

В. Н. Белов, П. П. Макарычев

## АНАЛИЗ БАЗЫ ДАННЫХ С ПРИМЕНЕНИЕМ АЛГЕБРЫ КОРТЕЖЕЙ

*Аннотация.* Рассматривается применение алгебры кортежей для определения возможностей OLAP транзакционной базы данных на примере анализа результатов деятельности сотрудников вуза.

*Ключевые слова:* алгебра кортежей, оперативный анализ данных, базы данных, C-кортеж.

*Abstract.* The article describes an approach to determine capabilities of OLAP through the instrumentality of cortege algebra by example of university staff activity results database.

*Key words:* cortege algebra, OLAP, database, C-cortege.

### Введение

Любая организация стремится увеличить эффективность своей деятельности за счет внедрения современных информационных систем и технологий. Как правило, со временем база данных информационной системы организации накапливает существенный объем данных, которые могут быть проанализированы с целью получения новых сведений. Однако при проектировании транзакционных баз данных, как правило, такая возможность не учитывается, соответственно возникает вопрос о применимости и возможностях существующей транзакционной базы данных для анализа данных. Для ответа на данный вопрос используют прототипирование [1]. Как правило, для самых сложных и критичных функций разрабатывается программный код, чтобы определить возможность их реализации. Однако ответ на данный вопрос также может быть получен с помощью исследования схемы базы данных с применением алгебры кортежей (АК), которая дает возможность строить модели обработки данных, позволяя избежать трудозатрат на разработку программного кода.

### 1. Анализ связей базы данных с помощью АК

АК может использоваться для анализа базы данных различными способами. Часть анализа можно проводить, не углубляясь в структуру отношений, а рассматривая связи между ними. Для этой цели структуру базы данных удобно представить в виде системного графа. Граф, описывающий структуру базы данных результатов деятельности сотрудников вуза, представлен на рис. 1.

Вершины графа  $F$ ,  $K$ ,  $P$ ,  $L$ ,  $H$  соответствуют отношениям «факультет», «кафедра», «сотрудник», «учетные записи», «права доступа». В терминальных таблицах, соответствующих отношениям  $T_1$  (кадровый потенциал),  $T_2$  (учебная работа),  $T_3$  (научная работа), хранятся значения меток шкал для оценки деятельности преподавателей. Отношения  $D_1$ ,  $D_2$ ,  $D_3$  соответствуют результатам деятельности. Взаимосвязи между отношениями показаны дугами графа. Данный граф можно представить как  $C$ -систему, изоморфную матрице смежности этого графа [2]:

$$G[XY] = \begin{bmatrix} \{L\} & \{H\} \\ \{P\} & \{K\} \\ \{K\} & \{F\} \\ \{D_1\} & \{T_1, P\} \\ \{D_2\} & \{T_2, P\} \\ \{D_3\} & \{T_3, P\} \end{bmatrix}$$

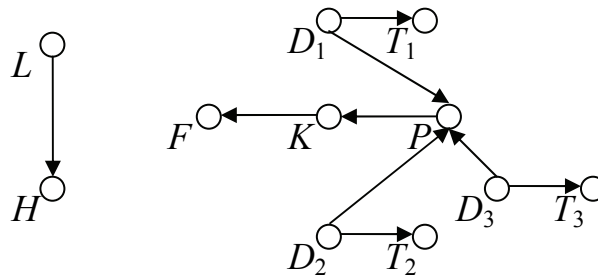


Рис. 1. Граф базы данных

На основе  $C$ -систем вычислим транзитивное замыкание графа:

$$G^+ = G \cup G^2 \cup G^3 = \begin{bmatrix} \{L\} & \{H\} \\ \{P\} & \{K\} \\ \{K\} & \{F\} \\ \{D_1\} & \{T_1, P\} \\ \{D_2\} & \{T_2, P\} \\ \{D_3\} & \{T_3, P\} \end{bmatrix} \cup \begin{bmatrix} \{P\} & \{F\} \\ \{D_1\} & \{K\} \\ \{D_{21}\} & \{K\} \\ \{D_3\} & \{K\} \end{bmatrix} \cup \begin{bmatrix} \{D_1\} & \{F\} \\ \{D_1\} & \{F\} \\ \{D_1\} & \{F\} \end{bmatrix} =$$

$$= \begin{bmatrix} \{L\} & \{H\} \\ \{P\} & \{K, F\} \\ \{K\} & \{F\} \\ \{D_1\} & \{T_1, P, K, F\} \\ \{D_2\} & \{T_2, P, K, F\} \\ \{D_3\} & \{T_3, P, K, F\} \end{bmatrix}$$

Транзитивное замыкание графа  $G^+$  представлено на рис. 2.

Каждый  $C$ -кортеж в результирующей  $C$ -системе показывает возможности разбиения объектов реального мира, описываемых отношением из первого столбца  $C$ -системы, на группы в соответствии с признаками, задаваемыми элементами из второго столбца, а каждый элементарный кортеж соответствует такому варианту разбиения. Таким образом, элементарные кортежи соответствуют простейшему варианту разбиения, а  $C$ -кортежи соответствуют более сложному разбиению. Количество элементарных кортежей  $n$ , входящих

в  $C$ -кортеж, соответствует количеству возможных простейших разбиений и определяет количество возможных более сложных разбиений, равное  $n!$ . Используя терминологию многомерной модели данных, можно сказать, что максимальная «степень» композиции графа с самим собой соответствует максимальному уровню измерения, задаваемого отношением во втором столбце, для фактов, задаваемых отношением в первом столбце [3].

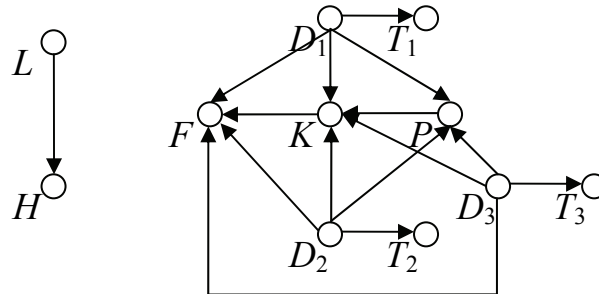


Рис. 2. Транзитивное замыкание графа базы данных

Данные, полученные с помощью АК, могут быть интерпретированы в терминах матроидов.  $C$ -системы  $G[XY]$  и  $G^+[XY]$  позволяют задать матроид графа  $G^+$ . Множество элементарных кортежей  $C$ -системы  $G^+[XY]$  задает множество дуг графа  $G^+$ , объединение доменов  $X \cup Y$  задает множество вершин графа. Первый элемент каждого элементарного кортежа  $C$ -системы  $G^+[XY]$  соответствует началу дуги, а второй элемент элементарного кортежа соответствует концу дуги. Таким образом, можно задать матрицу инцидентности для графа, что позволяет задать соответствующий графовый матроид [4]. Независимые множества матроида будут являться подмножествами множества элементарных кортежей  $G^+[XY]$ , соответствующего носителю матроида [5]. Множество элементарных кортежей  $C$ -системы  $G[XY]$  задает базис матроида, соответственно ранг матроида в данном случае равен количеству элементарных кортежей  $C$ -системы  $G[XY]$ . Об исследовании базы данных с помощью графового матроида будет написано в следующей статье.

## 2. Анализ отношений с помощью АК

Каждое отношение может быть представлено в алгебре кортежей как  $C$ -система, и становится возможным использование соответствующих операций. Каждое отношение в исследуемой базе данных имеет атрибут, соответствующий суррогатному ключу, обозначаемый как  $S^j$ , где  $j$  – имя отношения. Наличие в схемах двух отношений атрибутов, приписанных к одному сорту, позволяет вычислять их композицию и соединение. Для каждого элементарного кортежа  $C$ -системы, соответствующей транзитивному замыканию  $G^+$ , возможно построение частного универсума, определяющего один из вариантов разбиения множества объектов на подмножества с целью проведения OLAP. Каждый такой частный универсум может быть получен с помощью операций соединения, перестановки атрибутов и элиминации атрибутов. Так, элемен-

тарному кортежу  $(K, F)$  соответствует отношение  $R^{fk} [S^k, N^k, S^f, N^f]$ , задающее разбиение кафедр по факультетам, где  $N^k$  – название кафедры,  $N^f$  – название факультета. В структурах АК данное отношение может быть получено следующим образом: для  $C$ -системы  $K [S^k S^f N^k]$  выполним операцию перестановки атрибута, чтобы преобразовать его в  $C$ -систему  $K [S^k N^k S^f]$ , затем выполним операцию соединения отношения  $F [S^f N^f]$  и отношения, задаваемого полученной  $C$ -системой:

$$R^{fk} [S^k, N^k, S^f, N^f] = F [S^f, N^f] \oplus K [S^k, N^k, S^f].$$

Частный универсум, задаваемый отношением  $R^{fk}$ , может быть описан как декартово произведение его атрибутов  $R^{fk} = S^k \times N^k \times S^f \times N^f$ , что позволяет рассматривать его в качестве гиперкуба с измерением  $S^f$  и тремя атрибутами информационного назначения. При добавлении в схему отношения  $K [S^k S^f N^k]$  дополнительного атрибута, который может быть использован в качестве меры, становится возможным OLAP данного гиперкуба для получения более сложной информации, чем вычисление количества кафедр на факультете. При этом алгоритм задания отношения  $R^{fk}$  останется тем же, поскольку в этом случае отношение  $K [S^k S^f N^k]$  может быть получено с помощью элиминации атрибута. Большая часть кортежей данного декартова произведения не будет иметь места в базе данных и может быть задана  $C$ -системой  $R^{fk}$ . В связи с большим объемом данных, содержащимся в такой системе, фактор наличия возможности использования сокращенного описания в используемой СУБД может быть критичным.

Покажем, как полученная структура АК может быть использована для описания процесса формирования списка кафедр, имеющих на факультете с названием «ФВТ». Формулировка в АК будет иметь вид

$$Q_1^{fk} [S^k, N^k, S^f, N^f] = [* * * \{\text{ФВТ}\}],$$

где фиктивные компоненты «\*», обозначают множества, равные доменам соответствующих атрибутов [2].

Результирующей структурой является  $C$ -система, в каждом  $C$ -кортеже которой имеется единственный элементарный кортеж, поскольку между атрибутами  $S^k$  и  $N^k$  имеется функциональная связь.

Таким образом, можно определить  $C$ -систему, описывающую разбиение кафедр по факультетам, через объединение входящих в нее  $C$ -систем:

$$R^{fk} = \bigcup_{j=1}^m Q_j^{fk} [S^k, N^k, S^f, N^k] = [* * \{j\} *],$$

где  $m$  – количество факультетов в вузе;  $j$  – номер факультета.

Аналогичным образом построим отношения для остальных элементарных кортежей  $C$ -системы, соответствующей транзитивному замыканию  $G^+$ . Элементарному кортежу  $(K, P)$  соответствует отношение  $R^{kp} [S^p, N^p, O^d, O^p, O^r, S^k, S^f, N^k]$ , задающее разбиение преподавателей по кафедрам, где  $N^p$  – имя сотрудника,  $O^d$  – ученая степень,  $O^p$  – должность,  $O^r$  – ученое звание.

Отношение  $R^{kp}$  получается как соединение отношений  $K$  и  $P$ :

$$R^{kp} [S^p, N^p, O^d, O^p, O^r, S^k, S^f, N^k] = K [S^k, S^f, N^k] \oplus P [S^p, N^p, O^d, O^p, O^r, S^k].$$

Часто возникает ситуация, при которой возможны различные способы оперативного анализа данных единственного отношения. Рассмотрим отношение  $P [S^p, N^p, O^d, O^p, O^r, S^k]$ . Атрибуты  $O^d, O^p, O^r$  определяют текстовое описание соответствующих характеристик. Так уменьшается количество обращений к базе данных и, соответственно, время обработки запроса, что имеет немаловажное значение в информационных системах, ориентированных на транзакционные базы данных.  $C$ -система, задающая отношение  $P$ , может быть разбита на составляющее ее  $C$ -системы при использовании  $O^d, O^p, O^r$  в качестве измерений гиперкуба, задаваемого отношением  $P [S^p, N^p, O^d, O^p, O^r, S^k]$ .

Допустим, необходимо получить список всех доцентов вуза. Формулировка в АК будет иметь следующий вид:

$$Q_1^p [S^p, N^p, O^d, O^p, O^r, S^k] = [* * * \{ \text{доцент} \} * *].$$

Тогда  $C$ -система, соответствующая отношению  $P [S^p, N^p, O^d, O^p, O^r, S^k]$ , может быть представлена как

$$P [S^p, N^p, O^d, O^p, O^r, S^k] = \bigcup_{j=1}^m \bigcup_{i=1}^n \bigcup_{l=1}^q Q_1^p [S^p, N^p, O^d, O^p, O^r, S^k] = [* * o_j^d o_i^p o_l^r *],$$

где  $o_j^d$  –  $j$ -й элемент домена атрибута  $O^d$ ;  $o_i^p$  –  $i$ -й элемент домена атрибута  $O^p$ ;  $o_l^r$  –  $l$ -й элемент домена атрибута  $O^r$ ;  $m$  – кардинальность множества  $O^d$ ;  $n$  – кардинальность множества  $O^p$ ;  $q$  – кардинальность множества  $O^r$ .

Отношения, соответствующие  $C$ -системам, определяющим разбиение результатов деятельности на типы, можно записать следующим образом:

$$R^{d,t_i} [S_i^d, S^p, V_i^d, S_i^t, N_i^t, E, U] = T_i [S_i^t, N_i^t, E, U] \oplus D_i [S_i^d, S^p, V_i^d, S_i^t],$$

где  $V_i^d$  – некоторое числовое значение, характеризующее результат деятельности сотрудников вуза, которое может быть агрегировано;  $N_i^t$  – название вида деятельности сотрудников вуза, к которому относится результат;  $E$  – название подтипа деятельности, к которому относится вид деятельности сотрудников вуза;  $U$  – тип пользователя, который является ответственным за учет вида деятельности сотрудников вуза,  $i \in \{1, 2, 3\}$ . В дальнейшем индекс  $i$  будет использоваться для записи отношений

Последовательное соединение нескольких отношений позволяет строить кросс-таблицы. Так, например, результаты деятельности сотрудников вуза удобно представить в виде таблицы, в которой названия столбцов соответствуют видам работ, выполняемых сотрудниками вуза, а названиям строк соответствуют имена сотрудников вуза. В этом случае результатам деятельности сотрудников вуза будут соответствовать ячейки таблицы. Соответствующее отношение может быть получено следующим образом:

$$\begin{aligned} R^{pd,t_i} [S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p, N^p, O^d, O^p, O^r, S^k] = \\ = P [S^p, N^p, O^d, O^p, O^r, S^k] \oplus R_i^{d,t_i} [S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p]. \end{aligned}$$

Произведем соединение полученного отношения с отношением  $K$  :

$$\begin{aligned} R^{kpd,t_i} [S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p, N^p, O^d, O^p, O^r, S^k, S^f, N^k] = \\ = K [S^k, S^f, N^k] \oplus R^{pd,t_i} [S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p, N^p, O^d, O^p, O^r, S^k]. \end{aligned}$$

С помощью полученного отношения можно построить кросс-таблицу, где строки, заголовки которых соответствуют названиям кафедр, будут состоять из множества подстрок, являющегося подмножеством множества строк полученной ранее таблицы:

$$R^{kpd,t_i} = \bigcup_{j=1}^n Q_1^{kpd,t_i} [* * * * * * * * * * * \{j\} * *],$$

где  $n$  – количество кафедр в вузе,  $j \in \{1, 2, \dots, n\}$ .

Для каждой такой строки можно вычислить агрегированное значение. Пусть агрегирование  $V_i^d$  производится суммированием, тогда для хранения агрегированных данных потребуется задать отношение  $R^a [A, S^k]$ , где  $A$  – агрегированные значения для кафедр. Агрегированные данные будут представлять собой  $C$ -систему, полученную объединением  $C$ -кортежей:



Операция консолидации возможна при наличии иерархии измерения и осуществляется комбинацией операций соединения отношений, элиминации атрибутов и вычисления агрегированных значений соответствующих фактов некоторым заданным способом. Так, в гиперкубе, задаваемом отношением  $R^{kpd;t_i}$ , имеется измерение, для которого уровню измерения «кафедра» соответствует более высокий уровень «факультет». При выполнении операции консолидации сначала необходимо провести соединение с отношением, соответствующим более высокому уровню иерархии:

$$R^{fkpd;t_i} \left[ S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f \right] = \\ = F \left[ S^f, N^f \right] \oplus R^{kpd;t_i} \left[ S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f \right].$$

Затем выполняется операция элиминации атрибутов, соответствующих более низкому уровню иерархии:

$$\exists s^k \exists n^k R^{fkpd;t_i} \left[ S_i^d, V_i^d, S_i^t, N_i^t, E, U, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f \right].$$

На последнем этапе вычисляются новые агрегированные значения.

Осуществление операции консолидации возможно с помощью других алгоритмов, модель выполнения которых также можно построить с помощью АК.

Операция детализации производится аналогично операции консолидации с учетом того, что более высокий уровень иерархии заменяется более низким уровнем.

Аналогичным образом выполнение операций OLAP может быть показано на других данных.

Гиперкуб, задаваемый отношением  $R^{kpd;t_i}$ , содержит данные обо всех результатах деятельности одного типа сотрудников вуза.

Поскольку частный универсум, задаваемый декартовым произведением  $S_i^d \times S_i^t \times V_i^d \times S^p \times N^p \times O^d \times O^p \times O^r \times S^k \times N^k \times S^f \times N^f \times N_i^t \times E \times U$ , может содержать значительно большее число элементов, чем  $C$ -система, соответствующая отношению  $R^{kpd;t_i}$ , и требует для своего задания осуществления нескольких операций соединения, то при реализации сложных запросов к гиперкубу, задаваемому данным отношением, логично разбивать их на простые запросы. Так, запрос, формирующий список сотрудников факультета ФВТ, прошедших научную стажировку, может быть сформулирован как

$$Q_1^{fkpd;t_i} \left[ S_{3i}^d, S_3^t, V_3^d, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f, N_3^t, E, U \right] = \left[ \begin{array}{cc} * & * \\ * & * \\ * & * \\ * & * \\ * & * \\ * & * \\ * & * \\ * & * \\ * & * \\ * & * \end{array} \right] \{ \text{ФВТ} \} \{ \text{научная\_стажировка} \} \left[ \begin{array}{cc} * & * \\ * & * \end{array} \right].$$

При этом реализация может иметь вид  $Q_1^{fkpd;t_3} \cap R^{fkpd;t_3}$ , однако это потребует выполнения четырех операций соединения отношений, данные которых не были отфильтрованы, либо наличия сохраненного отношения  $R^{fkpd;t_3}$ . В первом случае для обработки данных потребуется большое коли-



чество вычислительных операций и оперативной памяти, а во втором случае в базе данных будет храниться большой объем избыточных данных. Однако в первом случае количество требуемых вычислительных ресурсов можно сократить, осуществляя операции среза перед соединением:

- 1)  $Q_2^f [S^f, N^f] = [* \{ФВТ\}] \cap R^f [S^f, N^f];$
- 2)  $Q_3^{fk} [S^k, N^k, S^f, N^f] = Q_2^f [S^f, N^f] \oplus K [S^k, N^k, S^f];$
- 3)  $Q_3^{fkp} [S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f] =$   
 $= Q_3^{fk} [S^k, N^k, S^f, N^f] \oplus P [S^p, N^p, O^d, O^p, O^r, S^k];$
- 4)  $Q_2^{t3} [S_3^t, N_3^t, E, U] = [* * \text{научная\_стажировка} *] \cap$   
 $\cap T_3 [S_3^t, N_3^t, E, U];$
- 5)  $Q_2^{d3t3} [S_3^d, S^p, V_3^d, S_3^t, N_3^t, E, U] = Q_2^{t3} [S_3^t, N_3^t, E, U] \oplus$   
 $\oplus D_3 [S_3^d, S^p, V_3^d, S_3^t];$
- 6)  $Q_2^{fkpd3t3} [S_3^d, S_3^t, V_3^d, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f, N_3^t, E, U] =$   
 $= Q_2^{d3t3} [S_3^d, V_3^d, S_3^t, N_3^t, E, U, S^p] \oplus$   
 $\oplus Q_3^{fkp} [S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f].$

Все доступные для анализа данные можно объединить в один OLAP-куб, в котором определены измерения  $S_1^d, S_2^d, S_3^d, S_1^t, S_2^t, S_3^t, S^p, O^d, O^p, O^r, S^k, S^f, E, U$  и меры  $V_1^d, V_2^d, V_3^d$ :

$$R^{fkpdt} [S_1^d, S_2^d, S_3^d, S_1^t, S_2^t, S_3^t, V_1^d, V_2^d, V_3^d, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f, N_1^t, N_2^t, N_3^t, E, U] =$$

$$= \bigcup_{i=1}^3 R^{fkpd;t_i} [S_i^d, S_i^t, V_i^d, S^p, N^p, O^d, O^p, O^r, S^k, N^k, S^f, N^f, N_i^t, E, U].$$

Данный гиперкуб можно использовать для моделирования наиболее сложных операций обработки данных, однако при реализации запросов необходимо помнить о возможности оптимизации, представленной выше.

### 3. Анализ сортов атрибутов с помощью АК

Подходы, рассматривающие только связи между отношениями, и их атрибуты не дают возможности определить все доступные варианты анализа данных. Их можно дополнить с помощью анализа сортов атрибутов отношений. Рассмотрим права доступа к данным в отношениях базы данных.

Определим отношение, задающее права всех пользователей системы:

$$R^{lh} [S^h, S, U, B, S^l, N^l, A] = L [S^l, N^l, A] \oplus H [S^h, S, U, B, S^l],$$

где  $N^l$  – имя сотрудника;  $A$  – данные, необходимые для идентификации и аутентификации;  $S$  – значение, равное значению атрибута, соответствующего суррогатному ключу одного из отношений  $F, K, P$ ;  $B$  – значение, определяющее отношение, значение суррогатного ключа которого определяет значение  $S$ .

Атрибуты  $S^k, S^f, S^p, S$  приписаны к одному сорту:

$$S^k \cup S^f \cup S^p = S.$$

Соответственно возможно выполнение операций соединения отношений, содержащих атрибут  $S$ , и отношений, содержащих атрибуты  $S^k, S^f, S^p$ .

Таким образом, отношение, соответствующее  $C$ -системе данных определенного типа, доступных пользователю, может быть получено с помощью АК.

На первом этапе получим  $C$ -системы, соответствующие правам доступа пользователей к данным факультетов, кафедр и сотрудников:

$$1) Q_1^{lh} [S, S^h, U, B, S^l, N^l, A] = [* * * \{\text{факультет}\} * * *];$$

$$2) Q_2^{lh} [S, S^h, U, B, S^l, N^l, A] = [* * * \{\text{кафедра}\} * * *];$$

$$3) Q_3^{lh} [S, S^h, U, B, S^l, N^l, A] = [* * * \{\text{сотрудник}\} * * *].$$

Над полученными  $C$ -системами проведем операции добавления фиктивных атрибутов:

$$4) Q_1^{lh} [S, S_i^d, V_i^d, S_i^t, N_i^t, E, S^h, U, N^p, O^d, O^p, O^r, N^k, N^f, B, S^l, N^l, A];$$

$$5) Q_2^{lh} [S, S_i^d, V_i^d, S_i^t, N_i^t, E, S^h, U, N^p, O^d, O^p, O^r, N^k, N^f, B, S^l, N^l, A];$$

$$6) Q_3^{lh} [S, S_i^d, V_i^d, S_i^t, N_i^t, E, S^h, U, N^p, O^d, O^p, O^r, N^k, N^f, B, S^l, N^l, A].$$

Проведем операции перестановки, элиминации атрибутов и добавления фиктивных атрибутов над  $C$ -системой, задаваемой отношением  $R^{fkpd_i t_i}$ :

$$7) R_1^{fkpd_i t_i} [S^f, S_i^d, V_i^d, S_i^t, N_i^t, E, U, N^p, O^d, O^p, O^r, N^k, N^f, B, S^l, N^l, A];$$

- 8)  $R_1^{fkpd_{it_i}} [S^k, S_i^d, V_i^d, S_i^t, N_i^t, E, U, N^p, O^d, O^p, O^r, N^k, N^f, B, S^l, N^l, A]$ ;
- 9)  $R_1^{fkpd_{it_i}} [S^p, S_i^d, V_i^d, S_i^t, N_i^t, E, U, N^p, O^d, O^p, O^r, N^k, N^f, B, S^l, N^l, A]$ .

Вычислим  $C$ -системы, определяющие доступные данные в зависимости от значений множества, соответствующих атрибуту  $U$ , определяющему, от имени какого типа пользователя осуществляется доступ к данным:

- 10)  $Q_1^{lhfkpd_{it_i}} = Q_1^{lh} \cap R_1^{fkpd_{it_i}}$ ;
- 11)  $Q_2^{lhfkpd_{it_i}} = Q_2^{lh} \cap R_2^{fkpd_{it_i}}$ ;
- 12)  $Q_3^{lhfkpd_{it_i}} = Q_3^{lh} \cap R_3^{fkpd_{it_i}}$ ;
- 13)  $Q_4^{lhfkpd_{it_i}} = Q_1^{lhfkpd_{it_i}} \cup Q_2^{lhfkpd_{it_i}} \cup Q_3^{lhfkpd_{it_i}}$ .

Полученные  $C$ -системы для каждого  $i$  определяют OLAP-кубы, разделяющие доступ к данным в зависимости от пользователя системы.

### Заключение

Таким образом, с помощью АК можно рассмотреть все возможности OLAP для транзакционной базы данных и принять решение о разработке соответствующих программных средств. В случае, если исследованных возможностей OLAP недостаточно, имеет смысл принять решение о разработке специального хранилища данных и исследовать его структуру с помощью АК, чтобы убедиться в соответствии возможностей анализа потребностям организации. При заполнении разработанного хранилища данных возможности OLAP транзакционной базы могут быть использованы при организации ETL-процесса. Таким образом, определение возможностей OLAP с помощью АК на этапе проектирования позволяет принимать решение о разработке, минуя часто встречающийся и трудоемкий этап прототипирования.

### Список литературы

1. **Overmyer, S. P.** Revolutionary vs. Evolutionary Rapid Prototyping: Balancing Software Productivity and HCI Design Concerns / S. P. Overmyer // Proceedings of the Fourth International Conference on Human-Computer Interaction. – Elsevier Science, 1991. – P. 303–307.
2. **Кулик, Б. А.** Вероятностная логика на основе алгебры кортежей / Б. А. Кулик // Моделирование и анализ безопасности и риска в сложных системах – 2005 : труды Международной научной школы (28 июня – 1 июля 2005 г.). – СПб. : Изд-во СПбГУАП, 2005. – С. 406–412.
3. **Педерсен, Т. Б.** Технология многомерных баз данных / Т. Б. Педерсен, К. С. Йенсен // Открытые системы. – 2002. – № 1. – С. 45–50.
4. **Липский, В.** Комбинаторика для программистов : пер. с польск. / В. Липский. – М. : Мир, 1988. – 200 с.
5. **Харари, Ф.** Теория графов / Ф. Харари. – М. : Мир, 1973. – 300 с.
6. **Барсегян, А. А.** Методы и модели анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб. : БХВ-Петербург, 2004. – 336 с.

**Белов Вадим Николаевич**  
аспирант, Пензенский  
государственный университет

E-mail: oracool@gmail.com

**Belov Vadim Nikolaevich**  
Postgraduate student,  
Penza State University

**Макарычев Петр Петрович**  
доктор технических наук, профессор,  
заведующий кафедрой математического  
обеспечения и применения ЭВМ,  
Пензенский государственный  
университет

E-mail: makpp@yandex.ru

**Makarychev Petr Petrovich**  
Doctor of engineering sciences, professor,  
head of sub-department of computer  
application and software,  
Penza State University

---

УДК 004.652

**Белов, В. Н.**

**Анализ базы данных с применением алгебры кортежей / В. Н. Белов,  
П. П. Макарычев // Известия высших учебных заведений. Поволжский регион.  
Технические науки. – 2011. – № 3 (19). – С. 25–36.**