



Серия «Математика»

2011. Т. 4, № 3. С. 68–73

Онлайн-доступ к журналу:

<http://isu.ru/izvestia>

---

---

ИЗВЕСТИЯ

Иркутского  
государственного  
университета

---

---

УДК 518.517

## Алгебры Кодда и дескриптивные логики

И. А. Казаков

*Иркутский государственный университет*

**Аннотация.** В статье рассматриваются вопросы моделирования алгебр Кодда в рамках дескриптивных логик. В частности, показана выразимость операций реляционной алгебры Кодда в терминах объектных теорий, моделирующих базы данных. Результаты работы имеют как теоретическое, так и практическое значение, поскольку разрабатываемый метод позволяет работать с базами данных как онтологиями в логических базах знаний.

**Ключевые слова:** онтологии; базы данных; объектные теории; дескриптивные логики; Libretto.

### 1. Введение

Одной из актуальных проблем является задача вовлечения информации, накопленной в базах данных, в системы логической обработки. Поскольку базы данных и логические системы основаны на очень разных формализмах, эта задача не имеет простого решения [6]. Здесь можно воспользоваться подходами, используемыми сегодня в программировании, которые больше опираются на методы погружения реляционных описаний в объектно-ориентированные структуры. Мы считаем, что такой подход может оказаться полезным и для погружения информации, представленной в реляционных форматах, в системы обработки знаний.

В рамках этого подхода в статье [1] нами было аксиоматизировано понятие реляционной базы данных в терминах объектных дескриптивных логик. Мы описали понятие реляционной БД с точки зрения объектных теорий [2] и определили правила формирования по базе данных специальной объектной теории, семантически эквивалентной этой базе. Также нами были описаны условия, реализующие замкнутость мира. Эти условия позволяют избавиться от нерелевантных моделей этой объектной теории, т. е. тех моделей, внутренняя структура которых противоречит природе реляционных баз.

В данной работе показывается, что можно ввести преобразования объектных теорий баз данных, которые соответствуют преобразованиям баз данных при применении операций алгебры Кодда. Кроме того, демонстрируется определимость этих операций на объектно-итерационном языке программирования Libretto [4] [5].

## 2. Операции алгебры Кодда

Под базой данных понимается набор таблиц, в [1] даются формальные определения таблицы (табличного отношения) и базы данных.

Ниже приведены основные операции реляционной алгебры Кодда. Известно (см. напр. [3]), что базовый список операций алгебры Кодда состоит из операций выборки ( $\sigma$ ), проекции ( $\pi$ ), объединения ( $\cup$ ), вычитания ( $-$ ) и декартова произведения ( $\times$ ). Все остальные операции (типа пересечения  $\cap$ , соединения  $\bowtie$  и деления  $\div$ ) выразимы через базовые.

### 2.1. ОПЕРАЦИЯ ВЫБОРКИ

Будем говорить, что *атрибуты*  $A$  и  $B$  некоторой базы данных  $\mathbb{DB}$  *согласованы*, если они имеют один и тот же тип данных в  $\mathbb{DB}$ . Также, будем говорить, что *атрибут*  $A$  *согласован* с некоторым элементом  $v \in D$ , где  $D \in \mathbb{D}$ , если атрибут  $A$  имеет тип данных  $D$  в  $\mathbb{DB}$ .

**Определение 1.** [*Предикат выборки*] Пусть  $\mathbb{DB}$  — база данных. Выражение  $A \text{ rel } x$  назовем элементарным предикатом выборки, где  $A \in \text{sign}(\mathbb{DB})$ ,  $x \in (\text{sign}(\mathbb{DB}) \cup \mathbb{D})$ ,  $A$  согласован с  $x$  и  $\text{rel} \in \{=, \neq, <, \leq, >, \geq\}$ . Предикат выборки — это либо элементарный предикат выборки, либо один из  $\neg p_1$ ,  $p_1 \vee p_2$ ,  $p_1 \wedge p_2$ , где  $p_1, p_2$  — предикаты выборки.

Предикат выборки  $p$  определен на отношении  $\mathbf{r} = \langle R_{\mathbf{r}}, \text{body}_{\mathbf{r}}, \text{header}_{\mathbf{r}} \rangle$ , если все атрибуты, входящие в  $p$ , принадлежат сигнатуре  $\mathbf{r}$ . Предикат  $p$  истинен на кортеже  $t \in \text{body}_{\mathbf{r}}$  (обозначаем  $\mathbb{DB} \models p(t)$ , где  $\mathbb{DB}$  — база данных, содержащая отношение  $\mathbf{r}$ ), если  $p$  определен на  $\mathbf{r}$ , и при подстановке вместо каждого атрибута  $A$ , входящего в  $p$ , значения  $t(A)$ , получившееся булево выражение является истинным, где истинность элементарных предикатов определяется истинностью полученных равенств и неравенств в соответствующих доменах.

Пусть предикат выборки  $p$  определен на отношении  $\mathbf{r}$ , в таком случае на  $\mathbf{r}$  определена *операция выборки*:

$$\sigma_p(\mathbf{r}) = \mathbf{r}'$$

где  $\mathbf{r}'$  – отношение с новым именем  $R_{\mathbf{r}'} \in \mathbb{ID}$ , такое, что

$$\begin{aligned} body_{\mathbf{r}'} &= \{t \mid t \in body_{\mathbf{r}} \text{ и } \{\mathbf{r}\} \models p(t)\}, \\ header_{\mathbf{r}'} &= header_{\mathbf{r}} \end{aligned}$$

Операция выборки  $\sigma_p$ , примененная к отношению  $\mathbf{r}$  базы данных  $\mathbb{DB}$ , преобразует ее в базу данных  $\mathbb{DB}' = \mathbb{DB} \cup \{\sigma_p(\mathbf{r})\}$ .

## 2.2. ОПЕРАЦИЯ ПРОЕКЦИИ

Теперь определим на табличных отношениях *операцию проекции*, позволяющую извлекать заданные атрибуты из отношения. Пусть  $\mathbf{r} = \langle R_{\mathbf{r}}, body_{\mathbf{r}}, header_{\mathbf{r}} \rangle$  – отношение и  $\{A_1, \dots, A_s\} \subseteq sign(\mathbf{r})$ . Тогда

$$\begin{aligned} \pi_{A_1, \dots, A_s}(\mathbf{r}) &= \mathbf{r}' \\ body_{\mathbf{r}'} &= \{\langle t(A_1), \dots, t(A_s) \rangle \mid t \in body_{\mathbf{r}}\} \\ header_{\mathbf{r}'} &= \langle \langle A_1, D_1 \rangle, \dots, \langle A_s, D_s \rangle \rangle, \end{aligned}$$

где  $D_i$  – типы данных атрибутов  $A_i$  в  $\mathbf{r}$ .

## 2.3. ОПЕРАЦИЯ ОБЪЕДИНЕНИЯ

Пусть  $\mathbf{r} = \langle R_{\mathbf{r}}, body_{\mathbf{r}}, header_{\mathbf{r}} \rangle$  – табличное отношение с заголовком  $header_{\mathbf{r}} = \langle \langle A_1, D_1 \rangle, \dots, \langle A_k, D_k \rangle \rangle$  и  $\pi : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$  – перестановка. Будем говорить, что *отношение*  $\mathbf{r}_{\pi} = \langle R_{\mathbf{r}_{\pi}}, body_{\mathbf{r}_{\pi}}, header_{\mathbf{r}_{\pi}} \rangle$  получено из отношения  $\mathbf{r}$  с помощью перестановки  $\pi$ , если  $header_{\mathbf{r}_{\pi}} = \langle \langle A_{\pi(1)}, D_{\pi(1)} \rangle, \dots, \langle A_{\pi(k)}, D_{\pi(k)} \rangle \rangle$ ,  $body_{\mathbf{r}_{\pi}} = \{t_{\pi} \mid t \in body_{\mathbf{r}}\}$ , где  $t_{\pi} = \langle t_{\pi(1)}, \dots, t_{\pi(k)} \rangle$  перестановка кортежа  $t = \langle t_1, \dots, t_k \rangle$ .

Операция объединения действует на табличных отношениях, таких, что заголовок одного из табличных отношений при некоторой перестановке совпадает с заголовком другого табличного отношения. Пусть  $\mathbf{r}^1$  и  $\mathbf{r}^2$  такие табличные отношения, что для некоторой перестановки  $\pi$  выполнено  $header_{\mathbf{r}^1} = header_{\mathbf{r}^2_{\pi}}$ . Определим операцию объединения:

$$\begin{aligned} \mathbf{r}^1 \cup \mathbf{r}^2 &= \mathbf{r}' \\ body_{\mathbf{r}'} &= body_{\mathbf{r}^1} \cup body_{\mathbf{r}^2_{\pi}} \\ header_{\mathbf{r}'} &= header_{\mathbf{r}^1} \end{aligned}$$

С точки зрения заголовка  $header_{\mathbf{r}'}$  операция объединения не является коммутативной (получаются разные порядки имен атрибутов). Но мы готовы это терпеть, чтобы избежать дублирования имен и появления синонимов.

## 2.4. ОПЕРАЦИЯ ВЫЧИТАНИЯ

Операция вычитания также действует на табличных отношениях, заголовки которых при некоторой перестановке совпадают. При подходящей перестановке  $\pi$  вычитание табличных отношений определяется как

$$\begin{aligned} \mathbf{r}^1 - \mathbf{r}^2 &= \mathbf{r}' \\ \text{body}_{\mathbf{r}'} &= \text{body}_{\mathbf{r}^1} \setminus \text{body}_{\mathbf{r}^2_\pi} \\ \text{header}_{\mathbf{r}'} &= \text{header}_{\mathbf{r}^1} = \text{header}_{\mathbf{r}^2_\pi} \end{aligned}$$

## 2.5. ОПЕРАЦИЯ ДЕКАРТОВА ПРОИЗВЕДЕНИЯ

Пусть  $\mathbf{r}^1$  и  $\mathbf{r}^2$  — отношения такие, что  $\text{sign}(\mathbf{r}^1) \cap \text{sign}(\mathbf{r}^2) = \emptyset$ . Тогда

$$\begin{aligned} \mathbf{r}^1 \times \mathbf{r}^2 &= \mathbf{r}' \\ \text{body}_{\mathbf{r}'} &= \{t_1 \cdot t_2 \mid t_1 \in \text{body}_{\mathbf{r}^1} \text{ и } t_2 \in \text{body}_{\mathbf{r}^2}\} \\ \text{header}_{\mathbf{r}'} &= \text{header}_{\mathbf{r}^1} \cdot \text{header}_{\mathbf{r}^2} \end{aligned}$$

где  $\cdot$  — операция конкатенации (присоединения) кортежей.

## 3. Выразимость операций алгебры Кодда

В работе [1] приведен алгоритм построения по базе данных  $\mathbb{DB}$  объектной теории базы данных  $\mathbb{O} + \mathbb{CW}$  с аксиомами замыкания (далее объектной теории).

**Теорема 1.** *Операции алгебры Кодда определимы в объектных теориях баз данных.*

*Доказательство.* Мы приведем доказательство этой теоремы для операции выборки.

Прежде всего нам необходимо формализовать в дескриптивной логике предикаты выборки из определения 1. Для этого введем два новых вида концептов, которые назовем атомарными концептами выборки:

$$P_1 \text{ rel } P_2 \quad P_1 \text{ rel } v$$

где  $\text{rel} \in \{=, \neq, <, \leq, >, \geq\}$ ,  $P_1, P_2$  — т-свойства,  $v \in \mathcal{D}$ . Определим семантику атомарных концептов следующим образом:

$$\begin{aligned} (P_1 \text{ rel } P_2)^I &= \{x \mid I \models \exists v_1 \exists v_2 (P_1(x, v_1) \wedge P_2(x, v_2) \wedge (v_1 \text{ rel } v_2))\} \\ (P_1 \text{ rel } v)^I &= \{x \mid I \models \exists v_1 (P_1(x, v_1) \wedge (v_1 \text{ rel } v))\} \end{aligned}$$

Произвольный концепт выборки получается из атомарных в виде их булевой комбинации с использованием связок  $\sqcup, \sqcap, \neg$ .

Пусть  $\mathbb{O} + \text{CW}$  — объектная теория базы данных  $\mathbb{DB}$  и  $p$  — предикат выборки, определенный на некотором отношении из  $\mathbb{DB}$ . Обозначим через  $E_p$  концепт выборки, полученный из  $p$  заменой всех атрибутов  $A$  на т-свойства  $P_A$ , а  $\vee, \wedge$  на  $\sqcup, \sqcap$  соответственно, в таком случае будем писать  $E_p = \text{tr}(p)$  (здесь и далее  $\text{tr}$  преобразование, определенное в [1]). Легко проверяется, что концепты выборки ведут себя аналогично предикатам выборки:

**Лемма 1.** *Если  $\mathbb{O} + \text{CW}$  — объектная теория базы данных  $\mathbb{DB}$ , то для любого  $t \in \text{body}_{\mathbf{r}}$ ,  $\mathbb{DB} \models p(t)$  тогда и только тогда, когда  $\mathbb{O} \vdash \{o\} \sqcap E_p \sqcap C$ , где  $E_p = \text{tr}(p)$ ,  $o = \text{tr}(t)$  и  $C = \text{tr}(\mathbf{r})$ .*

*Доказательство.* Лемма доказывается индукцией по длине концепта  $E_p$ . Напомним, что выводимость  $\mathbb{O} \vdash \{o\} \sqcap E_p \sqcap C$  в соответствии с семантикой дескриптивных логик означает, что в любой интерпретации  $I$  объектной теории  $\mathbb{O}$  выполняется  $(\{o\} \sqcap E_p \sqcap C)^I \neq \emptyset$ .

Пусть  $E$  — концепт выборки. Будем говорить, что операция выборки с концептом  $E$  (операция  $\sigma_E^*$ ) «применима» к концепту  $C$  в теории  $\mathbb{O} + \text{CW}$ , если все т-свойства, встречающиеся в определении концепта  $E$ , лежат в словаре теории  $\mathbb{O} + \text{CW}$ , и для любого такого т-свойства  $P$ , если  $\exists P \sqsubseteq B \in \mathbb{O}$  для некоторого концепта-столбца  $B$ , то  $C \sqsubseteq B \in \mathbb{O}$ .

Приступим к определению операции выборки в объектной теории  $\mathbb{O} + \text{CW}$ . Пусть  $\mathcal{W} = \langle \mathcal{C}, \mathcal{R}, \mathcal{P}, \mathcal{O} \rangle$  — словарь теории  $\mathbb{O} + \text{CW}$ . Пусть  $E$  — концепт выборки и операция выборки  $\sigma_E^*$  «применима» к концепту  $C$  теории  $\mathbb{O} + \text{CW}$ . Далее по шагам построим теорию  $\mathbb{O}' + \text{CW}'$  в словаре  $\mathcal{W}' = \langle \mathcal{C}', \mathcal{R}, \mathcal{P}, \mathcal{O}' \rangle$ :

Шаг 1. Расширяем словарь  $\mathcal{W}$  именем нового концепта  $C' \notin \mathcal{C}$ , положим  $C' = C \cup \{C'\}$ .

Шаг 2. Рассмотрим множество  $\mathcal{O}_{C,p} = \{o_1, \dots, o_l\}$ , состоящее из всех таких констант  $o \in \mathcal{O}$ , что  $\mathbb{O} + \text{CW} \vdash \{o\} \sqcap C \sqcap E$ . Для каждой константы  $o \in \mathcal{O}_{C,p}$  добавим в  $\mathcal{O}$  новую константу  $o' \notin \mathcal{O}$ , т. е. положим  $\mathcal{O}' = \mathcal{O} \cup \{o'_1, \dots, o'_l\}$ .

Шаг 3. Добавим в  $\mathbb{O}$  аксиомы  $C'(o'_1), \dots, C'(o'_l)$  и для каждой аксиомы  $P(o, v) \in \mathbb{O}$ , где  $o \in \mathcal{O}_{C,p}$ , добавим в  $\mathbb{O}$  аксиому  $P(o', v)$ .

Шаг 4. Добавим в  $\mathbb{O}$  все аксиомы для  $C'$  вида  $C \sqsubseteq B$  и  $P(o, t(A))$  (например, добавляем  $C' \sqsubseteq B$ , если  $C \sqsubseteq B \in \mathbb{O}$ ).

Шаг 5. Изменяем аксиому  $\top \sqsubseteq \bigsqcup_{o \in \mathcal{O}} \{o\}$  и аксиомы вида  $B \sqsubseteq C^1 \sqcup \dots \sqcup C^l$  и  $\exists P \sqsubseteq \{o_1\} \sqcup \dots \sqcup \{o_m\}$  в  $\text{CW}$  так, чтобы они учитывали появление  $C'$  и новых констант.

Определение остальных операций дается аналогично.

## Список литературы

1. Казаков И. А. Базы данных как онтологии / И. А. Казаков, А. В. Манцивода // Изв. Иркут. гос. ун-та. Сер. Математика. – 2011. – Т. 4, № 1. – С. 20–30.
2. Малых А. А. Объектно-ориентированная дескриптивная логика / А. А. Малых, А. В. Манцивода // Изв. Иркут. гос. ун-та. Сер. Математика. – 2011. – Т. 4, № 1. – С. 57–72.
3. Abiteboul S. Foundations of Databases / S. Abiteboul, R. B. Hull, V. Vianu. – Addison-Wesley, 1995. – 685 p.
4. Malykh A. A Query Language for Logic Architectures / A. Malykh, A. Mantsivoda // Proceedings of 7th International Conference «Perspectives of System Informatics». – Springer-Verlag Berlin Heidelberg, Lecture Notes in Computer Science 5947. – 2010. – P. 294–305.
5. Libretto: объектно-итерационный язык и система управления объектными базами знаний [Электронный ресурс]. – URL: <http://ontobox.org>.
6. Rosati R. On Combining Description Logic Ontologies and Nonrecursive Datalog Rules / R. Rosati // Lecture Notes in Computer Science. – 2008. – Vol. 341. Web Reasoning and Rule Systems. – P. 13–27.

---

**I. Kazakov****A formalization of the Codd algebra operations in logic  $SHOIN(D)$** 

**Abstract.** In this paper the problem of the interpretation of data bases as ontologies is investigated, in particular, a modeling of the Codd algebra operations via object theories. The problem of modeling the DB's closed world within the open world of description logics is solved. It is shown that within object theories the Codd's algebra can be defined. The results of the paper are also practically significant, because they offer a homogeneous and coherent method for manipulations with data bases as ontologies in logical knowledge bases.

**Keywords:** ontology; database; object theory; description logic, Libretto

Казаков Илья Анатольевич, Институт математики, экономики и информатики, Иркутский государственный университет, 664003, Иркутск, ул. К. Маркса, 1 тел.: (3952)242210 ([kazakov@baikal.ru](mailto:kazakov@baikal.ru))

Илья Kazakov, Irkutsk State University, 1, K. Marks St., Irkutsk, 664003  
Phone: (3952)242210 ([kazakov@baikal.ru](mailto:kazakov@baikal.ru))