернетики. М.: Энергия, 1973.

30. **Кузин Л. Т.** Основы кибернетики. Т. 2: Основы кибернетических моделей. М.: Энергия, 1979.

31. **Джордж Ф.** Основы кибернетики. М.: Радио и связь, 1984.

32. **Элти Дж., Кумбс М.** Экспертные системы. Концепции и примеры: Пер. с англ. М.: Финансы и статистика, 1987.

33. **Поспелов Д. А.** Ситуационное управление. Теория и практика. М.: Наука, 1986.

34. **Логика** рассуждений и ее моделирование: Сб. трудов/Под ред. Д. А. Поспелова//Вопросы кибернетики. М.: АН СССР, 1983.

35. **Поспелов Г. С., Поспелов Д. А.** Искусственный интеллект: состояние и перспективы//Материалы семинара «Искусственный интеллект. Итоги и перспективы». М.: МДНТП им. Ф. Э. Дзержинского, 1974. С. 16—29.

36. **Вольфенгаген В. Э., Кузин Л. Т., Саркисян В. И.** Реляционные методы проектирования банков данных. Киев: Выща школа, 1979.

37. **Roussopoulos N. D.** CSDL: A Concentual Sheme Definition Language for the Design of Data Base Applications//IEEE Transactions of Software Engineering. Vol. 5, № 5. September 1979.

38. **Вольфенгаген В. Э., Яцук В. Я.** Аппликативные вычислительные системы и концептуальный метод проектирования систем знаний. М.: Изд-во М-ва обороны СССР, 1987.

39. **Энгелер Э.** Метаматематика элементарной математики: Пер. с англ. М.: Мир, 1987.

40. **Consinean G.** The Cathegoricale Abstract Machine//Functional Programminn Languages and Computers Architecture. 1985, LINOS, № 201.

41. **Машинная** реализация систем искусственного интеллекта. Сб. статей/Под ред. Л. Т. Кузина. М.: Энергоатомиздат, 1987.

42. **Интеллектуальные** банки данных: Сб. трудов/Под ред. Л. Т. Кузина. М.: АН СССР, 1979. Вып. 55. Вопросы кибернетики.

43. **Пензов Ю. Э.** Элементы математической логики и теория множеств. Саратов: Изд-во Саратовского ун-та, 1968.

44. **Куратовский К., Мостовой А.** Теория множеств. М.: Мир, 1970.

45. **Курош А. Г.** Лекции по общей алгебре. М.: Наука, 1973.

46. **Гольдблатт Р.** Топосы: категорный анализ логики: Пер. с англ. М.: Мир, 1983.

47. **Цаленко М. Ш., Шульгейфер Е. Г.** Основы теории категорий. М.: Наука, 1972.

48. **Цаленко М. Ш.** Моделирование семантики в базах данных. М.: Наука, 1989.

49. **Вольдман Г. Ш.** Объяснение непроцедуральности средствами теории категорий//Программирование. 1980. № 4.

50. **Бенсон Д.** Синтаксис и семантика с точки зрения теории категорий//Кибернетический сборник. М.: Наука, 1974. Вып. 11. С. 101.

51. **Джонсон П. Г.** Теория топосов. М.: Наука, 1986.

52. **Жожикашвили А. В., Стефанюк В. Л.** Теория категорий в задачах представления знаний и обучении//Техническая кибернетика. 1986. № 2.

53. **Толстов В. Г.** Об основных концепциях и проблемах квантовой информатики//Программная инженерия. М.: Знание, 1988, с. 130—132.

54. **Ершов Ю. Л., Палютин Е. А.** Математическая логика. М.: Наука, 1979.

55. **Мальцев А. И.** Алгебраические системы. М.: Наука, 1970.

56. **Калужнин Л. А.** Введение в общую алгебру. М.: Наука, 1973.

57. **Андрунакиевич В. А., Рябухин Ю. М.** Радикалы алгебр и структурная теория. М.: Наука, 1979.

58. **Мазур М.** Качественная теория информации: Пер. с англ. М.: Мир, 1974.

59. **Пфанцагль И.** Теория измерений. М.: Мир, 1976.

60. **Хинтикка Я.** Логико-эпистемологические исследования. М.: Прогресс, 1980.

61. **Чень Ч., Ли Р.** Математическая логика и автоматическое доказательство теорем. М.: Наука, 1983.

62. **Такеути Г.** Теория доказательств. М.: Мир, 1978.

63. **Справочная** книга по математической логике. М.: Наука, 1983. Ч. 1—4.

64. **Проектирование** интегрированных баз данных/А. А. Стогний, В. Э. Вольфенгаген, В. А. Кушеверов и др. Киев: Техника, 1987.

65. **Смальян Р.** Теория формальных систем. М.: Наука, 1981.

66. **Колмогоров А. Н., Драгалин А. Г.** Математическая логика. Дополнительные главы. М.: МГУ, 1984.

67. **Колмогоров А. И., Драгалин А. Г.** Введение в математическую логику. М.: МГУ, 1982.

68. **Драгалин А. Г.** Математический интуиционизм. Введение в теорию доказательств. М.: Наука, 1979.

69. **Линдон Р.** Заметки по логике. М.: Мир, 1968.

70. **Клини С.** Математическая логика. М.: Мир, 1973.

71. **Черч А.** Введение в математическую логику. М.: Изд-во иностр. лит., 1960.

72. **Фейс Р.** Модальная логика. М.: Наука, 1974.

73. **Карри Х.** Основания математической логики. М.: Мир, 1969.

74. **Семантика** модальных и интенсиональных логик. М.: Прогресс, 1981.

75. **Модальные** и интенсиональные логики и их применение к проблемам методологии науки. М.: Наука, 1984.

76. **Вольфенгаген В. Э.** Вычислительная модель реляционного исчисления, ориентирования на представление знаний. М.: МИФИ, 1984.

77. **Крипке С.** Семантический анализ модальной логики.1. Нормальные модальные исчисления высказываний/Фейс Р. Модальная логика. М.: Наука, 1971.

78. **Абрамов Д. Ю.** Реляционная и тензорная модели данных. Методы и системы диагностики//Сб. трудов. Саратов: Изд-во Саратовского ун-та, 1986. С. 20—27.

79. **Цаленко М. Ш.** Реляционные базы данных//Алгоритмы и организация решений экономических задач. М.: Статистика. 1977. Вып. 9. С. 18—35. Вып. 10. С. 16—28.

80. **Нечеткие** множества в моделях управления и искусственного интеллекта/Под ред. Д. А. Поспелова, 1986.

81. **Заде Л.** Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976.

82. **Барендрегт Х.** Лямбда-исчисление. Его синтаксис и семантика. М.: Мир, 1985.

83. **Крон Г.** Тензорный анализ сетей: Пер. с англ. М.: Советское радио, 1978.

84. **Крон Г.** Исследование сложных систем по частям — диаоптика: Пер. с англ. М.: Наука, 1972.

85. **Веблен О., Уайтхед Д.** Основы дифференциальной геометрии. М.: Изд-во иностр. лит., 1949.

86. **Веблен О.** Инварианты дифференциальных квадратичных форм. М.: ГИИЛ, 1948.

87. **Сухотин Б. В.** Основы проблемы грамматики и семантики в тензорном исчислении//Проблемы структурной лингвистики. М.: Наука, 1978. С. 234—283.

88. **Петров А. Е.** Тензорная методология в теории систем. М.: Радио и связь, 1985.

89. **Рашевский И. К.** Риманова геометрия и тензорный анализ. М.: Наука, 1967.

90. **Кузин Л. Т., Арменский А. Е., Абрамов Д. Ю.** Тензорные банки данных//Изв. вузов. Сер. Приборостроение. 1986. Т. XXIX, № 1.

91. **Кузина И. В., Тимофеева В. В.** Введение уравнений обобщения для построения математической модели объектов электромеханики//Изв. вузов. Сер. Электромеханика. 1986. № 10.

92. **Понтрягин Л. С.** Основы комбинаторной логики. М.: Наука, 1976.

93. **Телеман К.** Элементы топологии и дифференцируемые многообразия: Пер. с англ. М.: Мир, 1967.

94. **Стернберг С.** Лекции по дифференциальной геометрии: Пер. с англ. М.: Мир, 1970.

95. **Грауэрт Г., Либ И., Фишер В.** Дифференциальное и интегральное исчисления: Пер. с англ. М.: Мир, 1971.

96. **Артин Э.** Геометрическая алгебра. М.: Наука, 1969.

97. **Лобановский М. Г.** Начало геометрической физики. М.: Высшая школа, 1974.

98. **Грюнбаум Б.** Этюды по комбинаторной геометрии и теории выпуклых тел. М.: Наука, 1971.

99. **Берже М.** Геометрия: Пер. с франц. М.: Мир, 1984. Т. 1, 2.

ОГЛАВЛЕНИЕ