

# ИНТЕЛЛЕКТУАЛИЗАЦИЯ АНАЛИЗА СЛОЖНЫХ ЗАПРОСОВ В РЕЛЯЦИОННЫХ СУБД

Зуенко А. А.<sup>1</sup>, Кулик Б. А.<sup>2</sup>, Фридман А. Я.<sup>1</sup>

<sup>1</sup> *Институт информатики и математического моделирования Кольского научного центра РАН  
г. Апатиты Мурманской области, 184209, Россия  
тел.: 81555-79782, e-mail: fridman@iimm.kolasc.net.ru*

<sup>2</sup> *Институт проблем машиноведения Российской академии наук (РАН)  
г. Санкт-Петербург, 199178, Россия  
тел.: 812- 5173498, e-mail: ba-kulik@yandex.ru*

*Аннотация* — Исследованы возможности применения предложенной авторами общей теории многоместных отношений для интеллектуализации процедур анализа сложных запросов к реляционным базам данных, представимых логическими формулами над элементарными одно- и двуместными предикатами без кванторов.

## I. Проблема учета семантики данных

Большинство фактографических систем, в частности, систем моделирования, используют для хранения и обработки данных реляционные СУБД (РСУБД), а также их расширения. Сильная сторона РСУБД заключается в том, что в них встроен язык запросов SQL, предоставляющий необходимые средства для выполнения незапланированных запросов. К недостаткам РСУБД относят слабое развитие в них средств семантического анализа данных. Это затрудняет использование РСУБД в предметных областях, имеющих сложно структурированные данные.[1]. Для представления семантики данных разработаны различные "расширенные" модели, в частности ER и RM/T [1 - 3], но и они не обладают развитыми средствами для задания ограничений, отличных от правил целостности.

Большинство операций реляционной алгебры заимствовано из алгебры множеств. Однако в рамках реляционной алгебры не определены операции алгебры множеств над отношениями, заданными в различных схемах, и операция взятия дополнения. Другими словами, реляционная алгебра не может рассматриваться как полноценный аналог алгебры множеств для случая, когда в качестве множеств берутся отношения (множества элементарных кортежей).

В то же время, известно, что не только данные, но и многие структуры знаний, используемые в формальных системах, представимы в виде отношений (графы, семантические сети, предикаты и т.п.). Следовательно, в качестве базовой структуры для унификации обработки данных и знаний целесообразно выбрать отношение, причем требуется принципиально новый, по сравнению с реляционной алгеброй, математический аппарат, обеспечивающий изоморфизм с алгеброй множеств. Такой математический аппарат – алгебра кортежей (АК) – описан в [4]. АК реализует общую теорию многоместных отношений, ориентированную на решение задач логического анализа (проверка правильности следствия, анализ гипотез и т.д.). Здесь мы остановимся на методах интеллектуализации реляционных СУБД методами АК. Цель работы – наделить СУБД дополнительными средствами логического анализа информации.

## II. Контекстный анализ ограничений

При моделировании сложных объектов модель предметной области (МПО) поэтапно уточняется и перестраивается, то есть она открыта для оператив-

ных модификаций структуры. Ввиду открытости МПО целесообразно автоматизировать ее контроль в статике и динамике, чтобы облегчить работу конечного пользователя и снизить негативную роль человеческого фактора. Для этого необходимо автоматически контролировать корректность процесса моделирования на всех этапах: от построения МПО до анализа результатов моделирования. Однако в большинстве систем моделирования контроль корректности сведен к жесткому программированию ограничений либо полностью возлагается на пользователя.

Мы предлагаем анализировать структурные особенности МПО посредством декларативного представления ограничений и их обработки методами искусственного интеллекта [5, 6]. В рамках подхода корректность модели контролируется на уровне семантического интерфейса реляционной БД системы моделирования путем оперативной подгрузки актуальных контекстных ограничений.

## III. Алгебра условных кортежей

АК ориентирована на анализ логических формул над элементарными одноместными предикатами, а большинство контекстных ограничений и запросов можно представить только в виде логических формул, содержащих элементарные двуместные предикаты. Как расширение АК на задачу логического анализа таких формул, авторами разработана алгебра условных кортежей (АУК) [5].

Пусть  $A = \{A_1, A_2, \dots, A_n\}$  – простые атрибуты, чьими доменами являются обычные множества, а  $C = \{A_1, A_2, \dots, A_n, A_1 \times A_2, A_1 \times A_3, \dots\}$  – множество, состоящее из простых атрибутов и всех их попарных декартовых произведений (сложных атрибутов).

АУК задается следующим образом:

$\Lambda = \langle \{R_{K_i}[Q_i]\}, \cup, \cap, +Atr, -Atr, Pr, \leftrightarrow Atr, \$ \rangle$ , где:  $R_{K_i}[Q_i]$  – некоторый АУК-объект (условный C-кортеж или условная C-система);  $Q_i$  – схема отношения, представляющая собой некоторое множество простых и сложных атрибутов, принадлежащих  $C$ ;  $K_i$  – множество сложных атрибутов, входящих в схему  $Q_i$ ;  $\$$  – операция наполнения АУК-объектов, остальные операции (+Atr – добавление атрибутов, -Atr – элиминация атрибутов, Pr – взятие проекции,  $\leftrightarrow Atr$  – перестановка атрибутов) аналогичны операциям АК.

Условным C-кортежем называется заданный в определенной схеме отношения кортеж множеств (компонент), представляющих собой либо подмножество одного из доменов простых атрибутов, либо совокупность бинарных отношений, заданных на соответствующем сложном атрибуте. В качестве основных бинарных отношений (двуместных предикатов) выступают: «<» (меньше), «>» (больше), «=» (равно), «≠» (не равно). Компоненты простых атрибутов формируются из открытых интервалов вида  $(a, b)$

и дискретных значений. В описании компонент сложных атрибутов используются лишь имена бинарных отношений, а конкретные множества значений (элементарных кортежей) формируются в результате операции наполнения  $\$$ .

Условной  $S$ -системой называется множество однотипных условных  $S$ -кортежей, записанных в виде матрицы, ограниченной прямыми скобками.

Чтобы отличать АУК-объекты от структур АК, в их именах явно указывается множество сложных атрибутов. Например,  $R_K[M]$  означает, что  $K$  – множество сложных атрибутов. АУК-объекты – это расширение  $S$ -систем и  $S$ -кортежей, в частности,  $S$ -система  $R[M]$  есть условная  $S$ -система  $R_K[M]$ , где  $K = \emptyset$ .

Пересечение компонент простых атрибутов сводится к пересечению между собой интервалов и дискретных значений. Пересечение компонент сложных атрибутов регламентируется специальной таблицей, исходя из свойств отношений «<», «>», «=», «≠».

Для реализации операции  $\$$  в АУК разработаны операторы взятия диагонали ( $\Delta$ ), взятия элементов, лежащих над ( $\bar{\Delta}$ ) и под диагональю ( $\underline{\Delta}$ ). Оператор  $\Delta([A, B])$  соответствует отношению «равно», т.е. находит диагональ множества  $A \cap B$ . Операторы  $\bar{\Delta}$  и  $\underline{\Delta}$  (сопоставляются отношениям «меньше» и «больше») определяют кортежи, которые принадлежат прямоугольнику  $A \times B$  и лежат соответственно над и под его диагональю. Отношение «не равно» можно выразить как дополнение отношения «равно» либо как объединение отношений «меньше» и «больше».

АУК позволяет автоматизировать анализ составляющих логической формулы, содержащей элементарные двуместные предикаты, и выявлять ее структурную противоречивость [5]. Если логическая формула соответствует некоторому SQL-запросу, то отрицательный результат ее проверки означает, что запрос всегда возвращает пустой набор данных, а положительный итог свидетельствует о возможности получения релевантных данных, при условии их наличия в базе данных.

#### IV. Заключение

В настоящей работе развивается подход к унифицированной обработке данных и знаний на основе их представления в виде системы многоместных отношений (АК-объектов).

АК создает единую методологическую основу для обработки данных и знаний и может рассматриваться как расширение реляционной алгебры на задачу обработки знаний.

Представление реляционных отношений, запросов к ним и ограничений на структуру базы данных в виде АУК-объектов позволяет расширить возможности логического анализа данных по сравнению с возможностями, предоставляемыми аппаратом алгебры кортежей. В частности, обеспечивается предварительная (до стадии исполнения запроса в СУБД) проверка корректности структуры запросов, селекторы которых являются логическими формулами над одно- и двуместными предикатами (бинарными отношениями). Таким образом, АУК поддерживает современный контекстный подход к управлению данными в системах моделирования.

Работа выполнена при поддержке РФФИ (проект № 09-07-00066), Отделения нанотехнологий и информационных технологий РАН (проект 2.3 текущей

Программы фундаментальных научных исследований) и Президиума РАН (проект 4.3 "Интеллектуальные базы данных" Программы № 15).

#### V. Список литературы

- [1] Дейт К. Дж. Введение в системы баз данных. М.; СПб.: Издательский дом "Вильямс", 2000. 848 с.
- [2] Цаленко М. Ш. Моделирование семантики в базах данных. М.: Наука, 1989. 288 с.
- [3] Codd E. F. Extending the Database Relational Model to Capture More Meaning // ACM Transactions on Database Systems. 1979. Vol. 4. No. 4. P. 397—434.
- [4] Кулик Б. А., Зуенко А. А., Фридман А. Я. Алгебраический подход к интеллектуальной обработке данных и знаний. СПб.: Изд-во Политехн. ун-та, 2010. 235 с.
- [5] Зуенко А. А., Фридман А. Я. Контекстный подход в системах сопровождения открытых моделей предметной области // Искусственный интеллект и принятие решений. 2008. Вып.3. С. 41—51.
- [6] Kulik B., Fridman A., Zuenko A. Logical Analysis of Intelligence Systems by Algebraic Method // Proc. of Twentieth European Meeting on Cybernetics and Systems Research EMCSR 2010 (Vienna, April 6-9, 2010). Vienna, 2010. P. 198—203.

### INTELLIGENT ANALYSIS OF COMPLEX ENQUERIES TO RELATIONAL DBMS

Zuenko A. A.<sup>1</sup>, Kulik B. A.<sup>2</sup>, Fridman A. Ya.<sup>1</sup>

<sup>1</sup> Institute for Informatics of KSC RAS

Apatity Murmansk region, 184209, Russia

Ph.: 81555-79782, e-mail: fridman@iimm.kolasc.net.ru

<sup>2</sup> Institute of Problems in Machine Science of RAS

St. Petersburg, 199178, Russia

Ph.: 812- 5173498, e-mail: ba-kulik@yandex.ru

**Abstract** — The paper introduces abilities of a developed general theory of multiple relations in the intelligent analysis of complex relational enquiries, which can be expressed by logical formulas containing elementary unary and binary predicates without quantifiers.

#### I. Semantics in Databases

Databases of conventional modelling systems supporting open subject domain models lack means for flexible synthesis and analysis of problem-dependent limitations in order to make data processing more intelligent. We see a solution of this problem in a declarative expression of such limitations and their processing by methods of artificial intelligence. This approach is implemented by a semantic interface for relational databases of a modelling system.

#### II, III. Algebra of Conditional N-tuples

Within frames of our algebraic approach to unified logical analysis of data and knowledge called n-tuple algebra (NTA), we have developed its extension to the problem of intelligent analysis of complex relational enquiries. This can be done by our algebra of conditional n-tuples (ACT) whose support is a set of both *simple attributes* (ordinary sets) and all their pairwise Cartesian products called *complex attributes*. The ACT operations return diagonal, subdiagonal and supradiagonal elements of these products, thus checking basic binary relations «<» (less than), «>» (more than), «=» (equal), «≠» (not equal) between simple attributes.

#### IV. Conclusion

ACT provides the automatic analysis of complex relational enquiries, which can be expressed by logical formulas containing elementary unary and binary predicates without quantifiers. In particular, it allows for preliminary (before execution of a query by a DBMS) revealing structure inconsistencies in a set of given limitations, thus increasing efficiency of query implementation.