

Алгебра многомерных матриц для обработки адаптируемой модели данных

07, июль 2011

авторы: Елисеев Д.В., Балдин А.В.

УДК. 004.652

МГТУ имени Н.Э. Баумана

d-eli@mail.ru

bal@bmstu.ru

Введение

Адаптируемая информационная система состоит из подсистемы хранения данных и интеллектуальной подсистемы [1]. 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 представления 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 [2]. Это связано с тем, что реляционная модель крайне проста в проектировании и реализации по сравнению с другими моделями данных (иерархическими, сетевыми, объектными) и имеет мощный математический аппарат, основанный на теории множеств и реляционной алгебре.

Одним из недостатков ограничивающих применение реляционных СУБД для создания адаптируемых систем является недостаточная гибкость структуры данных. Многомерное представление реляционной модели данных позволяет устранить этот недостаток [2]. 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 [5]. 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 [2]. Такая структура позволяет вводить как дополнительные оси и размерности, так и дополнять множества элементов существующих осей в любой момент времени.

Для реализации адаптируемой модели данных на основе многомерного пространства необходимо определить обобщённую структуру такого пространства и ввести ряд операций по обработке данных.

1 Многомерное пространство для адаптируемой модели данных

1.1 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 [3]. Каждое отношение включает в себя множество атрибутов и множество записей, которые определяются ключом отношения [2, 4]. 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.

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 [5]. 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.

1.2 Работа с многомерным пространством для адаптируемой модели данных

В соответствии с разделом 1.1 многомерное пространство для адаптируемой модели данных хранит отношения реляционной модели предметной области. Поэтому для получения результата запроса целесообразно разделить операции на два типа (рис. 1).

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) [3, 4].

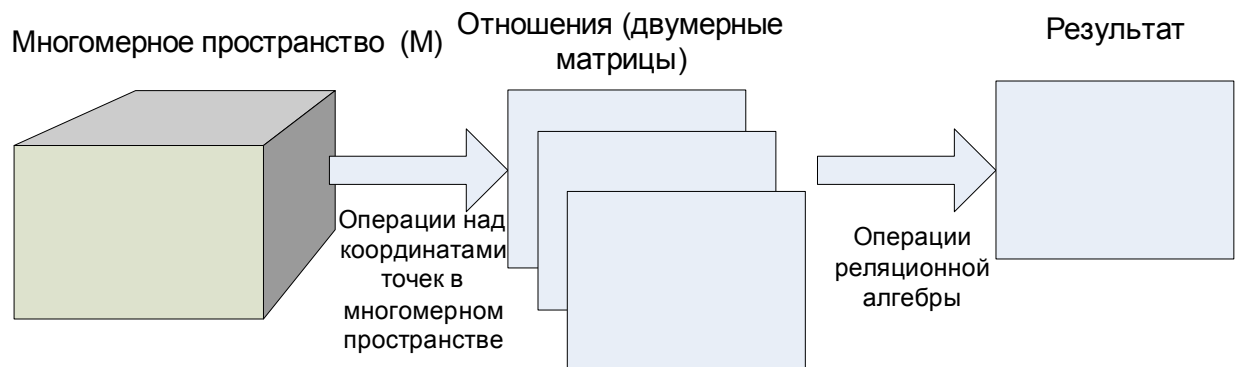


Рис. 1. Схема работы с многомерным информационным пространством.

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. 1).

2 Алгебра многомерных матриц

2.1 Определение многомерной матрицы

Определение: Р-мерная матрица (многомерная матрица) – совокупность элементов

$C_{i_1 \dots i_p}$, где индексы $i_1, \dots, i_p$ принимают значения от 1 до n_α , ($\alpha = 1, \dots, p$) соответственно.

Р-мерная матрица содержит $n_1 \times n_2 \times \dots \times n_p$ элементов и обозначается $C_M = \|c_{i_1 \dots i_p}\|$ [6].

Пример трёхмерной кубической матрицы представлен на рис. 2.

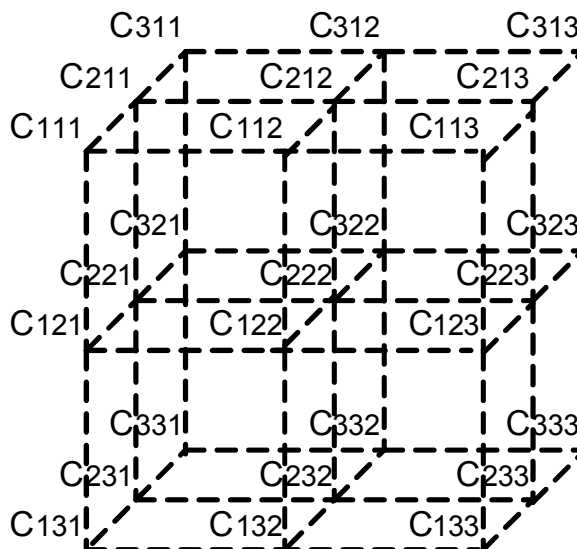


Рис. 2. Трёхмерная кубическая матрица.

Таким образом, многомерное информационное пространство C_M , описанное в разделе 1.1, может быть представлено в матричном виде следующим образом:

$C_M = \|c_{v,s,id,t,u}\|$. Матрица состоит из значений точек многомерного пространства. Если точка не входит в описание модели предметной области, тогда её значение равно «null», иначе точка содержит значение типа «скаляр», т.е. число, дата, строка или булево значение.

Введение алгебраической структуры многомерной матрицы позволяет проводить операции в многомерном пространстве над индексами (элементами осей пространства), не анализируя значения матрицы. Операции над многомерными матрицами выделяют отношения предметной области (двумерные матрицы), которые можно в дальнейшем обрабатывать с помощью операций реляционной алгебры (язык SQL). Схема работы с многомерным информационным пространством представлена на рис. 1.

2.2 Алгебра многомерных матриц

Определение: алгебра многомерных матриц – алгебра (E, Ω) , где E – множество многомерных матриц, а Ω – множество операций [6].

Определим множество многомерных матриц $E = \{C : C \cup C_M = C_M\}$, где C_M – многомерная матрица, описывающая многомерное пространство адаптируемой модели данных, а $\cup$ – операция объединения подпространств, описываемых многомерными матрицами, рассмотренная ниже. Такое определение множества многомерных матриц гарантирует, что при использовании введённых операций над многомерными матрицами мы не выйдем за пределы многомерного пространства адаптируемой модели данных.

Введём следующие операции над многомерными матрицами, в соответствии с представленной структурой многомерного пространства в разделе 1.1:

2.2.1. Область доступа.

Областью доступа для пользователя u_c будем называть многомерную матрицу следующего вида: $C_{u_c} = \Phi(C_M) = \|c_{v,s,id,t,u}\|$, $\phi(v,s,id,t,u) = \text{Истина}$.

$\phi(v,s,id,t,u)$ - функция проверки принадлежности элемента матрицы с координатами $m = \langle v,s,id,t,u \rangle$ области доступа, определённой у пользователя u_c . Область доступа делает доступным ту часть многомерного пространства, к которой есть доступ у пользователя.

2.2.2. Объединение подпространств многомерных матриц.

Пусть даны 2 многомерные матрицы

$$C_A = \|c_{v_A,s_A,id_A,t_A,u_A}\|, v_A \in V_A \subseteq V, s_A \in S_A \subseteq S, id_A \in ID_A \subseteq ID, t_A \in T_A \subseteq T, u_A \in U_A \subseteq U,$$

$$C_B = \|c_{v_B,s_B,id_B,t_B,u_B}\|, v_B \in V_B \subseteq V, s_B \in S_B \subseteq S, id_B \in ID_B \subseteq ID, t_B \in T_B \subseteq T, u_B \in U_B \subseteq U.$$

Тогда $C_D = C_A \cup C_B = \|c_{v_D,s_D,id_D,t_D,u_D}\|$, $v_D \in V_A \cup V_B$, $s_D \in S_A \cup S_B$, $id_D \in ID_A \cup ID_B$, $t_D \in T_A \cup T_B$, $u_D \in U_A \cup U_B$.

2.2.3. Пересечение подпространств многомерных матриц.

Пусть даны 2 многомерные матрицы

$$C_A = \|c_{v_A,s_A,id_A,t_A,u_A}\|, v_A \in V_A \subseteq V, s_A \in S_A \subseteq S, id_A \in ID_A \subseteq ID, t_A \in T_A \subseteq T, u_A \in U_A \subseteq U,$$

$$C_B = \|c_{v_B,s_B,id_B,t_B,u_B}\|, v_B \in V_B \subseteq V, s_B \in S_B \subseteq S, id_B \in ID_B \subseteq ID, t_B \in T_B \subseteq T, u_B \in U_B \subseteq U.$$

Тогда $C_D = C_A \cap C_B = \|c_{v_D,s_D,id_D,t_D,u_D}\|$, $v_D \in V_A \cap V_B$, $s_D \in S_A \cap S_B$, $id_D \in ID_A \cap ID_B$, $t_D \in T_A \cap T_B$, $u_D \in U_A \cap U_B$.

2.2.4. Разность подпространств многомерных матриц.

Пусть даны 2 многомерные матрицы

$$C_A = \|c_{v_A,s_A,id_A,t_A,u_A}\|, v_A \in V_A \subseteq V, s_A \in S_A \subseteq S, id_A \in ID_A \subseteq ID, t_A \in T_A \subseteq T, u_A \in U_A \subseteq U,$$

$$C_B = \|c_{v_B,s_B,id_B,t_B,u_B}\|, v_B \in V_B \subseteq V, s_B \in S_B \subseteq S, id_B \in ID_B \subseteq ID, t_B \in T_B \subseteq T, u_B \in U_B \subseteq U.$$

Тогда $C_D = C_A \setminus C_B = \|c_{v_D,s_D,id_D,t_D,u_D}\|$, $v_D \in V_A \setminus V_B$, $s_D \in S_A \setminus S_B$, $id_D \in ID_A \setminus ID_B$, $t_D \in T_A \setminus T_B$, $u_D \in U_A \setminus U_B$.

Т.к. все многомерные матрицы являются частями одного многомерного пространства, то в разных матрицах элементы, имеющие одинаковые индексы, имеют и одинаковые значения. Поэтому при объединении, пересечении и разности подпространств

многомерных матриц нет необходимости проверять значения точек пространства. Результатом выполнения этих операций будет всегда многомерная матрица, описывающая соответствующее определённую часть общего многомерного пространства.

2.2.5. Сечение

Определение: Совокупность элементов матрицы с фиксированным значением индексов $i_1, \dots, i_k, k \leq N$ называется сечением ориентации $i_1, \dots, i_k$ [6].

Таким образом, сечением многомерной матрицы $C = \|c_{v,s,id,t,u}\|$ будем называть многомерную матрицу:

$$C_\theta = \theta_f \|c_{i_1 i_2 \dots i_n}\|,$$

где f – функция проверки логического выражения вида: $i_1 = z_1$ и $i_2 = z_2$ и ... и $i_k = z_k, k \leq N$.

Тогда сечение многомерной матрицы будет иметь вид:

$$C_\theta = \|c_{z_1 z_2 \dots z_k i_{k+1} \dots i_n}\|.$$

Назовём сечением первого уровня – n -мерную матрицу, у которой зафиксировано значение одного индекса.

Сечение второго уровня – n -мерная матрица, у которой зафиксировано значения двух индексов и т.д.

Графическое изображение операции сечения представлено на рис. 3. изображен результат операции сечения: $C_\theta = \|c_{i_1 i_2}\|$, если фиксируется третья координата в трёхмерной матрице. В результате получаем двумерную матрицу.

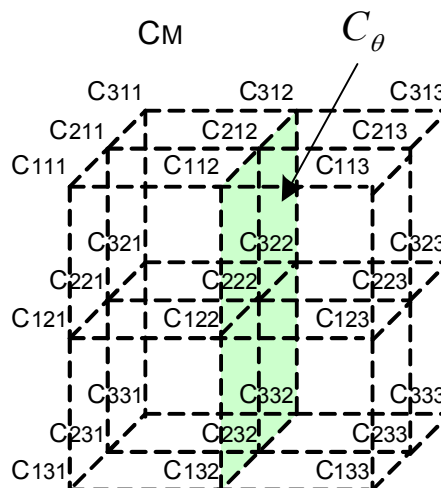


Рис. 3. Операция «Сечение»

2.2.6. Срез.

Срезом многомерной матрицы $C = \|c_{v,s,id,t,u}\|$ будем называть многомерную матрицу следующего вида:

$$C_\psi = \psi_f \|c_{i_1 i_2 \dots i_n}\|,$$

где $f(i_1, i_2, \dots, i_n)$ – функция проверки логического условия, не включающего равенство индексов определённому значению.

Графическое изображение операции среза представлено на рис. 4. представлен результат выполнения операции среза.

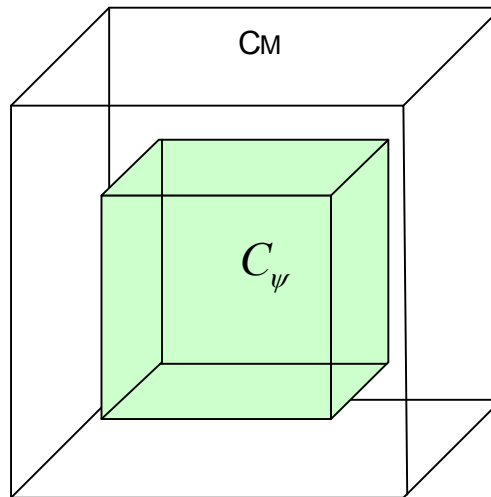


Рис. 4. Операция «Срез»

2.2.7. Срез первых.

Срезом первых многомерной матрицы $C = \|c_{v,s,id,t,u}\|$ будем называть многомерную матрицу следующего вида: $C_{\tau_F} = \tau_F(\|c_{v,s,id,t,u}\|, t_0)$.

Введём на множестве элементов многомерной матрицы отношение эквивалентности ρ_{VSId} :

$$m_1 = \langle v_1, s_1, id_1, t_1, u_1 \rangle, m_2 = \langle v_2, s_2, id_2, t_2, u_2 \rangle, \\ c_{m_2} = \rho_{VSId}(c_{m_1}) \Leftrightarrow v_1 = v_2 \text{ и } s_1 = s_2 \text{ и } id_1 = id_2.$$

Тогда множество $[c_{m_0}]_{\rho_{VSId}} = \{c_m : c_m = \rho_{VSId}(c_{m_0})\}$ – класс эквивалентности по отношению ρ_{VSId} . Множество классов эквивалентности по отношению ρ_{VSId} на множестве элементов многомерной матрицы образует разбиение этого множества и обозначается $\|c_{v,s,id,t,u}\| / \rho_{VSId}$.

Т.к. множество T является упорядоченным, элементы множества $[c_{m_0}]_{\rho_{VSId}}$ можно упорядочить по оси T и найти максимальный или минимальный элемент. Минимальный элемент множества $[c_{m_0}]_{\rho_{VSId}}$ обозначим $\min_T [c_{m_0}]_{\rho_{VSId}}$.

$$\text{Тогда } C_{\tau_F} = \tau_F(\|c_{v,s,id,t,u}\|, t_0) = \|\min_T [c_m]_{\rho_{VSId}}\|, \text{ где } \|c_m\| = \psi_{t \geq t_0} \|c_{v,s,id,t,u}\|.$$

Графическое изображение операции среза первых представлено на рис. 5. Сначала отбираются все точки многомерной матрицы, имеющие индекс по оси времени больше заданного значения t_0 . Выбранные точки обозначены на рис. 5. Затем выбранные точки разбиваются на классы эквивалентности по отношению ρ_{VSId} , и в каждом классе эквивалентности ищется наименьшая точка по отношению порядка, заданного на оси времени.

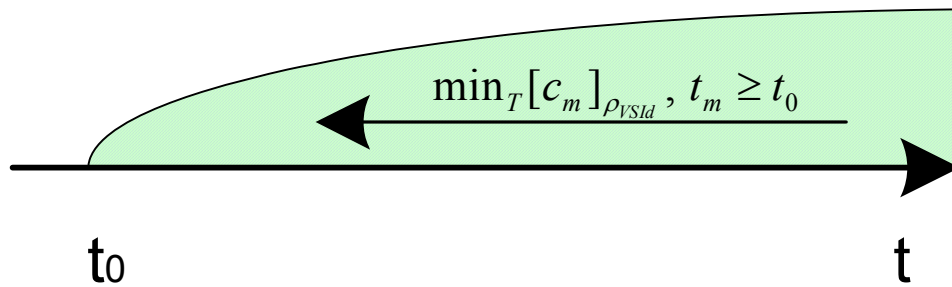


Рис. 5. Операция «Срез первых»

2.2.8. Срез последних.

Срезом последних многомерной матрицы $\|c_{v,s,id,t,u}\|$ будем называть многомерную матрицу следующего вида: $C_{\tau_L} = \tau_L(\|c_{v,s,id,t,u}\|, t_0)$.

По аналогии со срезом первых определим операцию срез последних:

$$C_{\tau_L} = \tau_L(\|c_{v,s,id,t,u}\|, t_0) = \|\max_T[c_m]_{\rho_{VSId}}\| = \|c_{v,s,id}\|, \text{ где } \|c_m\| = \psi_{t \leq t_0} \|c_{v,s,id,t,u}\|.$$

Графическое изображение операции среза последних представлено на рис. 6. Сначала отбираются все точки многомерной матрицы, имеющие индекс по оси времени меньше заданного значения t_0 . Выбранные точки обозначены на рис. 6. Затем выбранные точки разбиваются на классы эквивалентности по отношению ρ_{VSId} , и в каждом классе эквивалентности ищется наибольшая точка по отношению порядка, заданного на оси времени.

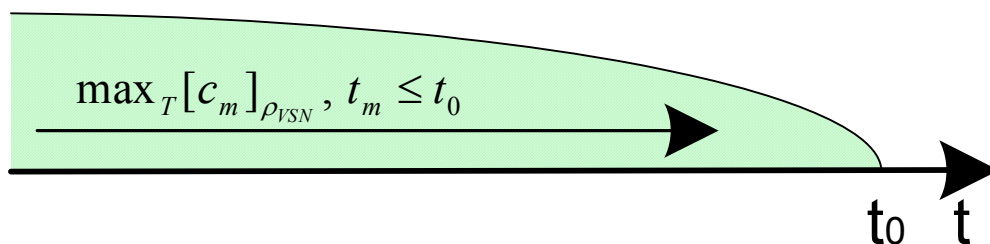


Рис. 6. Операция «Срез последних»

Дальнейшие действия с отношениями происходят с помощью операций реляционной алгебры.

Таким образом, алгебра многомерных матриц определяется следующим образом:

$$A = (D, \phi: D \rightarrow D, \cup: D \times D \rightarrow D, \cap: D \times D \rightarrow D, \setminus: D \times D \rightarrow D, \theta_f: D \rightarrow D, \psi_f: D \rightarrow D, \tau_F: D \times T \rightarrow D, \tau_L: D \times T \rightarrow D).$$

Заключение

В данной статье рассмотрена обобщённая структура многомерного пространства для адаптируемой модели данных и введена алгебра многомерных матриц для работы с этим пространством.

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.

Операции алгебры многомерных матриц можно рассматривать как расширение над операциями реляционной алгебры и существующим языком SQL. Эти операции используются для динамического формирования отношений реляционной модели, хранимых в многомерном пространстве, в зависимости от условий, накладываемых на координаты точек. Затем для получения окончательного результата запроса над полученными отношениями нужно применить операции реляционной алгебры.

Операции алгебры многомерных матриц предназначены для обработки индексов элементов матрицы. Эти операции не анализируют значения элементов матрицы, а формируют отношения реляционной модели в зависимости от условий, накладываемых на индексы. Алгебра определена на множестве многомерных матриц, получаемых из многомерного пространства, описывающего адаптируемую модель данных. Это гарантирует, что при применении введённых операций над многомерными матрицами мы не выйдем за пределы многомерного пространства адаптируемой модели данных.

Таким образом, алгебра многомерных матриц является математическим аппаратом, позволяющим выполнять операции с адаптируемой моделью данных на основе многомерного пространства.

Список литературы

1. Модель представления знаний при создании адаптивной информационной системы [Электронный ресурс] / Елисеев Д.В. - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2010 – Режим доступа: <http://technomag.edu.ru/doc/139659.html>, свободный, (Дата обращения: 02.07.2011)
2. Адаптируемая модель данных на основе многомерного пространства [Электронный ресурс] / Балдин А.В., Елисеев Д.В. - Электрон. журн. – М.: «Наука и образование: электронное научно-техническое издание», 2010 – Режим доступа: <http://technomag.edu.ru/doc/161410.html>, свободный, (Дата обращения: 02.07.2011)
3. Гарсиа-Молина, Гектор, Ульман, Джеффри, Д., Уидом, Дженнифер Г Системы баз данных. Полный курс. : Пер. с англ. — М.: Издательский дом "Вильямс", 2003. — 1088 с.
4. 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.
5. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. - М: Радио и связь, 2002.-286 с.
6. Реализация матричной модели данных в иерархических структурах [Электронный ресурс] / Сергеев В. П. - Электрон. журн. – Математическая морфология., 2007 – Режим доступа: <http://www.smolensk.ru/user/sgma/MMORPH/N-14-html/sergeev/sergeev.htm>, свободный, (Дата обращения: 02.07.2011)