

Средства OLAP-моделирования и их применение в задачах здравоохранения

Ноженкова Л. Ф.

expert@icm.krasn.ru

Красноярск, Институт вычислительного моделирования СО РАН

Технология оперативной аналитической обработки многомерных данных OLAP (On-line Analytical Processing) считается одним из разделов интеллектуального анализа данных. Аналитические OLAP-модули все чаще появляются в составе отечественных и зарубежных продуктов и финансово-производственных приложений. Существование аналитической обработки сводится к автоматизированной поддержке формирования аналитических запросов, агрегированию данных, операциям над многомерным кубом данных с использованием плоских представлений — кросс-таблиц, кросс-диаграмм, картограмм. Наибольшее применение технология OLAP получила в бизнес-среде, где, как правило, решение конкретной аналитической задачи укладывается в рамки одного многомерного куба. При этом классические OLAP-решения мало пригодны к использованию в прикладных областях, где необходим комплексный анализ данных, связанный с реализацией сложных аналитических алгоритмов. Примерами прикладных областей, в которых указанные проблемы не позволяют эффективно применять традиционные средства OLAP-технологии, являются здравоохранение, образование, социальная защита населения и множество других. Методы расчета аналитических показателей и решения задач планирования в этих прикладных областях представляют собой сложные многошаговые процессы анализа многомерных данных. Возникает необходимость поэтапной обработки данных. Например, задачи планирования медицинской помощи представляют собой сложные многошаговые процессы с большим количеством расчетов, многообразием входной и выходной информации, сложными внутренними взаимосвязями.

Коллективом отдела прикладной информатики Института вычислительного моделирования СО РАН предложен новый подход к решению разнообразных задач с применением OLAP-технологии, основанный на построении комплексов так называемых «OLAP-моделей» [1]. OLAP-модель строится пользователем и несет в себе описательную информацию о решении некоторой аналитической задачи. Структурно OLAP-модель состоит из информации, описывающей исходные данные и их взаимосвязи, измерения и показатели информационного куба, операции над кубом, способы представления результатов вычисления и способы сохранения результатов для последующего использования. Введение в модель операций сохранения результатов расчета в источник (хранилище) данных

позволило реализовать поэтапный анализ данных путем создания комплексов OLAP-моделей.

Комплекс представляет собой совокупность OLAP-моделей, связанных по данным. В рамках одного расчета модели образуют последовательно выполняемую цепочку операций, при этом данные, рассчитанные одной моделью, в дальнейшем используются другими моделями.

Ядром системы оперативной аналитической обработки данных является OLAP-машина, которая представляет собой механизм выполнения запросов пользователя на выборку многомерной информации и изменения ее представления. От архитектуры OLAP-машины зависит и уровень решаемых аналитическим инструментом задач, и степень свободы пользователя при решении этих задач. Предложена и реализована оригинальная архитектура OLAP-машины, отличительными особенностями которой являются выполнение нерегламентированных запросов пользователя, использование встроенного языка программирования для расчета значений аналитических объектов и возможность использования составных иерархий в качестве измерений. Составная иерархия позволяет упорядочить информационные объекты одновременно по нескольким измерениям. Использование иерархий в качестве измерений многомерного куба потребовало применения специальных структур данных и алгоритмов, направленных в первую очередь на уменьшение временных затрат. Разработаны средства автоматизации создания специализированных OLAP-приложений: инструментальное ядро в виде набора компонент, связанных с OLAP-машиной, среда проектирования экранных форм пользовательского интерфейса, мастер быстрого создания приложений.

Расширен функциональный состав хранилищ данных [2]. Введены новые конструктивные элементы, выполняющие функции поддержки связанных многошаговых аналитических расчетов: OLAP-модель, сложное иерархическое измерение, таблица расчетных значений (агрегатов) и группа отчетных форм. Применение оригинальных технологических компонентов позволило выполнять сложные многошаговые аналитические расчеты. На первом шаге на вход аналитического инструмента поступают исходные обрабатываемые данные. Далее идет последовательное выполнение шагов расчета, параметры которых описаны в репозитории хранилища в виде OLAP-моделей. Взаимодействие моделей между собой происходит путем передачи через хранилище информации в виде таблиц агрегатов и данных репозитория. Выполнение многошагового расчета сопровождается так называемым интерактивным аналитическим экспериментом — возможно вмешательство пользователя в выполнение расчета для модификации параметров и настройки модели. Про-

цесс формирования каждой из аналитических моделей сопровождается взаимодействием пользователя со средствами управления хранилищем. На любом из этапов построения модели возможен возврат к более ранним этапам.

Разработаны средства создания OLAP-приложений с адаптированным для специалистов предметной области интерфейсом. Инструментарий разработки адаптированных интерфейсов позволяет создавать ориентированные на конкретную задачу OLAP-приложения на основе инструментального ядра, полностью ограждая пользователя от сложной внутренней организации системы и сохраняя при этом весь функционал. Важным требованием к разрабатываемым приложениям является отражение специфики и традиций конкретной предметной области, в том числе, создание и применение словаря терминов соответствующей прикладной области.

Разработанные средства организации хранения и оперативной аналитической обработки данных послужили основой для создания и использования централизованного хранилища медицинских данных регионального уровня [3, 4, 5]. Фактологические данные в хранилище стекаются из множества информационных систем, работающих в учреждениях здравоохранения и обязательного медицинского страхования. Состав информации в хранилище данных направлен на анализ эффективности деятельности медицинских учреждений в системе здравоохранения региона и оказывает существенную помощь в решении задач оперативного управления и планирования. Медицинская информация разнородна и включает статистические (обобщенные), персонифицированные, пространственно-распределенные, и другие типы данных. Для увеличения наглядности представления результатов оперативного аналитического моделирования разработаны средства отображения результатов OLAP-анализа на электронной карте. Полученные результаты применены для информационно-аналитической поддержки задач национального проекта «Здоровье» по реструктуризации сети медицинских учреждений.

Работа выполнена при поддержке гранта Президента для ведущих научных школ № НШ-3428.2006.9, гранта РФФИ № 05-07-90244 и гранта по проекту № 7 Программы фундаментальных исследований Президиума РАН № 14.

Литература

- [1] Дудина Ю. В., Ишенин П. П., Ноженкова Л. Ф. Технология реализации аналитических моделей средствами системы «Аналитик» для решения задач планирования / Труды Всероссийской конференции «Информационно-аналитические системы и технологии в здравоохранении и ОМС». — Красноярск: КМИАЦ, 2002. — С. 246–254.

-
- [2] *Жучков Д. В.* Автоматизация обработки больших массивов данных // «Открытое образование». Приложение — Красноярск: ООО «Экспресс-Офсет», 2006. — С. 56–62.
- [3] *Евсюков А. А., Ноженкова Л. Ф.* Оперативное геомоделирование сети медицинских учреждений // Вестник КрасГАУ. — 2006. — № 13. — С. 114–118.
- [4] *Исаева О. С.* Построение информационных моделей для OLAP-анализа медико-демографических данных // Журнал «Открытое образование». Приложение. — Красноярск, 2006. — С. 65–73.
- [5] *Ноженков А. И., Коробко А. В., Никитина М. И.* Информационно-аналитическая поддержка формирования территориальной программы бесплатной медицинской помощи // Вестник КрасГАУ, 2006. — № 13. — С. 108–113.