

Использование OLAP-систем для аналитических исследований в системах дистанционного обучения

Житкевич А.П., Пашенко А.В.

Белорусский национальный технический университет

В настоящее время появляется все больше и больше различных обучающих систем, как для дистанционного обучения, так и для использования в учебных аудиториях. Функциональность этих систем может быть представлена такими возможностями как:

1. Удобное представление учебного материала с разбиением на темы, подтемы и другие иерархические элементы, а также возможность доступа к информации смежных и общих дисциплин.

2. Подсистема проверки знаний, включающая различные варианты тестирования, например, оценка знаний учащегося по результатам семестра либо, тест на степень освоенности материала одной темы.

3. Анализ результатов проверки знаний педагогом. Такой анализ обычно производится в разрезе групп, курсов, а также дисциплин.

Существующие обучающие системы не предусматривают в себе возможность контроля со стороны преподавателя за качеством учебного материала и подсистемы тестирования. Наиболее уязвимым элементом является именно тестирующая оболочка. Главным здесь является правильный выбор и формулировка заданий (как вопросов, так и ответов), а также оценка их выполнимости и репрезентативности. Для получения корректных результатов анализа исследование должно производиться на достаточно большом количестве измерений, желательно за продолжительный период времени.

В общем случае, большой объем информации, необходимой для хранения данных, их анализа с последующим принятием решения, ведет к необходимости применения специальных средств, называемых системами поддержки принятия решений, которые призваны облегчить работу людей, выполняющих анализ. В нашем случае аналитиком будет преподаватель каждой конкретной дисциплины.

На рисунке 1 изображена обобщенная структура системы поддержки принятия решений.

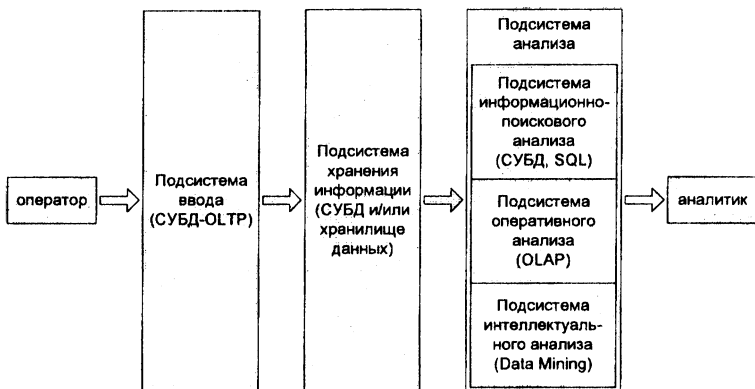


Рис. 1. Структура системы поддержки принятия решений.

Для реализации такой системы нужна определенная ретроспективная информационная база, в которой накапливаются сведения об уже выполненных учащимися заданиях. В качестве такой базы обычно используется реляционная база данных (или OLTP – система оперативной транзакционной обработки). Недостатком системы типа OLTP является отсутствие поддержки инструментов для аналитического представления данных и, как следствие, невозможность их дальнейшего анализа. Задачами анализа в разрабатываемой системе могут быть, например, определение зависимости количества правильных ответов от количества обращений к теоретическому материалу или от количества обращений к знаниям смежных дисциплин, определение решаемости конкретного задания, поиск некорректных формулировок как в содержимом тестов, так и в учебном материале и т.д.

Для проведения таких аналитических исследований исходная база данных должна быть сначала преобразована в хранилище данных (ХД). В основе концепции ХД лежит идея разделения данных, используемых для оперативной обработки и для решения задач анализа. Это позволяет применять структуры данных, которые удовлетворяют требованиям их хранения с

учетом использования в OLTP-системах и системе анализа. Такое разделение позволяет оптимизировать как структуры данных оперативного хранения для выполнения операций ввода, модификации, удаления и поиска, так и структуры данных, используемые для анализа.

ХД – это предметно-ориентированный, интегрированный, неизменяемый, поддерживающий хронологию набор данных, организованный для целей поддержки принятия решений.

В разрабатываемой системе данные, необходимые для определенных задач анализа и хранящиеся в OLTP-системе, сводятся в единую таблицу, представляющую хранилище данных. На рисунке 2 изображена структура базы данных для хранения преобразованных данных, которые имеют подходящую для анализа форму.

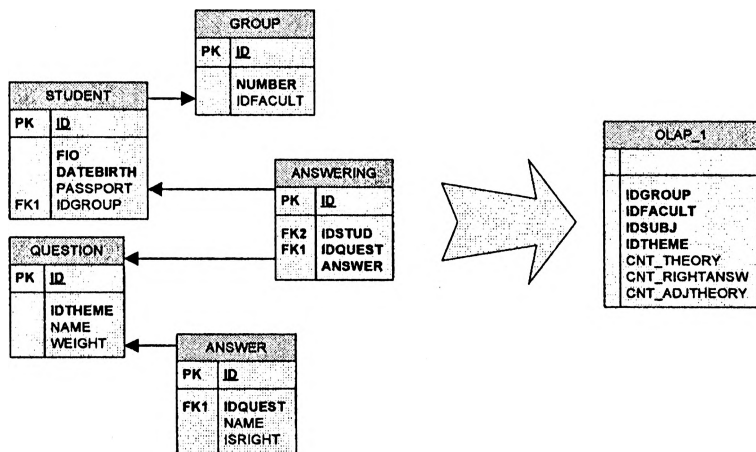


Рис. 2. Структура хранилища данных.

Затем к информации, хранящейся в хранилище данных, должны быть применены технологии OLAP. Применение этой технологии обусловлено недостатками реляционной модели базы данных, в частности, невозможностью объединять, просматривать и анализировать данные с точки зрения множественности измерений. В процессе анализа данных, для аналитика

важно построить и изучить зависимости между различными параметрами. Кроме того, число таких параметров может варьироваться в широких пределах.

Для приведения данных к удобной аналитику форме необходимо выделение одного из атрибутов, на основании которого и будет производиться анализ. После чего из реляционного (двухмерного) представления данных, на основании выделенного параметра, формируется, трехмерная структура, называемая куб решений. В общем случае эта структура может представлять собой гиперкуб.

Над гиперкубом могут быть произведены следующие операции:

1. Срез – формируется подмножество многомерного массива данных, соответствующие единственному значению одного или нескольких элементов измерений, не входящих в подмножество. Под измерением понимается последовательность значений одного из анализируемых параметров.

2. Вращение – изменение расположения измерений, представленных в отчете или на отображаемой странице. Например, операция вращения может заключаться в перестановке местами столбцов и строк таблицы, или перемещении интересующих измерений в столбцы или строки создаваемого отчета.

3. Консолидация и детализация – операции, которые определяют переход вверх по направлению от детального представления данных к агрегированному и наоборот.

Использование этих операций позволяет пользователю-аналитику создавать произвольные запросы к данным и наглядно отображать зависимости между требуемыми параметрами.

Еще одним важным дополнением для разрабатываемой системы может послужить введение в хранилище данных атрибута, содержащего дату произведенного измерения. Такая модификация позволит проводить анализ по показателям, которые могут изменяться с течением времени. А именно, можно определить семестр, в котором наиболее эффективно изучать данную дисциплину, эффективность изменений, вносимых в теоретический материал, а также в учебный процесс и распределение занятий по видам нагрузки.