

Белорусский государственный университет информатики и радиоэлектроники, кафедра ИТАС
ул. П. Бровки, г. Минск, Беларусь
телефон: + (375) 296 999 086; e-mail: eugenoleg@mail.ru
web: www.bsuir.by

В статье рассмотрены аспекты реализации и принципы построения информационных аналитических систем на основе многомерной модели данных. Так же описан программный продукт «АИС кафедры ВУЗа», представляющий типовое решение по созданию и анализу многомерных баз данных и позволяющий анализировать профессиональную деятельность сотрудников кафедры ВУЗа и деятельность кафедры в целом.

Ключевые слова – аналитическая информационная система, гиперпрямоугольник данных, многомерная база данных, многомерная модель данных.

1 МНОГОМЕРНОЕ ПРЕДСТАВЛЕНИЕ ДАННЫХ

В процессе принятия решений пользователь генерирует некоторые гипотезы. Для превращения этих гипотез в законченные решения они должны быть проверены. Проверка гипотез осуществляется на основании информации об анализируемой предметной области. Как правило, наиболее удобным способом представления такой информации для человека является зависимость между некоторыми параметрами.

В процессе анализа данных, поиска решений часто возникает необходимость в построении зависимостей между различными параметрами. Кроме того, число таких параметров может варьироваться в широких пределах. Традиционные средства анализа, оперирующие данными, которые представлены в виде таблиц реляционной базы данных (БД), не могут в полной мере удовлетворять таким требованиям. В 1993 г. Э.Ф.Кодд рассмотрел ее недостатки, указав в первую очередь на невозможность «объединять, просматривать и анализировать данные с точки зрения множественности измерений, т.е. самым понятным для аналитиков способом». Поэтому не удивительно то внимание, которое сегодня уделяется средствам реализации и концепциям построения аналитических информационных систем (АИС), ориентированных на аналитическую обработку данных, основанных на многомерном подходе.

Многомерное представление данных и OLAP (Online Analytical Processing) уже стали сегодня одними из наи-

более широко распространенных концепций построения аналитических систем [1].

Обратимся к многомерному представлению данных. Основными понятиями многомерной модели данных являются гиперпрямоугольник данных, оси (измерения), метки, ячейка и вектор мер (рисунк 1).

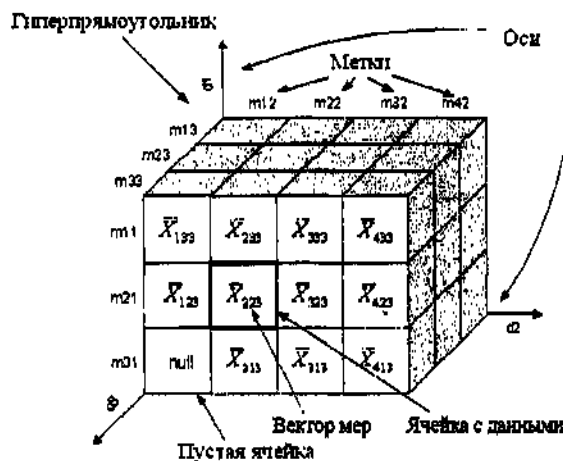


Рис.1. Многомерная модель данных

Гиперпрямоугольник данных содержит два или более измерений и представляет собой упорядоченный набор ячеек (упорядоченный многомерный массив). Каждая ячейка определяется одним и только одним набором значений измерений – меток. Ячейка может содержать данные, количественно характеризующие процесс – меры или быть пустой.

Под измерением будем понимать множество меток, образующих одну из граней гиперпрямоугольника. Измерениями служат основные атрибуты анализируемого процесса. Как правило, в качестве одного из измерений используется время. Примером временного измерения является список дней, месяцев, кварталов, лет.

Для получения доступа к данным пользователю необходимо указать одну или несколько ячеек путем выбора значений измерений, которым соответствуют необходимые ячейки. Процесс выбора значений измерений называется фиксацией меток, а множества выбранных значений измерений – множеством фиксированных меток. Подмножество гиперпрямоугольника, получившееся в результате фиксации меток одного или более измерений,

называется срезом. Операция построения среза проводится с целью получения требуемого подмножества ячеек и отсечения "ненужных" значений путем последовательной фиксации меток. Срез, как правило, представляет собой двумерный массив (таблицу) [2].

Очевидно, что физически данные, представленные в многомерном виде, хранятся в "плоских" файлах. При этом гиперпрямоугольник представляется в виде одной плоской таблицы, в которую построчно вписываются все комбинации членов всех измерений с соответствующими им значениями мер.

Итак, суть OLAP заключается в том, что исходная для анализа информация представляется в виде многомерного гиперпрямоугольника, и обеспечивается возможность произвольно манипулировать ею и получать нужные информационные разрезы – отчеты. При этом конечный пользователь видит гиперпрямоугольник как многомерную динамическую таблицу, которая автоматически суммирует данные в различных разрезах (измерениях), и позволяет интерактивно управлять вычислениями и формой отчета.

2 ПРОГРАММНЫЙ ПРОДУКТ «АИС КАФЕДРЫ ВУЗА»

Представляемый программный продукт, называемый «АИС кафедры ВУЗа», разработан в системе MatLab с применением принципов проектирования многомерных баз данных. В нем реализовано типовое решение по созданию и редактированию подсистем, состоящих из многомерных баз данных с заданной размерностью и с любым количеством параметров на пересечении измерений, а также реализован типовой модуль формирования текстовых и графических отчетов для созданных подсистем. Графические отчеты представляют собой графики и диаграммы, при построении которых существует возможность выбирать анализируемые атрибуты (измерения), которые будут помещены в легенду графического окна.

В программном средстве созданы следующие подсистемы: сотрудники кафедры; публикации; студенты; магистратура; аспирантура; успеваемость; дипломное проектирование; нагрузка сотрудников.

В базах данных подсистем «Успеваемость», «Публикации» и «Студенты» многомерность реализована 3-мя измерениями: года, курсы и семестры; года, сотрудники кафедры и виды публикаций; года, сотрудники кафедры и мероприятия, в которых участвовали студенты (таких как конференции, республиканские конкурсы).

В базах данных «Магистратура» и «Аспирантура» многомерность реализована 4-мя измерениями: года, сотрудники, формы обучения (очное или заочное) и формы оплаты (платное или бюджетное).

В базе данных «Дипломное проектирование» две оси (года и сотрудники кафедры).

В базе данных «Нагрузка сотрудников» шесть измерений: года, сотрудники кафедры, семестры, курсы, дисциплины и виды работ (такие как лабораторные, практические, лекции, курсовое проектирование, консультации, зачеты, экзамены, дипломное проектирование, руководство и занятия с аспирантами и магистрантами и др.).

Информация по запросам представляется в виде развернутых графических и текстовых отчетов.

На рисунке 2 изображен график распределения видов работ по годам, сотрудникам, курсам, семестрам и дисциплинам

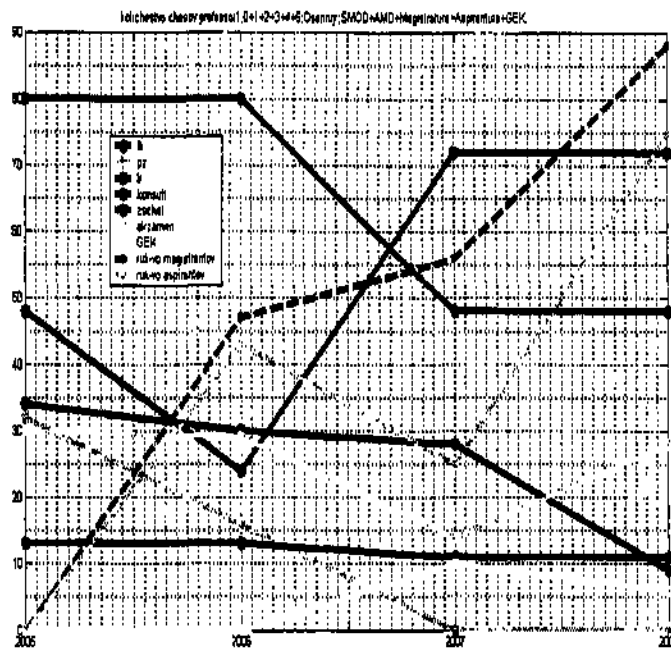


Рис. 2. Графический отчет подсистемы «Нагрузка сотрудников».

Данная система может успешно использоваться на кафедрах высших учебных заведений для учета личной информации о сотрудниках, для учета и детального анализа их профессиональной деятельности и деятельности кафедры в целом, и принятия на основании данного анализа управленческих решений.

ЛИТЕРАТУРА

- [1] Методы представления информации в разреженных гиперкубах данных [Электронный ресурс] – Режим доступа: <http://www.olap.ru/home.asp?artId=77>, свободный. – Загл. с экрана.
- [2] Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. – СПб.: БХВ-Петербург, 2004. – 336 с. ил.