

Адаптируемая модель данных на основе многомерного пространства

10, октябрь 2010

авторы: Балдин А. В., Елисеев Д. В.

УДК. 004.652

МГТУ имени Н.Э. Баумана

d-eli@mail.ru

Введение

Адаптируемость (характеристика, определяющая способность системы к развитию в соответствии с нуждами пользователей и бизнеса.) является одним из наиболее важных требований, предъявляемых к информационным системам различного назначения. Адаптируемость рассматривают достаточно широко, включая в это понятие такие взаимосвязанные нефункциональные требования как способность к развитию, гибкость, расширяемость, интероперабельность и т.п. [1].

Адаптация информационных систем - это процесс их настройки на меняющиеся условия эксплуатации, потребности пользователей и бизнес-процессы как при создании новых систем, так и при сопровождении существующих [1]. Этот итеративный процесс можно считать важнейшей частью жизненного цикла автоматизированной системы.

Адаптируемые системы - это «легко изменяемые» системы, включающие средства, которые обеспечивали бы их настройку на новые требования и условия динамически (в ходе эксплуатации), облегчали бы их сопровождение [1].

Потребность в адаптируемых информационных системах возникает в предметных областях, в которых постоянно меняются требования к системе, и от того, насколько быстро система сможет приспособиться к новым требованиям, зависит выполнение решаемых системой задач. Примерами таких предметных областей могут служить области, связанные с системой управления вузом [2, 3], системой управления персоналом, экономикой и т.д.

Адаптируемая информационная система состоит из подсистемы хранения данных и интеллектуальной подсистемы [4]. Функционирование интеллектуальной подсистемы зависит от хранимых знаний, т.е. подсистему можно изменять (модифицировать) через изменение хранящихся знаний. Т.к. интеллектуальная подсистема должна взаимодействовать с подсистемой хранения данных, то важной задачей, определяющей способность системы к развитию, является разработка адаптируемой модели данных.

1 Табличное представление данных

Адаптируемость автоматизированной информационной системы в первую очередь определяется свойствами модели представления данных. В настоящее время одним из основных средств описания и хранения данных являются реляционные СУБД. Это связано с тем, что реляционная модель проста в проектировании и реализации по сравнению с другими моделями данных (иерархическими, сетевыми, объектными) и имеет мощный математический аппарат, основанный на теории множеств, нормализации схем отношений и реляционной алгебре.

Одним из недостатков ограничивающих применение реляционных СУБД для создания адаптируемых систем является недостаточная гибкость структуры данных. Реляционная модель реализуется в СУБД в виде таблиц, описывающих отношения [5, 6].

Возьмем некоторую реляционную таблицу. Структура реляционной таблицы представлена на рис. 1.



Рис. 1. Структура реляционной таблицы

Для любой такой таблицы существует отдельное множество заголовков столбцов (атрибутов) описания отношения. При традиционном подходе к реляционным таблицам количество столбцов-атрибутов может быть любым, но это только при проектировании такой таблицы. Как только описание таблицы задано, т.е. зафиксировано некоторое количество атрибутов отношения, после этого изменять структуру реляционной таблицы: добавлять, удалять или изменять атрибуты отношения нельзя. При этом наименьшей адресуемой частью реляционной таблицы, при выполнении операций добавления, удаления или изменения, является не отдельная клеточка, а вся строка-запись, т.е. все клетки одной строки таблицы. Таким образом, в реляционных таблицах можно лишь накапливать и обрабатывать целые строки-записи таблицы. Следовательно, реляционные таблицы являются одномерными, так как добавление, удаление или изменение производится сразу для всех атрибутов одновременно, т.е. «в длину», а изменение таблицы «в ширину» (изменение структуры таблицы при функционировании системы) - запрещено.

В результате, с помощью табличного представления данных могут быть описаны только такие предметные области, в которых можно однозначно зафиксировать, определить некоторые сущности, затем, для каждой сущности однозначно зафиксировать определенные атрибуты и некоторые связи между различными сущностями. В случае

выявления либо новой сущности, либо нового свойства, либо новой связи система представления данных должна быть спроектирована заново или перепроектирована, т.е. практически каждый раз создается новая модель (структура представления) предметной области [7].

2 Многомерное дискретное информационное пространство

2.1 Описание трёхмерного дискретного информационного пространства

Как правило, в адаптируемой информационной системе происходит накопление и обработка разнородной информации [2, 3]. Накопление информации приводит к необходимости обучения, постоянного уточнения и изменения системы представления данных: выявляются новые объекты, появляются новые свойства, возникают новые отношения между объектами. Кроме того, возможно удаление или замена устаревших понятий. Отметим, что поведение системы обычно рассматривается в фиксированном пространстве, причем поведение системы - это изменение ее состояний во времени, исходом которого является некоторый результат [7]. Однако, использование фиксированного пространства состояний не позволяет описать поведение сложной развивающейся системы. Поэтому для создания моделей адаптируемых (развивающихся) систем, способных более адекватно отображать изменение системы, ее развитие, необходимо применять другой способ представления данных (отличный от табличного представления).

Для решения этой задачи введем три равнозначных понятия: вещь, (объект, сущность); свойство, (атрибут, характеристика); отношение, (связь, взаимодействие). Этим трех выделенных понятий-категорий достаточно для представления любой информации о любой предметной области [7].

В процессе анализа выделяется некоторая сущность в предметной области, а затем присваивается ей некоторое уникальное название. Если требование уникальности не будет выполнено, то это приведет к неоднозначности и к ошибкам. После этого, определяются атрибуты сущности, на основе которых и произошло ее выделение из предметной области. Каждому атрибуту присваивается определенное уникальное название (имя, идентификатор). Затем, изучаются все отношения выделенной сущности с другими сущностями данной предметной области. Каждому отношению присваивается уникальное название (идентификатор).

Допустим, что все три множества V , S и O (сущности, атрибуты, отношения) для некоторой конкретной предметной области являются счетными множествами. Предположим, что границы такой предметной области достаточно большие, а в некоторых случаях эти границы могут постепенно увеличиваться. Отметим, что опыт научных исследований и практической работы показывает, что таких больших и актуальных предметных областей достаточно много [7]. Тогда можно построить трехмерное дискретное информационное пространство: $\langle \text{сущность, атрибут, отношение} \rangle$, или $\langle V, S, \theta \rangle$.

Наименьший элемент этого пространства, то есть трехмерную точку, будем называть *мивар*, а само пространство – *миварным пространством* [7].

Мивар — это наименьшая адресуемая точка трехмерного пространства $\langle \text{сущность, атрибут, отношение} \rangle$. Миварное пространство представлено на рис. 2.

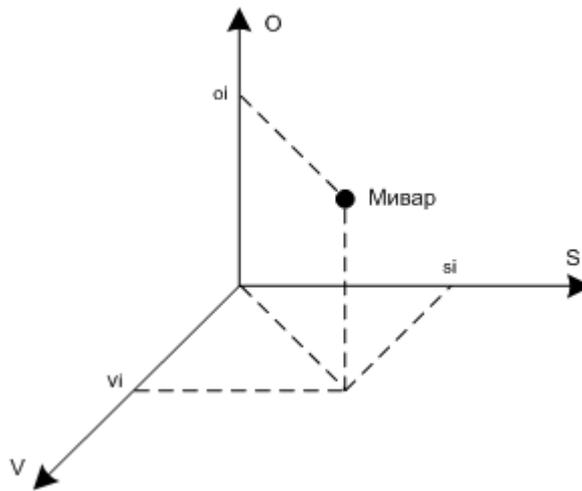


Рис.2. Миварное пространство представления данных

2.2 Формализованное описание многомерного пространства представления данных

Формализованное описание многомерного пространства представления данных можно представить в виде совокупности множеств.

$$A = \{a_i\},$$

$i = \overline{1, N}$; A – множество названий осей многомерного пространства,

N – количество осей многомерного пространства.

Множество A содержит структуру многомерного пространства, т.е. совокупность необходимого числа осей.

$$MM = B_1 \times B_2 \times \dots \times B_N,$$

где MM – многомерное пространство, B_n – множество значений оси a_n ,

$B_n = \{b_{n i_k}\}$, $n = \overline{1, N}$, $i_k = \overline{1, I_n}$, i_k – идентификатор элемента множества оси a_n .

Множество MM – декартово произведение множеств значений осей многомерного пространства. Именно с помощью элементов множеств B_n происходит описание представления данных предметной области.

Тогда модель данных описывается множеством

$$M = \{c_{i_1, i_2, \dots, i_N} \mid i_1 = \overline{1, I_1}; \dots; i_N = \overline{1, I_N}\},$$

где M – модель данных,

c_{i_1, i_2, \dots, i_N} – значениеточки многомерного пространства с координатой $\langle i_1, i_2, \dots, i_N \rangle$.

2.3 Основные возможности многомерного представления данных

Многомерная концепция представления данных позволяет работать с динамическими (изменяющимися, эволюционными) структурами хранения данных, что открывает новые возможности по созданию адаптируемых систем сбора и обработки информации. Многомерное представление данных позволяет использовать неявные

ассоциативные связи различных понятий и объектов [7]. Это означает, что на основе анализа структуры хранящихся данных можно получать дополнительную информацию, которая не содержится в явном виде в базе данных. В многомерном пространстве можно ввести понятие меры близости - расстояния либо между отдельными точками, либо между их совокупностями. Можно использовать меру подобия, схожести различных структур. Это открывает принципиально новые возможности для систем, использующих механизм логического вывода или обработку изображений.

Миварное представление включает в себя идеи объектно-ориентированного программирования, более того, оно раскрывает новые возможности и перспективы. В любое время можно либо расширить описание исследуемых объектов, либо уменьшить его [7].

Ещё одна особенность многомерного информационного пространства заключается в том, что при необходимости можно вводить дополнительные оси и размерности. Многомерное пространство может быть разбито на отдельные подпространства, можно и, наоборот, по определенным правилам соединять отдельные представления в более общее.

С точки зрения сетевых моделей (семантических сетей), миварное представление - это некоторая сеть, находящаяся в координатном трехмерном пространстве, что только расширяет возможности сетевых моделей. С точки зрения реляционных моделей, миварное представление - это трехмерная (или N-мерная) реляционная таблица, находящаяся в трехмерном пространстве, в котором собраны все обычные реляционные таблицы.

3 Многомерное пространство для адаптируемой модели данных

3.1 Структура многомерного пространства для адаптируемой модели данных

Определим общую структуру многомерного пространства для описания адаптируемой модели данных, основанной на реляционной модели.

Реляционная модель данных – это множество нормализованных отношений (таблиц), к которым применимы операции реляционной алгебры [5]. В соответствии с разделом 1 данной статьи каждое отношение включает в себя множество атрибутов и множество записей, которые определяются ключом отношения. Таким образом, для описания реляционной модели данных в многомерном пространстве необходимо ввести три оси: ось сущностей предметной области (название отношения), ось атрибутов сущностей и множество идентификаторов записи сущности.

Тогда для реляционной модели данных трёхмерное дискретное информационное пространство, описанное в разделе 2.1, преобразуется в пространство *<сущность, атрибут, идентификатор>*.

Назовём пространство *<сущность, атрибут, идентификатор>* определяющим, и соответственно образующие это пространство оси – определяющими осями. Тогда значение каждой точки в определяющем пространстве зависит от параметров - других осей, образующих многомерное пространство.

Адаптируемая система изменяется (развиваться) со временем. Поэтому необходимо ввести ось времени, которая будет определять состояние модели данных при описании поведения адаптируемой информационной системы.

Состояние модели данных в момент времени t – совокупность точек многомерного пространства, значения которых в определяющем пространстве по параметру оси времени являются последними до заданного момента времени t . Таким образом, для реляционной модели состоянием модели данных является совокупность отношений в определённый момент времени.

Для обеспечения многопользовательского режима работы и безопасности хранимой информации следует определить ось модификаторов системы, т.е. множество пользователей системы. Данная ось определяет, какой пользователь изменил состояние модели данных.

Таким образом, структура многомерного пространства для адаптируемой модели данных состоит из пяти основных множеств (осей)

- 1) V – множество сущностей предметной области. $V = \{v_i\}, i = \overline{1, I_V}, I_V = |V|$
- 2) S – множество атрибутов сущностей предметной области. $S = \{s_i\}, i = \overline{1, I_S}, I_S = |S|$.
- 3) ID – множество идентификаторов записи сущности. Множество идентификаторов записей сущности будем считать множество натуральных чисел. В каждой сущности идентификатор выделяет конкретную запись из отношения. Поэтому в каждом отношении идентификатор записи не должен повторяться, должен быть уникальным.
- 4) T – множество времён изменений состояний модели данных. Введём на множестве T отношение порядка (T, \leq) , т.е. $t_1 \leq t_2$, когда событие t_2 произошло не раньше события t_1 .
- 5) U – множество модификаторов. Содержит идентификатор пользователя, который произвёл изменение значения точки многомерного пространства. $U = \{u_i\}, i = \overline{1, I_U}, I_U = |U|$.

Тогда многомерное пространство будет иметь следующий вид:

$$MM = V \times S \times ID \times T \times U.$$

Если $m \in MM$, то $m = \langle v, s, id, t, u \rangle$ – точка многомерного пространства.

3.2 Работа с многомерным пространством для адаптируемой модели данных

В соответствии с разделом 3.1 многомерное пространство для адаптируемой модели данных хранит отношения реляционной модели предметной области. Поэтому для получения результата запроса целесообразно разделить операции на два типа (рис. 3).

1. Операции над элементами осей многомерного пространства (операции с координатами точек). Данные операции выполняются над координатами точек многомерного пространства, не анализируя значения этих точек. Это операции по выделению различных подпространств, соответствующих определённым условиям по осям (операции по получению срезов многомерного пространства).
2. операции реляционной алгебры (язык SQL) [5, 6].

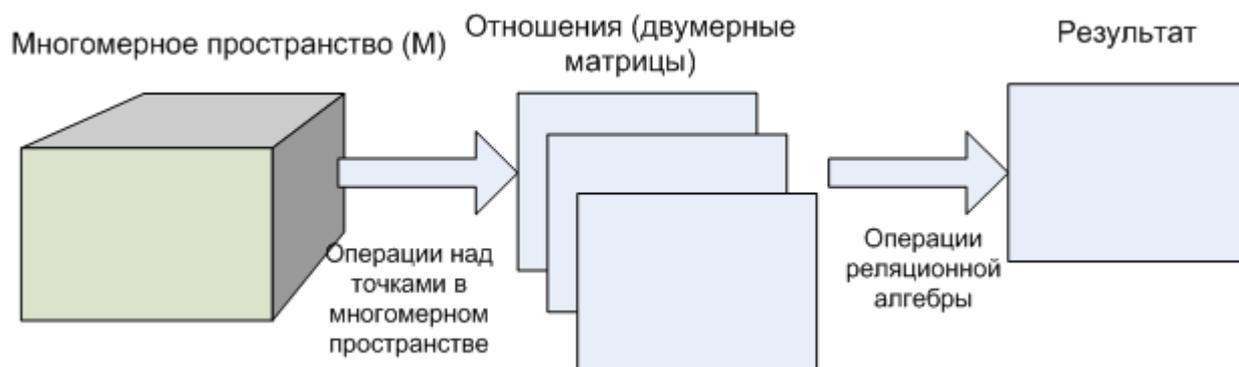


Рис. 3. Схема работы с многомерным информационным пространством.

Введём функцию получения значения точки многомерного пространства:

$$c_m = Cx(m).$$

Операции с координатами точек выделяют требуемое подпространство. Затем выделенное подпространство преобразуется в отношения предметной области, с помощью многократного применения функции $Cx(m)$ ко всем точкам выделенного подпространства. К полученным отношениям можно в дальнейшем применять операции реляционной алгебры для определения результата запроса (рис. 3).

Заключение

В данной статье рассмотрена адаптируемая модель данных на основе многомерного пространства. Главным отличием многомерного представления данных от табличного представления, используемого в реляционных СУБД, является то, что можно в любой момент времени добавлять новые элементы во множества, определяющие оси многомерного пространства. Тем самым происходит изменение модели данных. Табличное представление реляционной модели не позволяет изменять саму модель при функционировании автоматизированной системы.

Многомерное пространство представления данных можно разделить на три уровня: оси многомерного пространства, декартово произведение множеств значений осей и значения точек пространства.

Определённые в статье пять осей многомерного пространства (сущности, атрибуты, идентификаторы, время, модификаторы) являются основными для проектирования адаптируемой автоматизированной системы, но не запрещается использовать дополнительные оси для большей детализации хранимых данных.

Определённая в статье структура многомерного пространства для адаптируемой модели данных является гибкой средой для хранения реляционной модели предметной области, т.к. с помощью многомерного пространства отношения реляционной модели формируются динамически в зависимости от условий, накладываемых на координаты точек.

Таким образом, адаптируемую модель представление данных на основе многомерного пространства можно использовать для создания гибких динамических автоматизированных информационных систем.

Список литературы

1. Лядова Л.Н. Метамоделирование и многоуровневые метаданные как основа технологии создания адаптируемых информационных систем // <http://www.foibg.com>: сервер международных статей по информатике. URL. www.foibg.com/ibs_isc/ibs-04/IBS-04-p20.pdf
2. Информационная управляющая система МГТУ им. Н.Э. Баумана «Электронный университет»: концепция и реализация / Т.И. Агеева, А.В. Балдин, В.А. Барышников и др.; [под ред. И.Б. Фёдорова, В.М. Черненко]. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. – 376 с.
3. Информационные технологии в инженерном образовании / под ред. С.В. Коршунова, В.Н. Гузненкова. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2007. – 432 с.
4. Модель представления знаний при создании адаптивной информационной системы [Электронный ресурс] / Елисеев Д.В. - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2009 – Режим доступа: <http://technomag.edu.ru/doc/139659.html>, свободный
5. Гарсиа-Молина, Гектор, Ульман, Джеффри, Д., Уидом, Дженнифер Г Системы баз данных. Полный курс. : Пер. с англ. — М.: Издательский дом "Вильямс", 2003. — 1088 с.
6. Четвериков В.Н., Ревунков Г.И., Самохвалов Э.Н. Базы и банки данных: Учебник для вузов/Ред. Четвериков В.Н.-М.:Высш.шк.,1987.-245 с.
7. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. - М: Радио и связь, 2002.-286 с.