РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ
УНИВЕРСИТЕТ

На правах рукописи

Бениаминов Евгений Михайлович

АЛГЕБРАИЧЕСКИЕ МЕТОДЫ В ТЕОРИИ БАЗ ДАННЫХ И ПРЕДСТАВЛЕНИИ ЗНАНИЙ

Специальность — 05.13.17 —
Теоретические основы информатики

Диссертация на соискание ученой степени
доктора физико-математических наук

Москва – 1996

Содержание:

Введение

Актуальность темы Одним из последствий стремительного развития средств вычислительной техники явилось изменение технологии обработки информации во всех сферах человеческой деятельности. Современные ЭВМ позволяют накапливать огромные объемы информации и производить ее эффективную обработку. Одни и те же данные могут многократно использоваться в различных прикладных задачах. Централизованное хранение, накопление и ведение данных при одновременном доступе к данным большого числа пользователей позволило во много раз повысить эффективность функционирования информационных систем. Развитие новой информационной технологии, основанной на широком использовании компьютеров, современных средств передачи данных, баз данных и средств представления знаний потребовало решения ряда задач моделирования процессов организации, хранения, обработки данных и привело к постановкам новых математических задач.

Представление данных в виде алгебраических структур является естественным для многих специалистов в области программирования. Начиная с известной работы Е.Кодда [43] применение алгебраических методов распространилось на моделирование баз данных. Объектом изучения стали алгебраические системы с операциями над отношениями — реляционные алгебры и цилиндрические алгебры. Среди авторов, исследовавших эти объекты можно назвать H. Andreka, I. Nemeti, F. Banchilhon, Б. И. Плоткина и др. (см.[33, 34, 36, 28]). В настоящее время реляционные алгебры хорошо изучены в работах зарубежных и отечественных специалистов, включая работы автора [3, 4, 6].

Следующим шагом явилось изучение алгебраическими методами схем баз данных, расширение систем операций операциями над типами данных и над схемами. Это направление, с одной стороны, восходит к алгебраическим моделям абстрактных типов данных (R. M. Burstall [41], H. Ehrig [46], J. A. Goguen [50], S. N. Zilles [72] и др.), которые легли в основу языков программирования с развитым механизмами определения типов данных, а с другой стороны, привело к построению алгебраических моделей систем представления знаний, в которых существенна неполнота знаний, модульность организации знаний. Системы последнего типа условно можно назвать базами понятий.

Базы понятий — это новая складывающаяся область в программировании, связанная с использованием и построением больших библиотек типов объектов в объектно-ориентированных языках программирования, а также с использованием CASE–технологий и построением специализированных программных систем типа TOOLKIT, элементы которых предназначены для многоразового многоцелевого использования в различных приложениях.

Практика и опыт использования растущих библиотек подобного рода требует некоторой стандартизации в их организации и формах обращения к ним, которая была бы основана на ясной модели отдельных понятий и базы понятий в целом.

Состояние работ и исследований по базам понятий очень напоминает ситуацию с базами данных в конце 60-х начале 70-х годов, когда была осознана необходимость отделения процессов ведения данных от их использования, были созданы файловые системы, но не было ясной концепции баз данных. Для формулировки одной из таких концепций баз данных был успешно использован алгебраический язык в работах Е. Кодда.

Алгебраические методы, с учетом опыта их применения в теории баз данных и абстрактных типов данных, оказываются полезными и в моделировании баз понятий. Среди алгебраических средств в таких моделях особую роль играют понятия и методы теории категорий.

Один из важных вопросов приложения алгебры к информатике — это какого сорта проблемы в области баз данных и понятий могут решаться алгебраическими средствами, какие достижения в алгебре могут быть использованы для приложений в этой области, какие имеются ограничения по применению алгебраических методов в чистом виде.

Основное, что используется из алгебры в приложениях, — это алгебраический язык. Следующее — это алгебраический стиль мышления: использование понятий изоморфизма и гомоморфизма алгебраических систем, задание алгебраических систем с помощью образующих и соотношений. Дальнейшее использование алгебры приводит к использованию специфических понятий и результатов, введенных и исследованных в алгебре математиками. К ним относятся специфические операции и алгебраические системы, теоремы о разрешимости в алгебрах, специфические алгоритмы.

Результатами алгебраического исследования, помимо построения языка, на котором могут ставиться вопросы в области теории баз данных и понятий, являются теоремы классификации построенных алгебр и, в идеале, алгоритмы для распознавания тождеств, изоморфизмов, гомоморфизмов систем и решений уравнений, которые могут быть использованы в системах проектирования, алгоритмах оптимизации запросов, получения ответов на запросы.

Кроме того, как и в базах данных, так и в базах понятий существенный вопрос — это полнота использованной системы операций, который в некоторой постановке тоже может быть сведен к алгебраической задаче.

Цель работы. 1) Разработка алгебраических средств моделирования баз данных реляционного типа. 2) Полное описание и классификация реляционных алгебр. 3) Разработка алгебраических средств представления понятий и моделирования баз понятий. 4) Исследование математических объектов, используемых для представления понятий.

Методы работы. В исследовании использованы методы теории универсальных алгебр, математической логики, теории категорий, теории топосов.

Научная новизна. В диссертации предложены новые алгебраические объекты, служащие для моделирования баз данных и понятий.

Автором решены следующие задачи, определяющие научную новизну работы:

Дана формализация понятия реляционная алгебра.

Доказана теорема о представлении полных реляционных алгебр в виде прямой суммы неприводимых.

Дано полное описание неприводимых реляционных алгебр в духе теории Галуа.

Предложен подход к представлению понятий, основанный на методах теории категорий и средствах теории топосов.

Доказана теорема классификации булевых топосов конечного типа.

Введено понятие рефлексивного топоса, требуемого задачами представления понятий, и доказана теорема существования непротиворечивого рефлексивного топоса.

Рассмотрено приложение методов алгебраических систем с условными операциями и условными соотношениями к задачам представления понятий.

Теоретическая и практическая ценность. Полученные в диссертации результаты позволяют глубоко исследовать математические модели, возникающие при проектировании информационных систем и автоматизированных систем представления понятий, дают возможность анализировать структуру получающихся при этом алгебраических объектов. Алгебраические и, в частности, теоретико категорные методы находят все более широкое применение в математической логике и информатике. Теорема о классификации неприводимых алгебр отражает связь между законами предметной области и принципами инвариантности (симметриями). Эта проблема давно интересовала ученых в различных областях исследований. Предложенные в диссертации подходы могут быть использованы при проектирования баз данных и баз понятий. Алгебраические методы, предложенные здесь для представления понятий, легли в основу системы представления понятий, макет которой на языке TurboProlog 2.0. был разработан в Международном центре по информатике и электронике в 1990г.

Апробация работы Результаты работы докладывались:

- на Всесоюзном симпозиуме "Методы логики в проблемах искусственного интеллекта и информатики" (Телави, 1978);

- на Первой Всесоюзной конференции "Банки данных" (Тбилиси, 1980);

- на IV Всесоюзной конференции "Применение методов математической логики" (Таллин, 1986);

- на Всесоюзной конференции "ДИАЛОГ — 87" (Тбилиси, 1987);

- на II Всесоюзной конференции "Искусственный интеллект — 90" (Минск, 1990);

- на Российской конференции по логическому программированию (RCLP'91) (Ленинград, 1991);

- на расширенном семинаре Московской и Киевской секции ACM SIGMOD (Москва, 1993);

- на международной конференции (Япония и страны СНГ) "Проектирование программного обеспечения, основанного на знаниях (JCKBSE'94)" (Переславль - Залесский, 1994);

- на международном симпозиуме "Перспективы развития систем баз данных и информационных систем (ADBIS'94, ADBIS'95)" (Москва, 1994, 1995);

- на научных семинарах в МГУ по алгебре (руководитель проф. А. В. Михалев (1980)) и по математической логике (руководитель проф. А. Л. Семенов (1986)).

Структура диссертации Диссертация состоит из введения, трех глав и списка литературы. Нумерация определений, утверждений и примеров в каждой главе проводится отдельно. При ссылках внутри главы на определения, утверждения или примеры других глав к соответствующему номеру добавляется номер главы, выраженный римскими цифрами. Например, Определение II.1.3 — третье определение в разделе 1 второй главы.

СОДЕРЖАНИЕ РАБОТЫ

Во введении дан обзор литературы, обсуждена актуальность темы, сформулирована цель и определены задачи исследования, описаны основные результаты.

В главе 1 определяются и исследуются реляционные алгебры, служащие для моделирования баз данных реляционного типа.

Реляционный подход к моделированию баз данных на логическом уровне, введенный Е.Ф. Коддом, в настоящее время широко освещен в литературе.

Суть реляционного подхода к моделированию данных состоит в том, что состояние описываемого мира (или рассматриваемой его части) можно представить в виде набора отношений (таблиц), а обработка данных сводится к операциям над отношениями. Поэтому основными понятиями подхода являются понятия схемы отношений, состояние схемы отношений и операций над отношениями.

Обработка информации с точки зрения реляционной модели заключается в получении новых отношений из отношений, хранящихся в ЭВМ, с помощью точно описанных процедур, которые называются операциями над отношениями. В различных информационных системах может использоваться разный класс операций над отношениями. В традиционном подходе Е.Ф. Кодда этот класс порождается (путем композиции) операциями проекции, соединения отношений, переименования атрибутов отношений и теоретико–множественными операциями объединения, пересечения и разности над отношениями с одинаковыми наборами атрибутов.

Операции над отношениями понимаются в рассматриваемом здесь алгебраическом подходе к моделям баз данных как алгебраические операции. Далее, дается точное аксиоматическое определение реляционной алгебры базы данных как многоосновной алгебры .

Грубо говоря, под реляционной алгеброй понимается множество, замкнутое относительно объявленных операций над отношениями. Так как эти операции имеют разные области определения, то алгебраическая система получается многоосновной. Соотношения между операциями, которые выполняются на любых отношениях, называются аксиомами реляционной алгебры.

В разных информационных системах, вообще говоря, могут быть выбраны разные классы операций над отношениями, но, как правило, эти классы замкнуты относительно композиции операций и содержат ряд простых необходимых операций так, что к исследованию возникающих при моделировании таких информационных систем алгебр могут быть с успехом применены методы современной алгебры.

База данных в реляционном подходе задается схемой и состоянием. Схема базы данных — это набор схем отношений $R^1(a_1^1, ..., a_{n_1}^1), ...,$ $R^k(a_1^k, ..., a_{nk}^k)$ (они называются базисными), состояние которых описывает предметную область, и набор соотношений (зависимостей, ограничений целостности базы данных) между этими отношениями вида:

$$\begin{array}{lcr} f_1(R^1, \ldots, R^k) & = & g_1(R^1, \ldots, R^k) \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ f_l(R^1, \ldots, R^k) & = & g_l(R^1, \ldots, R^k), \end{array}$$

где f_i и g_i — операции над отношениями, выраженные в виде композиций основных операций.

В частности, в виде подобных соотношений можно записать условие функциональной зависимости какого–либо атрибута от набора других атрибутов, вхождение одного отношения в другое как подмножества, пустоту пересечения каких–либо отношений и т.д. Соотношения, зафиксированные в схеме базы данных, отражают условия целостности базы данных, которые обязательно должны выполняться. Они отражают смысл представленных данных и могут используются в приложениях.

Состоянием S базы данных называется такое сопоставление схемам отношений с именами $R^1, ..., R^k$ настоящих отношений $S(R^1), ..., S(R^k)$ с теми же наборами атрибутов, что подстановка их в соотношения схемы делает соотношения верными.

Множество всех схем отношений, которые получаются из базисных схем посредством формального применения к ним операций над отношениями из данного класса, как легко заметить, замкнуто относительно операций, т. е. является реляционной алгеброй (алгеброй правильно построенных выражений — термов). Это множество можно также рассматривать как множество всех возможных запросов к базе данных.

Применяя к выделенным соотношениям схемы базы данных операции над отношениями и аксиомы реляционной алгебры, строится отношение эквивалентности на множестве запросов (термов). Множество классов эквивалентностей относительно этого отношения является реляционной алгеброй, которую естественно назвать реляционной алгеброй базы данных.

Пользователь базы данных имеет свое представление о данных, то есть свои базисные отношения и свои определяющие соотношения между ними и, следовательно, свою алгебру запросов и, соответственно свою реляционную алгебру пользователя. Если пользователь "подключен" к базе данных, то его запросы, выраженные через его базисные схемы отношений, интерпретируются запросами к базе данных так, что выполняются соотношения, объявленные пользователем. Эти условия определяют отображение алгебры пользователя в алгебру базы данных, совместимое с рассматриваемыми операциями, то есть определяют гомоморфизм реляционных алгебр. Наоборот, если отображение на множестве запросов определяет гомоморфизм алгебр, то это отображение допускает "подключение" пользователя к базе данных.

Две схемы базы данных имеют одинаковые информационные возможности, если определяемые ими реляционные алгебры изоморфны. Однако эффективность работы базы данных может существенно зависеть от ее схемы, т. е. выбора образующих и определяющих соотношений. Таким образом, после построения реляционной алгебры базы данных перед проектировщиком системы стоит задача оптимального выбора образующих этой алгебры.

Введение точных понятий позволило установить основные структурные теоремы о строении алгебр баз данных.

Обозначим через $\mathcal{K}$ систему Кодда операций над отношениями с бесконечными пересечениями и объединениями отношений.

Пусть $\mathcal{R}$ и $\mathcal{R}'$ — две $\mathcal{K}$-алгебры. Через $\mathcal{R}\oplus\mathcal{R}'$ обозначается $\mathcal{K}$-алгебра, которая называется прямой суммой $\mathcal{K}$-алгебр. Ее элементами с набором атрибутов A являются пары (R, R'), где $R \in \mathcal{R}(A)$, $R' \in \mathcal{R}'(A)$, а все операции выполняются по каждой координате отдельно.

О п р е д е л е н и е I.5.2. $\mathcal{K}$-алгебра $\mathcal{R}$ называется неприводимой, если она не представляется в виде суммы своих нетривиальных подалгебр, то есть $\mathcal{R} \neq \mathcal{R}_1 \oplus \mathcal{R}_2$, где $\mathcal{R}_1 \neq \emptyset$ и $\mathcal{R}_2 \neq \emptyset$.

Т е о р е м а I.5.15. Пусть $\mathcal{R}$ — произвольная $\mathcal{K}$-алгебра с конечным множеством образующих. Тогда $\mathcal{R}$ представляется в виде прямой суммы неприводимых $\mathcal{K}$-алгебр $\mathcal{R} = \bigoplus\limits_{s\in Spec\mathcal{R}} \mathcal{R}_s$ по некоторому множеству $Spec\mathcal{R}$, однозначно определяемому алгеброй $\mathcal{R}$. Если $j : \mathcal{R} \to \mathcal{R}'$ — гомоморфизм

$\mathcal{K}$-алгебр, то он индуцирует естественное частичное отображение

$$j : Spec\mathcal{R}' \to Spec\mathcal{R}$$

и множество гомоморфизмов-вложений

$$j_{s'} : \mathcal{R}_{j(s')} \to \mathcal{R}'_{s'},$$

для всех $s' \in Spec\mathcal{R}'$, для которых определено отображение j так, что $j(R) = j(\bigoplus\limits_{s \in Spec\mathcal{R}} R_s) = \bigoplus\limits_{s' \in Spec\mathcal{R}'} j_{s'}(R_{j(s')})$, где R — произвольный элемент $\mathcal{R}$.

По этой теореме, произвольная реляционная алгебра с конечным множеством образующих представляется в виде канонической прямой суммы неприводимых реляционных алгебр. Поэтому, чтобы завершить описание реляционных алгебр, нужно дать описание неприводимых реляционных алгебр.

Пусть $\mathcal{D}$ — $\mathcal{K}$-алгебра действительных отношений на множестве D, где D — дизъюнктное объединение областей значений всех типов атрибутов базы данных.

Т е о р е м а I.6.1. Если $\mathcal{R}$ — неприводимая $\mathcal{K}$-алгебра с конечным множеством образующих, тогда существует гомоморфизм - вложение $\mathcal{R} \to \mathcal{D}$.

Так как, согласно следствиям I.5.13 и I.5.14, каждая подалгебра алгебры реальных отношений $\mathcal{D}$ неприводима, то, отсюда следует, что для того, чтобы описать все неприводимые реляционные алгебры, нужно описать все реляционные подалгебры алгебры $\mathcal{D}$.

Пусть A и B — наборы атрибутов. Через $\mathcal{A}(A, B)$ обозначается множество всех отображений (согласований) из A в B, сохраняющих тип атрибутов. Тогда каждое отношение R с набором атрибутов A можно преставить в виде подмножества $R \subset \mathcal{A}(A, D)$, где каждая строка отношения представляется в виде соответствующего отображения набора атрибутов A в объединение доменов D, сохраняющего типы атрибутов.

Обозначим через $S(D)$ множество всех взаимно однозначных отображений D на себя, сохраняющих типы атрибутов. Очевидно, что $S(D)$ образует группу относительно композиции отображений. Будем говорить, что отношение R симметрично относительно $h \in S(H)$, если отношение R

не изменяется после замены значений атрибутов под действием отображения h, то есть $h \circ R = R$, где через $\circ$, обозначена операция композиции отображений из $R \subset \mathcal{A}(A, D)$ и отображения $h \in \mathcal{D}$.

Пусть $\{R_i\}_{i \in I}$ — произвольное множество отношений. Обозначим через H множество всех отображений $h \in S(D)$, относительно которых симметричны отношения R_i, при $i \in I$. Очевидно, что множество отображений H образуют подгруппу в $S(D)$, которая называется группой симметрии множества отношений $\{R_i\}_{i \in I}$.

Т е о р е м а I.6.7 Полные реляционные подалгебры $\mathcal{R} \subset \mathcal{D}$ алгебры отношений $\mathcal{D}$ над категорией наборов атрибутов $\bar{\mathcal{A}}$ (без ограничения их конечности) находятся во взаимно однозначном соответствии с подгруппами $H \subset S(D)$ группы всех перестановок элементов области значений атрибутов D, сохраняющих тип элементов. Подалгебре $\mathcal{R}$ сопоставляется подгруппа H всех перестановок, относительно которых симметрично каждое отношение подалгебры $\mathcal{R}$. Подгруппе H сопоставляется подалгебра R, состоящая из всех отношений симметричных относительно элементов группы H.

В случае конечного множества $\mathcal{D}$, эта теорема I.6.7. впервые была получена М. Краснером (см. [62]), а также независимо получена в работе F. Banchilhon (см.[36]).

Т е о р е м а I.6.8. Пусть $\mathcal{R} \subset \mathcal{D}$ и $\mathcal{R}' \subset \mathcal{D}'$ — полные подалгебры алгебр отношений с областями значений атрибутов, соответственно, D и D', и $j : \mathcal{R} \to \mathcal{R}'$ — изоморфизм реляционных алгебр. Тогда найдется взаимно однозначное отображение $g : D \to D'$ области D на область D', сохраняющее типы элементов, такое, что $H = g^{-1} \circ H' \circ g$, где H и H' — подгруппы, соответствующие подалгебрам $\mathcal{R}$ и $\mathcal{R}'$. В этом случае для любого отношения $R \in \mathcal{R}$ соответствующее ему отношение $j(R) \in \mathcal{R}'$ имеет вид: $j(R) = g \circ R$.

Для описания реляционных подалгебр алгебры отношений с конечными множествами атрибутов введем в множестве всех обратимых отображений $S(D)$ области значений атрибутов на себя следующую топологию:

для любого подмножества $K \subset S(D)$ обратимое отображение $\bar{h} \in S(D)$ принадлежит замыканию $\bar{K}$ тогда и только тогда, когда ограничение отображения $\bar{h}$ на любое конечное подмножество $e_L : L \subset D$ принадлежит множеству отображений $K \circ e_L$.

Т е о р е м а I.6.14. Реляционные подалгебры $\mathcal{R} \subset \mathcal{D}$ алгебры отношений с конечными наборами атрибутов и счетной или конечной областью значений атрибутов D находятся во взаимно однозначном соответствии с замкнутыми подгруппами $\bar{H} \subset S(D)$ группы $S(D)$ в топологии, определенной выше. Подгруппа $\bar{H}$ является группой симметрии отношений подалгебры $\mathcal{R}$.

Эти теоремы дают полную классификацию неприводимых полных реляционных алгебр и являются обобщением соответствующих результатов М. Краснера и Ф. Банчилона (F. Banchilhon) для алгебр конечных отношений.

Глава II посвящена алгебраическому подходу к моделированию понятий и их представлению для автоматизированной работы с ними. В качестве примеров представляемых понятий рассматриваются абстрактные типы данных, схемы баз данных и общие понятия предметных областей.

Известно, что алгебраические методы явились теоретической основой для определения, исследования и построения языков спецификаций абстрактных типов данных. В этой главе показывается, что алгебраические методы, дополненные средствами теории категорий являются удобным инструментом представления и более общих понятий.

Категорный подход к представлению знаний является естественным развитием реляционного подхода к базам данных, дополненного идеями и методами абстрактных типов данных в программировании, при этом используются достижения приложений математической теории категорий к математической логике.

В категорном подходе к представлению знаний предполагается, что представляемые знания прикладной области могут быть структурированы в виде системы понятий. С каждым понятием связывается имя понятия, определение и знания, вытекающие из определения. Теория категорий предоставляет математические средства для отражения семантики и логики понятий. Каждому представляемому понятию в этом подходе сопоставляется алгебраическая модель понятия (категория), в которой потенциально отражается полное знание о предмете понятия, вытекающее из определения этого понятия, а также строится конечная аппроксимация категории (фрагмент категории), в которой отражаются уже имеющиеся знания о представляемом понятии, отображенные в системе представления и непосредственно доступные пользователю.

Средства категорного подхода к представлению знаний условно можно разделить на три основные части:

– средства построения алгебраической модели (категории);

– средства построения конечной аппроксимации категории;

– язык представления знаний.

Все три составляющие подхода сильно связаны между собой:

– категория и ее аппроксимация задаются языковыми средствами;

– корректность новых языковых выражений определяется по уже построенной аппроксимации;

– категория, отражающая полные знания о представляемом понятии, — это идеальный объект, который является пределом своих конечных аппроксимаций.

В основе категорного подхода к представлению знаний также, как и в основе реляционного подхода к моделированию баз данных, лежит некоторый набор операций. Они называются категорными операциями, так как они были введены в математической теории категорий.

Категорные операции позволяют выразить представляемые понятия и определить требуемую их обработку. Особенностью категорных операций является возможность построения любых теоретико множественных конструкций из областей, не все элементы которых известны.

Единая категорная интерпретация понятий, основанная на общем наборе операций, позволяет достичь высокой степени интеграции понятий в систему знаний.

Глава III посвящена изучению алгебраических объектов, используемых для задач представления понятий.

Требование представления сложных понятий с использованием декартовых произведений, выделения элементов в областях с использованием финального объекта, выделением подобъектов и отношений для областей, в которых не все элементы известны, приводит к необходимости использования категорных конструкций топоса.

В приложениях топосов к моделированию баз данных особую роль играют булевы топосы, в которых реализуется классическая логика, и которые могут быть конструктивно заданы. В связи с этим рассматривается понятие алгебраического булевого топоса, заданного конечным множеством образующих и соотношений, в котором образующие-области (объекты) удовлетворяют условию конечности их мощности, выраженном на языке теории категорий. Такие топосы называются алгебраическими топосами конечного типа.

Топос называется непротиворечивым, если элементы его классификатора подобъектов (области истин) $true : I \to \Omega$ и $false : I \to \Omega$ не совпадают.

Топос называется полным или двузначным, если множество элементов его классификатора подобъектов Ω состоит лишь из двух элементов $true : I \to \Omega$ и $false : I \to \Omega$. Условие полноты топоса соответствует тому, что любое утверждение в полном топосе либо истинно, либо ложно, то есть полнота топоса соответствует полноте теории в исчислении предикатов.

Примерами полных булевых топосов являются категории $GFinset$, где G — произвольная конечная группа. Объектами категории $GFinset$ являются конечные множества M вместе с выделенным действием группы G на них перестановками элементов множества M. Морфизмами категории $GFinset$ являются все отображения G-множеств $f : M \to N$, сохраняющие действия группы G.

Оказывается, категории вида $GFinset$ исчерпывают примеры полных топосов конечного типа.

Т е о р е м а III.1.3. Если $\mathcal{C}$ — полный топос конечного типа, то существует такая конечная группа G, что категория $\mathcal{C}$ эквивалентна категории $GFinset$.

В частности, отсюда следует следующий результат:

Т е о р е м а III.1.19. Любой подтопос $\mathcal{C}$ топоса всех конечных множеств $Finset$, порожденный конечным множеством несущих множеств $D_1, ..., D_n$, и конечными множествами функций и отношений $f_1, ..., f_k; r_1, ..., r_m$ на множествах, построенных из несущих множеств операциями топоса, совпадает с подтопосом G-множеств, где $G = S(\mathcal{C})$ — группа симметрии этих образующих топоса $\mathcal{C}$.

Известно, что категории $G_1Finset$ и $G_2Finset$ конечных групп G_1 и G_2 эквивалентны тогда и только тогда когда эти группы изоморфны.

Эти результаты позволяет описать произвольные топосы конечного типа.

Т е о р е м а III.1.20. Пусть $\mathcal{C}$ — произвольный топос конечного типа. Тогда существует такой конечный набор конечных групп $G_1, ..., G_k$, что категория $\mathcal{C}$ эквивалентна конечному произведению топосов $G_iFinset$, для $i = 1, ..., k$, то есть $\mathcal{C} \approx G_1Finset \times ... \times G_kFinset$.

Если категорными средствами нужно отражать знания не только о внешнем мире, но и о самих средствах представления знаний, и о всей системе представления знаний в целом, то требуется расширение операций топоса, приводящее к понятию рефлексивного топоса, определяемого ниже.

Пусть в топосе $\mathcal{E}$ выделены два объекта NOb и $NMor$, которые называются область имен объектов и область имен морфизмов, соответственно. Множества элементов объектов NOb и $NMor$, то есть множества $\mathcal{E}(I, NOb)$ и $\mathcal{E}(I, NMor)$, где I — финальный объект категории $\mathcal{E}$, называются множеством имен объектов и множеством имен морфизмов, соответственно.

Предполагается, что между объектами NOb и $NMor$ заданы такие морфизмы, что они определяют внутреннюю категорию в $\mathcal{E}$ и индуцируют на множестве элементов объектов NOb и $NMor$ структуру категории, которая называется категорией имен и обозначается $NCat$.

Определение. Топос $\mathcal{E}$ вместе с внутренней категорией и категорией имен $NCat$ называется рефлексивным топосом, если существуют функторы $Name : \mathcal{E} \to NCat$ и $Denote : NCat \to \mathcal{E}$ такие, что их композиция $Denote \circ Name = 1\mathcal{E}$ является тождественным функтором на $\mathcal{E}$. Функтор $Name$ называется функтором именования, а функтор $Denote$ — денотатом.

Теорема III.2.1 Существует непротиворечивый рефлексивный топос.

В отличие от абстрактных типов данных, которые моделируются многоосновными алгебраическими системами, общие понятия и схемы баз данных, которые моделируются категориями с категорными операциями, требуют привлечения алгебраических средств, в которых операции действуют при выполнении некоторых условий. Примером такой операции является операция композиции отображений: композиция двух отображений f_1 и f_2 определена, если только область значений f_1 совпадает с областью определения f_2, то есть выполняется равенство $codom(f_1) = dom(f_2)$. Многие другие категорные операции также являются условными. Более того, условные операции требуются и для представления понятий, возникающих в предметных областях.

В диссертации исследуется понятие алгебры с условными операциями, порожденной словарем терминов и множеством условных операций и соотношений E. Частными случаями таких алгебр являются многоосновные алгебры, алгебраические топосы, рефлексивные топосы, алгебраические топосы с выделенными подобъектами и наследованием морфизмов. Рассмотрены понятие линейно излагаемой спецификации алгебры с условными операциями и понятие конечной аппроксимации таких алгебр.

Публикации Результаты диссертации опубликованы в 14 работах [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 37, 38, 39].

Благодарности. Я благодарю своих близких - мою жену и детей, за терпение, своих коллег и друзей за постоянное многолетнее сотрудничество, Российский Государственный Гуманитарный Университет за предоставленную возможность работать над диссертацией и Российский Фонд Фундаментальных Исследований за частичную материальную поддержку во время работы над диссертацией (грант номер 94-01-01479).

Литература

[1] Агафонов В. Н. *Спецификация программ: понятийные средства и их организация.* Новосибирск: Наука, Сиб.отд-ние, 1990.

[2] Агафонов В. Н. *Типы и абстракция данных в языках программирования.* В кн. Данные в языках программирования. М.:Мир,1982, с.263–327.

[3] Бениаминов Е. М. *Алгебраический подход к моделям баз данных реляционного типа.* В кн.:Семиотика и информатика, 1980, вып.14, с.44-80.

[4] Бениаминов Е. М. *Алгебраическая стуктура реляционных моделей баз данных.* НТИ, сер.2, 1980, N9, с. 23-25.

[5] Бениаминов Е. М., Березина Н. А. *Об алгебраическом подходе к описанию схем баз данных.*// В сб. Вопросы создания Автоматизированной системы НТИ по документам ГАФ СССР, Москва: ГАУ при Совете Министров СССР, ВНИИДАД, 1981, с.69-77.

[6] Бениаминов Е. М. *О роли симметрии в реляционных моделях баз данных и логических структурах.* НТИ, сер.2, 1984, N5, с.17-25.

[7] Бениаминов Е. М., Березина Н. А., Дунская М. В. *Разработка методов моделирования автоматизированной обработки, поиска и размещения данных в больших информационных системах.* СИФ ОЦНТИ, ВНИИДАД, депонированная рукопись, инв. N031–85,М.,1985.

[8] Бениаминов Е. М. *О некотором подходе к представлению знаний.* Тезисы докладов 4 Всесоюзн.конференции "Применение методов мат. логики" г. Таллин, 1986, с.34-36.

[9] Бениаминов Е. М. *Основания категорного подхода к представлению знаний. Категорные средства.* Изв. АН СССР Техн. кибернет.,N 2, 1988 , с.21–33.

[10] Бениаминов Е. М. *Рефлексивные топосы в категорном подходе к представлению знаний.* Тезисы докладов Всесоюзн. школы-семинара

"Семиотические аспекты формализации интеллектуальной деятельности" г. Боржоми, 1988, с.111-113.

[11] Бениаминов Е.М., Вайнтроб А. Ю. *Основные принципы диалогового языка для представления знаний средствами категорного подхода.* Материалы конференции ДИАЛОГ-87, г. Тбилиси, 1988, с.174-177.

[12] Бениаминов Е. М. *Алгебраические системы и типы данных.* //В кн.:Системное и теоретическое программирование, Ростов-н-Д: РГУ, 1988, с.83-92.

[13] Бениаминов Е. М. *Система представления и обработки понятий, основанная на алгебраическом (категорном) подходе.* Труды II Всесоюзной конференции "Искусственный интеллект- 90", т.2, 1990, с.8-11.

[14] Вигнер П. *Программирование на языке АДА.* М.:Мир, 1983.

[15] Глушков В. М., Цейтлин Г. Е., Ющенко Е. Л. *Алгебра, логика, языки, программирование.* Киев:Наукова думка, 1974.

[16] Голдблатт Р. *Топосы. Категорный анализ логики.* М.:Мир, 1983.

[17] Гончаров С. С., Ершов Ю. Л., Свириденко Д. И. *Семантические основы логического программирования*// В сб. Концептуализация и смысл, под ред. Полякова И. В., Новосибирск:Наука, 1990, с.6-20.

[18] Джонстон П. Т. *Теория топосов.* М.:Наука, 1986.

[19] Жожикашвили А.В., Стефанюк В.Л. *Теория категорий в задачах представления знаний и обучения.* Изв. АН СССР. Техн. кибернет., N 2, 1986.

[20] Замулин А. В. *Системы программирования баз данных и знаний.* Новосибирск:Наука, 1990.

[21] Калиниченко Л. А. *Методы и средства интеграции неоднородных баз данных.* М.:Наука, 1983.

[22] Калиниченко Л. А., Рывкин В. М. *Машины баз данных и знаний.* М.:Наука, 1990.

[23] Капитонова Ю. В., Летичевский А. А. *Математическая теория проектирования вычислительных систем.* М.:Наука, 1988.

[24] Кондрашина Е. Ю., Литвинцева Л. В., Поспелов Д. А. *Представление знаний о времени и пространстве в интеллектуальных системах.* М.:Наука, 1989.

[25] Кучеров Г. А. *Системы подстановок термов.* Препринт 601, ВЦ АН СССР Сиб. отделение, Новосибирск, 1985.

[26] Мальцев А. И. *К общей теории алгебраических систем.* Мат. сборник, 1954, т.35, вып.1.

[27] Массер Д. *Спецификация абстрактных типов данных в системе AFFIRM.*// В сб.: Требования и спецификации в разработке программ, М.:Мир, 1984, с.199-222.

[28] Плоткин Б. И. *Универсальная алгебра, алгебраическая логика и базы данных.* М.:Наука, 1991.

[29] Поспелов Д. А. *Логико-лингвистические модели в системах управления.* М.:Энергоиздат, 1981.

[30] Свириденко Д. И. *Проект Сигма. Цели и задачи.* // В сб. Логические методы в программировании под ред. Ершова Ю. Л. (Вычислительные системы, вып. 133), РАН, Сиб. отд.-ние, Ин.-т математики, Новосибирск, 1990, с.68-94.

[31] Цаленко М. Ш. *Моделирование семантики в базах данных.* М.:Наука, 1989.

[32] Шенфилд Дж. *Математическая логика.* М.:Наука, 1975.

[33] Andréka H., Németi I. *Applications of universal algebra, model theory and categories in computer science (Part 1)* Comput. Lingust. Comput. Lang., 13, 1979, p. 152-282.

[34] Andréka H., Németi I. *Applications of universal algebra, model theory and categories in computer science (Part 2)* Comput. Lingust. Comput. Lang., 14, 1980, p. 43-65.

[35] Andréka H., Németi I. *Applications of universal algebra, model theory and categories in computer science (Part 3)* Lect. Notes in Comp. Sci., V.117, Springer-Verlag, Berlin, 1981, p. 281-290.

[36] Banchilhon F. *On the Compliteness of Query Language for Relational Data Bases.* Lect. Notes in Comp. Sci., V.64, Springer-Verlag, 1978, pp.76-98.

[37] Beniaminov E. M. *Concept Bases and Algebraic Modeling Methods.* Proceedings of the International Workshop on Advances in Databases and Information Systems (ADBIS'94), 1994, p.133-135.

[38] Beniaminov E. M. *A Categorical Approach to Knowledge Representation.* Japan-CIS Symposium on Knowledge Based Software Engineering'94 (JCKBSE'94), 1994, p.181-182.

[39] Beniaminov E. M. *Algebraic Invariants of Database Schemes.* Proceedings of the Second International Workshop on Advances in Databases and Information Systems (ADBIS'95), V.1, 1995, p.259-263.

[40] Breu R. *Algebraic Specification Techniques in OOP Environments.* Lect. Notes in Comp. Sci., V.562, Springer-Verlag, 1991.

[41] Burstall R. M., Goguen J. A. *The Semantics of Clear, a Specification Language* In: Proceedings, of the 1979 Copenhagen Winter Shcool on Abstrakt Software Specification, Springer-Verlad, Berlin, 1980, 292-332.

[42] Chandra A. K., Harel D. *Computable queries for relational data bases.* J. Comput. & Syst. Sci., V.21, N2, 1980, p.156-172.

[43] Codd E. F. *A relational model for large shared data banks.* Comm. of ACM 13, 6, 1970, p.377-387.

[44] Dershowitz N. *Orderings for term-rewriting systems.* Theor. Comput. Sci., 1982, V.17, 3, pp.279-302.

[45] Elmasri R., Weeldrever J., Hevner A. *The category concept: an extension to the entity-relatconship model* Data & Knowledge Engineering, V.1, N1,1985.

[46] Ehrig H., Mahr B. *Fundamentals of Algebraic Specification.* V.1, Springer-Verlag, 1985.

[47] Ehrig H., Mahr B. *Fundamentals of Algebraic Specification.* V.2, Springer-Verlag, 1990.

[48] *First International Symposium on Category Theory Applied to Computation and Control* Lect. Notes in Com. Sci, V.25, Springer - Verlad, Berlin, 1975.

[49] Georgescu I. *A Categorial approach to knowledge-based systems.* Computers and Artifical Intelligence, V.3, N2, 1984, pp.105-113.

[50] Goguen J. A., Thatchar J. W., Wagner E. G. *An inititial algebra approach to the specification, correctness and implementation of abstract data types.* In Current Trends in Programming Methodology IV: Data Structuring, Prentice Hall, 1978, pp.80-144.

[51] Goguen J. A. *Some design principles and theory for OBJ-O, a language to express and execute algebraic specifications of problems.* In: Lect. Notes Comput. Sci.,V.75, 1979, pp.425-473.

[52] Goguen J. A., Burstall R. M. *Some fundamental algebraic tools for the semantics of computation.* Theoretical Computer Science, V.31, N2, 3, 1984.

[53] Goguen J. A., Burstall R. M. *Introducing Institution.* Lect. Not. Comp.Sci., V.164, 1984.

[54] Goguen J. A., Meseguer J. *Equality, types, modules, and (why not?) generics for logic programming.* Conf. of Logical Prog., Uppsala, Sweden, 1984, 179-210.

[55] Goguen J. A., Meseguer J. *Eqlog: Equality, types, and generic modules for logic programming.* In Douglas SeGroot and Gary Lindstrom, eds. "Logic Programming: Functions, Relations and Equations", Prentice-Hall, Englewood Cliffs, N.J., 1986, pp.295-363.

[56] Goguen J. A., Meseguer J. *Order-sorted algebra solves the constructor selector, multiple representation and coercion problems.* Symposium on Logic in Comp. Sci., IEEE Comp. Society Press, 1987, pp.18-29.

[57] Goguen J. A., Burstall R. M. *Institutions: Abstract Model Theory for Specification and Programming.* Jorn. of ACM, V.39, 1, 1992, pp.95-146.

[58] Grothendieck A., Verdier J. L. *Théorie des Topos.* (SGA 4, exposés I-VI).–Second edition.–Berlin; Heidelberg; N. Y.:Springer, 1972.

[59] Huet G. *Confluent reductions: abstract properties and applications to term rewriting systems.* Journ. of ACM, 1980, v.27, N4, pp.797-821.

[60] Kaplan S. *Symplifying conditional term rewriting systems: unification, termination and confluence.* Jorn. Symbolic Computation 4(3), 1987, pp.295-334.

[61] Knuth D., Bendix P. *Simple word problems in universal algebras.* In: Conputational Problems in Abstract Algebra, Pergamon Press, Elmsford, New York, 1970, pp.263-297.

[62] Krasner M. I. *Generalization et analogues de la theorie de Galois.* Comptes Rendus de Congress de la Victorie de I'Ass. Franc. pour I'Avancem. Sci., 1945, pp. 54-58.

[63] Lawvere F. W. *Functorial semantics of algebraic theories.* Proc. Nat. Acad. Sci., 1963, V.50, N5, pp.869-872.

[64] Lawvere F. W. *Introduction* In: Toposes, Algebraic Geometry and Logic, Lect. Notes in Math., V.174, 1972.

[65] Maltsev A. *Algebraic Systems.* Springer-Verlag, 1973.

[66] Melton A., Schmidt D., and Strecher G. *Galois connections and computer science applications.* Lect. Notes in Comp. Sci., V. 240, Springer-Verlag, 1986.

[67] *Proceedings, Category Theory and Computer Programming,* Lect. Notes in Comp. Sci., V. 240, Springer-Verlag, 1986.

[68] *Proceedings, Category Theory and Computer Science,* Lect. Notes in Comp. Sci. V. 283, Springer-Verlag, 1987.

[69] Rydeheard D.F., Burstall R.M. *Computational category theory.* Prentice Hall, 1988.

[70] Tuijn C., Gyssens M., Paredaens J. *A Categorical Approach to Object-Oriented Data Modelling.* Proceedings of Third Workshop on Foundation of Models and Languages for Data and Objects, Aigen, 1991, pp.187–196.

[71] Rusinowitch M., Rémy J. L. (Eds.) *Conditional Term Rewriting Systems.* Lect. Notes in Comp. Sci., V. 656, Springer-Verlag, 1993.

[72] Zilles S.N. *Introduction to data algebras.* Lect. Notes Comput. Sci., V.86, Springer-Verlag, 1980.