

Процесс OLAP-анализа представляется совокупностью операций с многомерными данными – детализации, консолидации (группировки), формирования среза и поворота. Операции консолидации определяют переход от детального представления данных к агрегированному, а в случае детализации осуществляется обратный переход. Формирование среза куба заключается в фиксации значения (значений) определенного измерения, при этом сокращается размерность куба. Срез представляет собой подкуб, в который входят все остальные измерения. Операция поворота заключается в изменении положения осей куба – измерений. В результате вращения меняется «точка зрения» на данные.

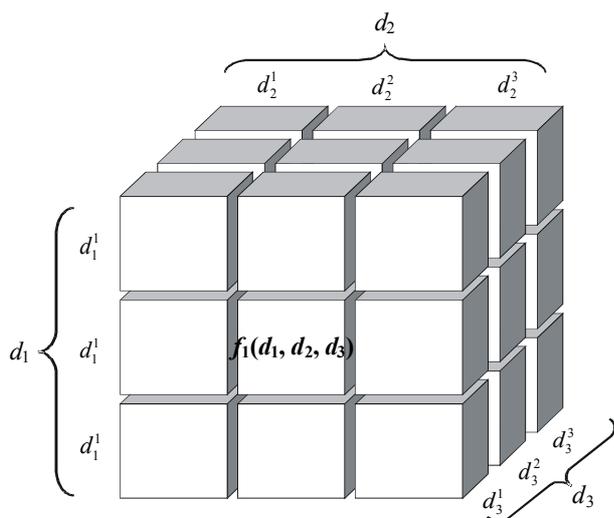


Рис. 1. OLAP-куб

Для конечного пользователя применение OLAP обеспечивает высокую гибкость получаемых решений за счет возможности изменения отображения результата. Аналитик получает не жестко регламентированный отчет, а инструментарий для творческого исследования задачи. Возможность свободной манипуляции данными упрощает получение необходимых наборов данных. Важным преимуществом OLAP является предоставление пользователю возможности оперировать знакомыми терминами из предметной области [11].

OLAP-куб можно рассматривать как абстрактное представление выбранного подмножества реляционной базы данных [12]. Процесс формирования OLAP-куба включает выбор необходимых таблиц, восстановление связей между ними, выбор полей данных и сопоставление их с терминами предметной области, что требует специальных знаний о предметной области и структуре исходных данных.

Построение OLAP-куба для каждой частной аналитической задачи ведет к формированию фрагментарной аналитической модели предметной области. Для повышения эффективности оперативной аналитической обработки данных необходима интегральная OLAP-модель, построенная на основе всех объектов анализа предметной области. Возможность манипулирования всеми объектами анализа предметной области одновременно сопряжена с необходимостью использования экспертных знаний об объектах анализа и о возможности их совместной аналитической обработки.

Построение интегральной OLAP-модели предметной области на основе формального концептуального анализа. С целью построения интегральной OLAP-модели предметной области на основе всех объектов анализа необходимо выделить группы объектов, имеющих общие структурные признаки, исходя из отношения сопоставимости и возможности их совместной аналитической обработки. Для этого целесообразно использовать методы бинарной объектно-признаковой кластеризации, в которых сходство объединяемых в один кластер объектов выражается через общие элементы описания всех объектов из данного кластера [13]. К таким методам относятся методы, основанные на формальных концептах и решетках формальных концептов [14].

Формальный концептуальный анализ (Formal Concept Analysis) впервые был предложен Р. Вилле (R. Wille) в 1981 г. [15] и активно развивается сегодня. Метод заключается в следующем.

Формальным контекстом называется тройка $K = (G, M, I)$, где G – множество объектов, M – множество атрибутов, $I \subseteq G \times M$ – отношение такое, что gIm , где $g \in G, m \in M$ означает, что объект g обладает атрибутом m . Формальный контекст может быть представлен в виде бинарной матрицы, строки которой помечены именами объектов, а столбцы – значениями атрибутов. Для произвольных $A \subseteq G$ и $B \subseteq M$ определяются A' и B' :

$$A' = \{m \in M \mid gIm \text{ для всех } g \in A\},$$

$$B' = \{g \in G \mid gIm \text{ для всех } m \in B\}.$$

Пара множеств (A, B) таких, что $A \subseteq G, B \subseteq M, A' = B$ и $B' = A$, называется формальным концептом контекста K . Множество объектов A представляет объем формального концепта, а множество атрибутов B – содержание формального концепта. Таким образом, формальный концепт – это множество объектов предметной области, каждый из которых обладает всеми атрибутами из некоторого подмножества атрибутов, присущих всем этим объектам.

Применение формального концептуального анализа к объектам оперативной аналитической обработки многомерных данных позволяет построить интегральную OLAP-модель на основе экспертных знаний об объектах анализа предметной области и возможности их совместной аналитической обработки. На основе интеграции технологии OLAP и формального концептуального анализа разработан метод построения концептуальной модели предметной области в виде формальной решетки многомерных кубов [16]. Контекстная диаграмма IDEF0 процесса построения интегральной OLAP-модели представлена на рис. 2.

Предлагаемый метод включает следующие основные этапы.

1. Формирование множества объектов анализа предметной области.
2. Построение формального контекста.
3. Определение формальных кубов-концептов.
4. Построение формальной концептуальной решетки OLAP-кубов.

На первом этапе путем интервьюирования конечного пользователя и изучения отчетных форм эксперт определяет аналитические задачи предметной области, фор-

мирует множество аналитических запросов. Сформулированные запросы позволяют определить множество терминов предметной области – объектов анализа, которые используются для построения OLAP-кубов концептуальной аналитической модели. Объекты концептуального анализа, в соответствии с моделью многомерного представления данных, делятся на множество измерений $D = \{d_1, d_2, \dots, d_n\}$ и множество показателей $F = \{f_1, f_2, \dots, f_m\}$. Термины, определяющие аспекты анализа предметной области, образуют множество измерений, а термины, представляющие количественные характеристики анализируемого процесса, образуют множество показателей. Затем на основе структуры хранилища данных выполняется сопоставление объектов анализа с полями таблиц, тем самым определяется физическая составляющая измерений и показателей. Объекты анализа могут быть связаны с полями таблиц хранилища данных напрямую или рассчитываться на основе нескольких полей по заданному алгоритму расчета.

На втором этапе построения интегральной OLAP-модели экспертом структурируется множество объектов анализа, сформированное на предыдущем этапе. На основе экспертных знаний об аналитических особенностях информационных объектов и функциональных зависимостей хранилища данных определяется сопоставимость показателей и измерений – возможность их совместной аналитической обработки.

Отношение сопоставимости между элементами множеств F и D обозначим R . $R \subseteq F \times D$, $(f_i, d_j) \in R$, если показатель f_i может быть проанализирован по измере-

нию d_j . Тройка (F, D, R) , в соответствии с теорией формального концептуального анализа, представляет собой формальный контекст K . Формальный контекст отражает знания эксперта об объектах анализа предметной области и о возможности их совместной аналитической обработки. Формальный контекст может быть представлен в виде бинарной матрицы, строки которой соответствуют показателям, а столбцы – измерениям.

На третьем этапе на основе сформированного формального контекста определяется множество кубов-концептов по признаку сопоставимости объектов анализа.

Теория формального концептуального анализа позволяет объединять характеристики анализируемого процесса по признаку общности аспектов анализа в кластеры – концепты. Множество показателей одинаковой размерности A , которые могут быть проанализированы по всем измерениям из B , образуют куб-концепт (A, B) . Множество показателей A представляет объем формального куба-концепта, а множество измерений B – содержание формального куба-концепта.

В соответствии с моделью многомерного представления данных, формальный куб-концепт – это аналитический многомерный куб, полный относительно добавления показателей той же размерности и состава измерений. Это означает, что невозможно включить в такой OLAP-куб дополнительный показатель без уменьшения числа измерений, т. е. в рамках построенного формального контекста не существует других показателей, сопоставимых с тем же набором измерений.

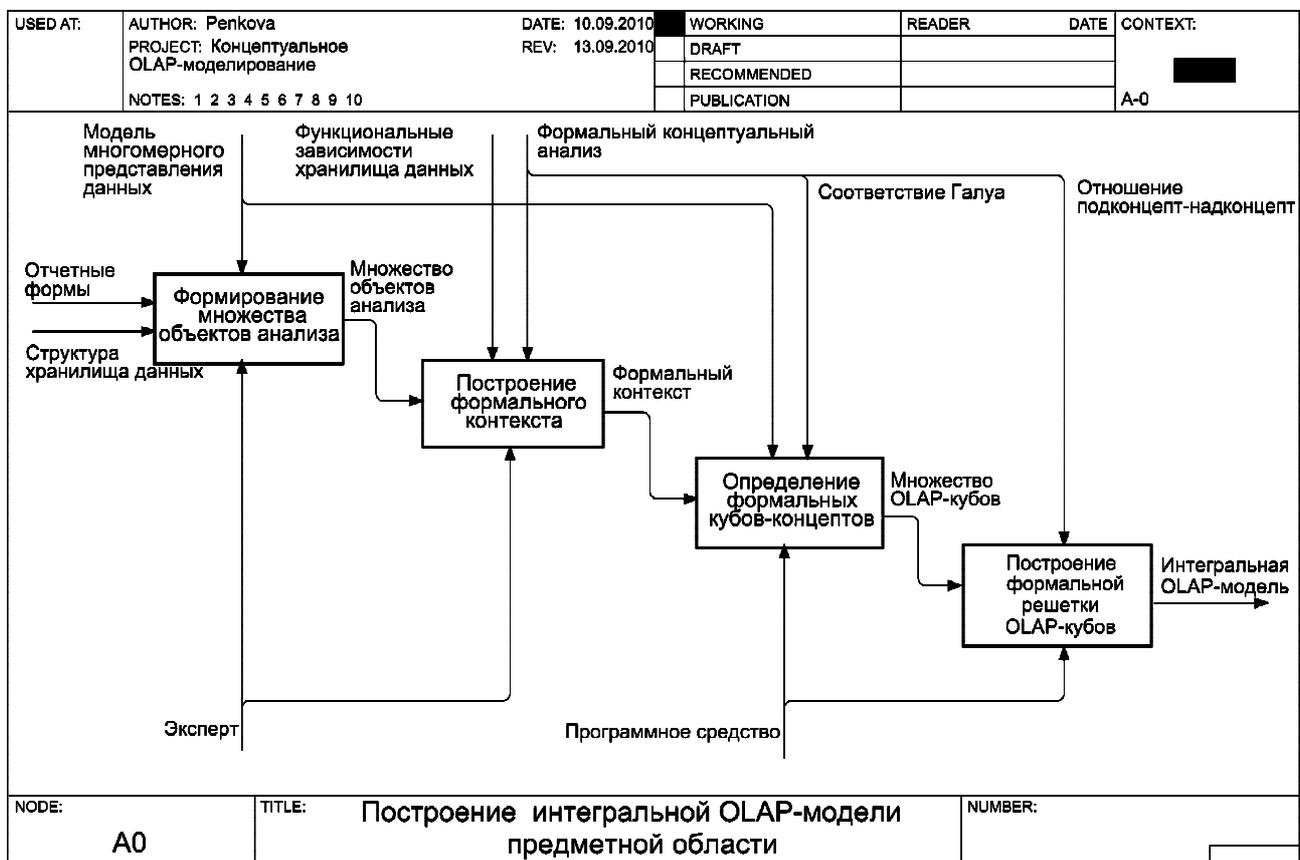


Рис. 2. Построение интегральной OLAP-модели предметной области

На заключительном этапе строится решетка кубов-концептов, которая позволяет оперировать всеми объектами анализа и охватывает максимально возможное число решаемых аналитических задач рассматриваемой предметной области.

Согласно методу формального концептуального анализа, множество всех концептов частично упорядочено отношением подконцепт–надконцепт: $(A_1, B_1) \leq (A_2, B_2)$ если $A_1 \subseteq A_2$ (что эквивалентно $B_2 \subseteq B_1$). В этом случае (A_1, B_1) называют подконцептом (A_2, B_2) , а (A_2, B_2) – надконцептом (A_1, B_1) . Упорядоченное множество всех концептов контекста образует полную решетку – решетку концептов [17].

Для концептуальной OLAP-модели отношение подконцепт–надконцепт определяется как подкуб–надкуб: множество показателей родительского куба включает множество показателей дочернего куба, а, в свою очередь, множество измерений дочернего куба включает множество измерений родительского куба. Решетка кубов-концептов представляет собой визуализацию интегральной OLAP-модели предметной области.

Построение формальной концептуальной решетки OLAP-кубов научной деятельности организации. Рассмотрим применение метода концептуального OLAP-моделирования для построения интегральной OLAP-модели научной деятельности организации в виде формальной концептуальной решетки многомерных кубов.

Научная деятельность – одна из основных форм функционирования научно-образовательных центров, научных и образовательных учреждений. Научная деятельность – это интеллектуальная деятельность, направленная на получение и применение новых знаний для решения технологических, инженерных, экономических, социальных, гуманитарных и иных проблем, обеспечения функционирования науки, техники и производства как единой системы.

Анализ эффективности научной деятельности организации связан с решением таких аналитических задач, как:

- оценка результативности научных исследований;

- оценка результативности аспирантуры и докторантуры;
 - оценка публикационной активности сотрудников;
 - мониторинг показателей интеграции науки и образования;
 - оценка кадрового обеспечения организации и др.
- Решение аналитических задач, в свою очередь, связано с выполнением следующих запросов:
- количество публикаций, подготовленных сотрудниками;
 - количество публикаций в российской и зарубежной печати за год;
 - количество патентов, полученных подразделениями организации за год;
 - количество учебных пособий, подготовленных сотрудниками организации;
 - количество российских и зарубежных грантов;
 - количество объектов интеллектуальной собственности;
 - количество сотрудников, ведущих преподавательскую деятельность;
 - количество сотрудников, осуществляющих научное руководство;
 - количество научных мероприятий, проведенных подразделениями организации и др.

Анализ запросов позволяет выделить множество терминов предметной области и на их основе определить множество объектов анализа научной деятельности. В том числе определяются:

- показатели: число публикаций; число патентов; число статей; число учебных пособий; число грантов; число проведенных конференций; количество сотрудников и т. д.
- измерения: год; подразделение; тип пособия; город; название журнала; тип публикации; тип патента; статус конференции; автор и т. д.

С учетом сопоставимости показателей и измерений (совместной аналитической обработки) строится формальный контекст. Фрагмент формального контекста научной деятельности организации представлен на рис. 3.

		d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
		Год	Подразделение	Тип пособия	Город	Название журнала	Тип публикации	Тип патента	Статус конференции	Автор
f_1	Число публикаций	×	×				×			×
f_2	Число проведенных конференций	×	×		×				×	
f_3	Число патентов	×	×					×		×
f_4	Число статей	×	×		×	×	×			×
f_5	Число учебных пособий	×	×	×	×		×			×

Рис. 3. Фрагмент формального контекста научной деятельности организации

A. V. Korobko, T. G. Penkova

METHOD OF CONCEPTUAL OLAP-MODELLING BASED ON FORMAL CONCEPTUAL ANALYSIS

A method of domain conceptual OLAP-modeling is suggested. The way of constructing the integral analytical model as a formal conceptual OLAP-cubes lattice is formally described. Advantages of on-line analytical processing based on conceptual OLAP-model are presented. Suggested method is implemented for integral analytical model constructing for scientific activities of the organization.

Keywords: conceptual OLAP-modeling, on-line analytical processing, formal conceptual analysis, decision support.

© Коробко А. В., Пенькова Т. Г., 2010

УДК 004.04:004.421

А. И. Ноженков

ПОСТРОЕНИЕ УПРАВЛЯЕМЫХ КОМПЛЕКСОВ OLAP-МОДЕЛЕЙ*

Представлены средства метаописания комплексов OLAP-моделей для решения сложных аналитических задач. Описаны алгоритмы управления расчетами. Полученные результаты применены в задачах территориально-планирования медицинской помощи.

Ключевые слова: OLAP-модель, комплекс OLAP-моделей, управление, многомерные данные, гибридная модель знаний, алгоритмы управления расчетами.

На сегодняшний день существует достаточно большое количество программных продуктов, реализующих функции OLAP-анализа. Такие крупные компании, как Hyperion Solutions Corp., IBM, Oracle, Microsoft, Sybase, Panorama Software, Cognos Inc. и другие ведут разработки в этой области, их решения охватывают практически все существующие задачи. Однако традиционные OLAP-решения малоэффективны в прикладных областях организационного управления, где необходим комплексный анализ данных, связанный с реализацией сложных аналитических алгоритмов. Методы расчета аналитических показателей для решения задач оперативного управления и планирования в здравоохранении, образовании, в сфере социальной защиты населения и множестве других прикладных областей представляют собой сложные многошаговые процессы анализа многомерных данных. Потребовался новый подход в OLAP-технологии, ориентированный на решение сложных аналитических задач, использующих связные многошаговые расчеты с множеством информационных объектов.

В статье представлены результаты работ по развитию функциональности OLAP-продуктов в рамках нового подхода, названного комплексным OLAP-моделированием [1; 2]. Описаны средства автоматизации разработки OLAP-приложений для сложных аналитических задач на основе построения управляемых комплексов OLAP-моделей.

Метаописание комплекса OLAP-моделей для сложных задач. Идея технологии OLAP хорошо описана [3–5]. В основе OLAP лежит многомерная логическая модель данных [6]. OLAP-модель (далее – модель) представляет собой описание процессов расчета некоторой аналитической задачи. Структурно аналитическая модель состоит из исходных данных (витрины данных), информационного куба и операций над ним, а также способов представления результатов вычисления.

Для решения сложных аналитических задач обычно строится не одна модель, а целый комплекс OLAP-моделей, каждая из которых решает отдельную подзадачу. Например, чтобы применить эту технологию к задаче планирования медицинской помощи, требуется построить модели анализа и прогноза демографической ситуации в регионе, анализа заболеваемости, анализа соответствия необходимых объемов медицинской помощи возможностям существующих медицинских учреждений и т. п. Каждая из перечисленных подзадач имеет самостоятельное значение, однако чтобы осуществить ежегодное планирование медицинской помощи в регионе, необходимо выполнить целенаправленный связный расчет с применением всего комплекса моделей. Для этого помимо средств сохранения результатов расчетов каждой модели необходимо обеспечить управление расчетами в зависимости от состояния задачи.

Построение управляемых комплексов OLAP-моделей для сложных аналитических задач предлагается осуше-

* Работа выполнена при финансовой поддержке гранта ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009–2013 гг. (ГК № 02.740.11.0621) и гранта ККФПН и НТД (доп. соглашение № 01/10 от 17 мая 2010 г.).