

**сообщения
объединенного
института
ядерных
исследований
дубна**

P10-87-62

А.Г.Романенко *, Ю.В.Столярский

**ДОКУМЕНТАЛЬНО-ФАКТОГРАФИЧЕСКАЯ
МОДЕЛЬ ДАННЫХ**

*Московский государственный историко-архивный институт

1987

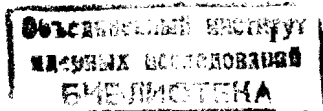
Современные автоматизированные информационные системы (АИС) можно условно разделить на два класса: фактографические и документальные.

Фактографические информационные системы в ответ на конкретные запросы о данных выдают конкретные ответы, содержащие по мере возможности только действительно запрашиваемые данные. Документальные системы обслуживают и тех потребителей, которых могут интересовать не фактические данные, а, например, анализ положения в той или иной области. Обычно в ответ на запросы такие системы выдают подборки документов.

Системы управления базами данных (СУБД), предназначенные для работы с фактографическими АИС, основываются на использовании определенных моделей данных. В настоящее время не существует общепринятого определения модели данных, однако в работе^{1/} делается попытка обобщить существующие представления о моделях данных. Большинство современных реализаций фактографических АИС применяет иерархическую (например, системы IMS и System2000) и сетевую (например, системы IDMS, IDS-II, DBMS-10, DMS-1100) модель данных. Разработан также ряд экспериментальных (System R, INGRES, MRS, Pascal/R) и несколько коммерческих систем, использующих реляционную модель данных.

Документальные (или информационно-поисковые) АИС, базирующиеся на программных системах (ПС), строятся на основе использования семантических моделей слабоструктурируемой информации. Основные теоретические положения, используемые при построении документальных систем, а также обзор существующих коммерческих и экспериментальных документальных ПС, таких как DIALOG, STAIRS, BRS, MEDLARS, MEDLINE, ORBIT, The Information Bank, LEXIS, WESTLAW, SMART и других, приводятся в работах^{2,3/}.

В работе^{4/} показано, что в современных АИС, базирующихся на СУБД, методы, модели и, соответственно, языки описания, обработки и поиска данных существенно отличаются от методов, моделей и языков систем, базирующихся на программных системах. Основной причиной таких различий является несовместимость существующих моделей структурируемых данных (иерархической, сетевой, реляционной), используемых в СУБД, с семантическими моделями слабоструктурируемой информа-



ции, положенными в основу ИС, реализующих информационно-поисковые системы. Задачи ускоренного развития научно-технического прогресса требуют создания интегрированных АИС, способных обрабатывать как структурируемую, так и неструктурируемую информацию. Проблема интеграции неоднородных баз данных на основе обеспечения совместимости существующих моделей данных или разработки обобщенной модели или системы совместимых моделей становится крайне актуальной. В работе приводится попытка разработки обобщенной модели представления документально-фактографической информации.

АИС представляет собой модель некоторой области реального мира (предметной среды). Совокупность моделей данных, положенных в основу АИС, определяет адекватность АИС предметной среде.

Для построения модели данных сделаем следующие предположения относительно предметной среды, отображаемой в АИС:

- а) предметная среда адекватно отображается в АИС совокупностью введенных и подлежащих вводу документов и отношений между ними;
- б) каждый документ может быть описан конечным набором свойств, в том числе свойств, отображающих свойства предметной среды с конечным числом их значений и отношений между ними;
- в) основное смысловое содержание свойств предметной среды, отображаемых в документе, может быть адекватно описано конечной совокупностью дескрипторов с заданными отношениями между ними.

Для формализованного представления документов разобьем множество свойств предметной среды, отображаемой в документах, X_0 на два подмножества: подмножество жесткоструктурируемых свойств X , назовем их фактографическими свойствами; подмножество неструктурируемых и слабоструктурируемых свойств X' , выражающих смысловое содержание процессов и явлений предметной среды, отображаемых в документах, назовем их семантическими свойствами. Обозначим далее:

$D = \{d_1, d_2, \dots, d_i, \dots, d_M\}$ - множество документов, вводимых в АИС;

$X = \{x_1, x_2, \dots, x_j, \dots, x_n\}$ - подмножество фактографических свойств, отображающих жесткоструктурируемые свойства предметной среды и самих документов (характеристик и объектов) - данные, приводимые в документах, элементы библиографического описания документов и т.д.;

$\hat{x} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_j, \dots, \hat{x}_n\}$ - семейство множеств значений элементов подмножества X , конечное множество всех значений x_j -го свойства предметной среды или документа, $j = 1, \dots, n$;

$T = \{t_1, t_2, \dots, t_k, \dots, t_N\}$ - множество дескрипторов некоторого тезауруса T_0 предметной среды;

$\hat{t} = \{t_1, t_2, \dots, t_k, \dots, t_N\}$ - семейство множеств значений смысловыразительных способностей дескрипторов множества T , t_k - конечное множество всех смысловыразительных значений дескриптора t_k в документах множества D , $k = 1, \dots, N$.

В соответствии с гипотезой "б" каждый документ d_i может быть описан набором свойств $X(d_i)$ и $X'(d_i)$ с конечным числом их значений и отношений между ними, т.е.

$$d_i = S_0(S[X(d_i)], S'[X'(d_i)])$$

Здесь S - семантическая модель отношений между структурируемыми свойствами предметной среды, S' - семантическая модель отношений между неструктурируемыми свойствами предметной среды, S_0 - семантическая модель отношений между структурируемыми и неструктурируемыми свойствами предметной среды.

Пусть документ d_i характеризуется набором свойств (x_{i1}, \dots, x_{in}) с их значениями $(\hat{x}_{i1}, \dots, \hat{x}_{in})$, тогда

$$S[X(d_i)] = S[x_{i1}\hat{x}_{i1}, \dots, x_{ij}\hat{x}_{ij}, \dots, x_{in}\hat{x}_{in}]$$

В соответствии с гипотезой "в" основное смысловое содержание документа d_i , отображающее подмножество семантических свойств предметной среды $X'(d_i)$, может быть описано конечной совокупностью дескрипторов T_i с заданными отношениями между ними, т.е.

$$S'[X'(d_i)] = T_i[t_{i1}\hat{t}_{i1}, \dots, t_{ik}\hat{t}_{ik}, \dots, t_{in}\hat{t}_{in}]$$

где T_i - семантическая модель, отображающая отношения между дескрипторами T_i ; \hat{t}_{ik} - смысловыразительное значение дескриптора t_k в документе d_i . Для простоты будем считать, что $0 \leq \hat{t}_{ik} \leq 1$, $i = 1, \dots, M$; $k = 1, \dots, N$. В частности, если $\hat{t}_{ik} = 0$, дескриптор t_k не несет никакой смысловой нагрузки в d_i ; $\hat{t}_{ik} = 1$ - дескриптор t_k несет максимальную смысловую нагрузку в d_i .

С учетом сказанного документ d_i описывается моделью

$$S_0(S[x_{i1}\hat{x}_{i1}, \dots, x_{in}\hat{x}_{in}]; T_i[t_{i1}\hat{t}_{i1}, \dots, t_{in}\hat{t}_{in}])$$

Множество документов $D = \{d_1, \dots, d_i, \dots, d_M\}$, каждый из которых описывается приведенной моделью, представим в виде таблицы I. Каждому документу $d_i \in D$ в этой таблице поставлен в соответствие кортеж значений фактографических свойств $\langle \hat{x}_{i1}, \dots, \hat{x}_{in} \rangle$, $i = 1, \dots, M$, и кортеж семантических значений дескрипторов $\langle \hat{t}_{i1}, \dots, \hat{t}_{in} \rangle$, $i = 1, \dots, M$.

Первый столбец этой таблицы, содержащий названия документов, назовем схемой документов 0-го ранга, а сами названия документов

$\hat{d}_1, \dots, \hat{d}_i, \dots, \hat{d}_M$ - документами 0-го ранга. Столбцу \hat{d} присвоим имя "названия документов":

Таблица I

\hat{d}	$x_1 \dots x_j \dots x_n$	$t_1 \dots t_l \dots t_N$
\hat{d}_1	$\hat{x}_{11} \hat{x}_{1j} \hat{x}_{1n}$	$\hat{t}_{11} \hat{t}_{1l} \hat{t}_{1N}$
.
\hat{d}_i	$\hat{x}_{i1} \hat{x}_{ij} \hat{x}_{in}$	$\hat{t}_{i1} \hat{t}_{il} \hat{t}_{iN}$
.
\hat{d}_M	$\hat{x}_{M1} \hat{x}_{Mj} \hat{x}_{Mn}$	$\hat{t}_{M1} \hat{t}_{Ml} \hat{t}_{MN}$

Первую строку, содержащую названия фактографических свойств x_1, \dots, x_n и дескрипторов t_1, \dots, t_N , назовем схемой свойств 0-го ранга, а элементы строки - свойствами 0-го ранга. Будем различать: схему фактографических свойств 0-го ранга - x_1, \dots, x_n ; схему дескрипторов 0-го ранга - t_1, \dots, t_N (заметьте, что принципиально столбец "названия документов" можно рассматривать как один из столбцов фактографических свойств документов! Тогда $\hat{d}_1, \dots, \hat{d}_M$ - совокупность значений этого свойства, а \hat{d} - название свойства).

Документы $\hat{d}_1, \dots, \hat{d}_M$ (фактографические свойства документов x_1, \dots, x_n , дескрипторы t_1, \dots, t_N) 0-го ранга объединим по какому-либо признаку (основанию) в непересекающиеся классы. Имена получившихся классов документов (фактографических свойств, дескрипторов) назовем документами (фактографическими свойствами, дескрипторами) 1-го ранга. Те документы (фактографические свойства, дескрипторы), которые не вошли ни в один класс 1-го ранга, образуют самостоятельные классы документов (фактографических свойств, дескрипторов) 1-го ранга со своими прежними именами. Объединение в классы документов (фактографических свойств, дескрипторов) 1-го ранга образует документы (свойства, дескрипторы) 2-го ранга. Необъединившиеся объекты, как и ранее, образуют самостоятельные классы объектов 2-го ранга с их прежними именами. Продолжим этот процесс до тех пор, пока во всех поддеревах получим единственный класс наивысшего ранга. Объединим единичные классы всех ветвей, поставив им в соответствие некоторый условный класс документов (фактографических свойств, дескрипторов) и присвоив ему некоторое имя.

Полученная структура, каждый класс 1-го ранга которой связан бинарным отношением с классом (i+1)-го ранга, представляет собой ие-

рархический классификатор документов (фактографических свойств, дескрипторов) с жесткой иерархией^{4/}.

Определение I. Построенный классификатор назовем одноаспектной иерархической схемой документов (фактографических свойств, дескрипторов), а имя класса высшего ранга классификатора назовем именем схемы документов (фактографических свойств, дескрипторов).

Заметим, что некоторые элементы классификатора до i-го ранга могут составлять самостоятельные классы, а на (i+1)-м шаге войти в состав более общего класса (рис. I).

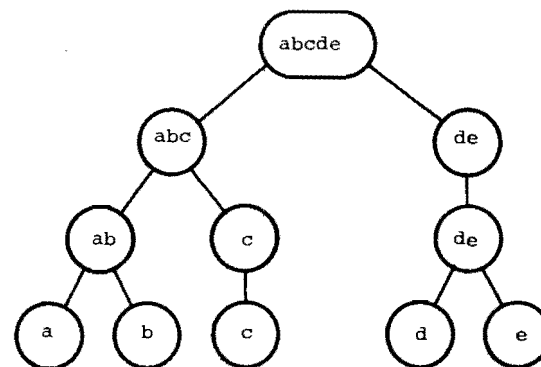


Рис. I

(Такое расширение классификатора необходимо для обеспечения индексирования и поиска информации с использованием дескрипторов любых уровней иерархии реальных тезаурусов. Оно необходимо в силу того факта, что для любого документа смысловыразительное значение дескриптора (i+1)-го уровня иерархии тезауруса в общем случае не равно сумме смысловыразительных значений всех дескрипторов i-го уровня, связанных бинарным отношением с дескриптором (i+1)-го уровня. Например, для любого документа, заиндексированного в соответствии с тезаурусом (рис. I), в общем случае $\hat{a}\hat{b} \neq \hat{a} + \hat{b}$).

Для одних и тех же множеств документов $\{\hat{d}_1\}$, фактографических свойств $\{x_1\}$ и дескрипторов $\{t_k\}$ можно построить несколько иерархических схем документов, фактографических свойств и дескрипторов по различным аспектам классификации.

Определение 2. Совокупность нескольких одноаспектных иерархических схем документов (фактографических свойств, дескрипторов), построенных по разным основаниям, назовем многоаспектной иерархической схемой документов S_0^* (фактографических свойств S^* , дескрипторов T^{**}). Нетрудно видеть, что многоаспектная иерархическая схема фактографических свойств документов есть некоторая семантическая модель отношений между структурируемыми свойствами предметной среды; многоаспектная иерархическая схема дескрипторов - некоторая семантическая модель отношений между неструктурируемыми свойствами предметной среды; многоаспектная иерархическая схема документов - некоторая семантическая модель отношений между структурируемыми и неструктурируемыми свойствами предметной среды. То есть в пределах решаемых задач указанные схемы можно рассматривать в качестве соответствующих моделей: $S_0^* = S_0; T^{**} = T^*$; $S^* = S$. Адекватность этих моделей предметной среде увеличивается с ростом числа аспектов схем.

Пусть задана одноаспектная или многоаспектная схема (документов, фактографических свойств и/или дескрипторов). Выделим несколько групп ее элементов (группы могут пересекаться, а элементы можно брать любых рангов). Свяжем каждую из выделенных групп с некоторой символической конструкцией, имеющей название. Получившиеся схемы групп назовем схемами отношений, а имена соответствующих конструкций - именами отношений.

Определение 3. Набор получившихся схем отношений назовем реляционной схемой (документов, фактографических свойств, дескрипторов). Построенные схемы отношений удовлетворяют требованиям реляционной модели данных^{4/}.

Если при построении одноаспектных схем исключить требования о непересекаемости классов и о наличии в схеме только одной вершины (класса) высшего ранга, а также допустить наличие связей между элементами любых рангов, то построенные схемы будут сетевыми схемами (документов, фактографических свойств, дескрипторов).

Определение 4. Одноаспектную, многоаспектную, сетевую или реляционную схему назовем тезаурусом (документов, фактографических свойств, дескрипторов). Данное определение соответствует общепринятому понятию тезауруса^{5,6,7/}. Действительно, каждой одноаспектной или иной схеме соответствует некоторое отношение (вид - род, часть - целое, причина - следствие и т.д.) или их совокупность, заданные на множествах документов, их фактографических свойств или дескрипторов.

Определение 5. Совокупность схем (тезауруса) документов, схем (тезауруса) фактографических свойств и схем (тезауруса) дескрипторов назовем схемой данных.

Определение 6. Совокупность тезауруса документопотока, тезауруса фактографических свойств документопотока, тезауруса дескрипторов и связанной с ними матрицы значений характеристик назовем документально-фактографической базой данных (ДФБД).

Определение 7. Иерархической (сетевой, реляционной) базой данных назовем совокупность данных (информации) предметной среды, представленных в соответствии с иерархической (сетевой, реляционной) схемами данных.

В соответствии с введенными определениями дадим интерпретацию иерархической, сетевой, реляционной, объектно-фактографической и дескрипторной моделей данных в терминах документально-фактографической модели.

Объектно-фактографическая БД - документально-фактографическая БД, в которой схемой БД является совокупность схемы документов и схемы фактографических свойств документов, схема дескрипторов отсутствует.

Дескрипторная БД - документально-фактографическая БД, в которой:

Вариант 1. Схемой БД является схема дескрипторов (тезаурус); схемы документов и фактографических характеристик отсутствуют, а элементы матрицы несут информацию о документах (инверсная БД).

Вариант 2. Схемой БД является схема документов; схемы фактографических свойств документов и дескрипторов отсутствуют, а элементы матрицы несут информацию о дескрипторах документов (прямая БД).

Реальные АИС часто включают в себя как прямые, так и инверсные БД. В этом случае дескрипторная БД есть совокупность прямой и инверсной БД.

Иерархическая БД - документально-фактографическая БД, в которой:

Вариант 1. Схемой БД является иерархическая схема фактографических свойств документов; схема документов и схема дескрипторов отсутствуют, а элементы матрицы несут в себе информацию об иерархической схеме документов (инверсная БД).

Вариант 2. Схемой БД является иерархическая схема документов; схема фактографических свойств документов и схема дескрипторов отсутствуют, а элементы матрицы несут информацию об иерархической схеме фактографических свойств документов (прямая БД).

Как и в случае дескрипторных БД, реальные иерархические БД объединяют прямые и инверсные составляющие.

Сетевая БД - документально-фактографическая БД, в которой:

Вариант 1. Схемой БД является сетевая схема фактографических свойств документов; схемы документов и дескрипторов отсутствуют,

а элементы матрицы несут информацию о сетевой схеме документов (инверсная БД).

Вариант 2. Схемой БД является сетевая схема документов; схемы фактографических свойств и дескрипторов отсутствуют, а элементы матрицы несут информацию о схеме фактографических свойств документов (прямая БД).

Сетевые БД могут включать прямую и инверсные составляющие.

Реляционная БД – документально-фактографическая БД, в которой:

Вариант 1. Схемой БД является реляционная схема фактографических свойств документов; схемы документов и дескрипторов отсутствуют (инверсная БД).

Вариант 2. Схемой БД является реляционная схема документов; схемы фактографических свойств документов и дескрипторов отсутствуют (прямая БД).

Понятия "прямая БД" и "инверсная БД", использованные выше для всех типов БД, введены по аналогии с близкими по смыслу понятиями, принятыми в дескрипторных БД. Можно показать, что все свойства прямой и инверсной организации данных в дескрипторных БД сохраняются и в БД всех других рассмотренных типов.

Обработка и поиск информации в ДФБД

Предложенная модель объединяет достоинства реляционной, иерархической и сетевой моделей данных в сфере операций обработки и манипулирования данными (простота операций манипулирования данными, поддержание целостности данных, обеспечение независимости данных от программ и т.д.) с возможностями семантических моделей данных по обработке и поиску текстовой информации.

Действительно, манипулирование фактографическими данными (левая часть таблицы ДФБД) допускает использование любых операций реляционной алгебры. При манипулировании семантической информацией (правая часть таблицы ДФБД, табл. I) возможно применение любых операций, используемых в семантических моделях.

Рассмотрим операции обработки информации, поиска и выборки в ДФБД.

Запросом в ДФБД будем называть выражение вида

$$S_{oq}(S_q[Q_{qx}(\hat{x}_1, \dots, \hat{x}_n)]; T'_q[Q_{qt}(\hat{t}_1, \dots, \hat{t}_N)]),$$

где q – индекс запроса,

S_{oq} – схема документов запроса,

S_q – схема фактографических свойств запроса,

T'_q – схема дескрипторов запроса,

$Q_{qx}(\hat{x}_1, \dots, \hat{x}_n)$ – функционал фактографических характеристик запроса – любой имеющий смысл для данной БД набор алгебраических и/или булевских выражений с переменными $\hat{x}_1, \dots, \hat{x}_n$ и константами K_1, \dots, K_L .

$Q_{qt}(\hat{t}_1, \dots, \hat{t}_N)$ – функционал семантических характеристик запроса – любой имеющий смысл для данной БД набор алгебраических и/или булевских выражений с переменными $\hat{t}_1, \dots, \hat{t}_N$ и константами K', \dots, K^L .

В частности, могут быть использованы термины сравнения $\hat{x}_\ell Q K_i$, $\hat{x}_\ell Q x_i$, где $Q \in \{=, \neq, >, <, \geq, \leq\}$, булевские выражения типа дизъюнкции или конъюнкции термов, весовая логика, например: $\sum_{\ell=1}^N K^\ell \hat{t}_\ell \geq K'$ и т.д.

Нетрудно видеть, что функционал Q_{qt} определяет меру смыслового соответствия документов в БД запросу \mathcal{U} , дополненный некоторым пороговым значением, является критерием смыслового соответствия документов запросу. Следовательно, в качестве Q_{qt} могут быть использованы любые известные меры смыслового соответствия.

Фрагмент тезауруса
дескрипторов:

ВНЕШНОСТЬ
цвет волос
белый
коричневый
черный
цвет глаз
зеленый
карий
синий

ХАРАКТЕР
веселый
вспыльчивый
добрый
⋮

Фрагмент тезауруса
документов:

ХОББИ
искусство
живопись
музыка
⋮
спорт
теннис
туризм
⋮

Фрагмент тезауруса
фактографических свойств:

СВЕДЕНИЯ О РОЖДЕНИИ
дата рождения
пол

ФИЗИЧЕСКИЕ ДАННЫЕ
рост
вес

ТЕЛЕФОН
⋮

Рис. 2

Реализация

Для подтверждения работоспособности предлагаемой модели создана тестовая ДФБД, содержащая сведения о некоторой вымышленной группе людей. ДФБД поддерживается на ЭВМ ЕС-1061. Для физического хранения и поиска информации в базе данных используется система ДИСОД ЕС ЭВМ. Преобразование языка запросов в команды ДИСОД в соответствии с тезаурусом документов, фактографических свойств и дескрипторов производится программой, написанной на языке PL/I. Возможность работы с ДФБД можно проиллюстрировать, используя таблицу 2 и рис.2. Таблица 2 изображает фрагмент структуры, аналогичной структуре тестовой ДФБД. На рис.2 приведены фрагменты тезаурусов (без высших уровней иерархии) документов, фактографических свойств и дескрипторов, имеющих структуру, аналогичную структуре тезаурусов тестовой ДФБД. Смысловые значения дескрипторов равны 1, если дескриптор присутствует в документе, либо 0 - в противном случае. Поиск информации в базе данных осуществляется в интерактивном режиме.

Рассмотрим некоторые варианты поиска в соответствии с представленной ранее структурой запросов. Например, результатом поиска в ответ на запрос:

```

FIND ХОББИ.СПОРТ AND СВЕД-РОЖД(ПОЛ=МУЖ,ДАТА-РОЖД>1947)
    AND ФИЗ-ДАН(РОСТ>180) AND ТЕЛЕФОН
    AND ВНЕШН(ЦВ-ГЛАЗ.КАРИЙ=1);
  
```

будут сведения о рождении, физических данных и внешности, а также номер телефона всех желающих завести знакомство мужчин, занимающихся спортом, родившихся после 1947 года, имеющих рост более 180 и карие глаза. С помощью команды SHOW запрошенные сведения появятся на экране в виде:

```

СИДОРОВ ИВАН ПЕТРОВИЧ
ПОЛ = МУЖ
ДАТА РОЖДЕНИЯ = 10/10/50
РОСТ = 185
ВЕС = 90
ЦВЕТ ГЛАЗ = КАРИЙ
ЦВЕТ ВОЛОС = ЧЕРНЫЙ
ХОББИ = ТЕННИС
ТЕЛЕФОН = 505
  
```

В ответ на запрос:

```

FIND СВЕД-РОЖД(ПОЛ=ЖЕН) AND ФИЗИЧ-ДАН(ВЕС<80) AND ТЕЛЕФОН
    AND ХАРАКТЕР(ВСПЫЛ=0);
  
```

будут получены сведения о рождении, физических данных и характере, а также номер телефона всех желающих завести знакомство женщин, не имеющих вспыльчивого характера и имеющих вес менее 80 килограммов. На экране полученная информация появится в таком виде:

Таблица 2

ИВАНОВ	ПЕТРОВА	СИДОРОВ	КОЗЛОВА	СВЕД-РОЖД		ФИЗ-ДАН		ТЕЛЕФОН	ВНЕШН						ХАРАКТЕР		
				ПОЛ	ДАТА-РОЖД	РОСТ	ВЕС		ЦВ-ВОЛОС	ЦВ-ГЛАЗ	КАРИЙ	КАРИЙ	КАРИЙ	КАРИЙ		КАРИЙ	КАРИЙ
1	2	3	4	x ₁	x ₂	x ₃	x ₄	x ₅	t ₁	t ₂	t ₃	t ₄	t ₅	t ₆	t ₇	t ₈	t ₉
1	2	3	4	МУЖ	01/01/47	180	85	555	0	1	0	1	0	0	1	0	0
1	2	3	4	ЖЕН	10/10/50	170	65	444	1	0	0	0	0	1	0	0	1
1	2	3	4	МУЖ	05/05/55	185	90	505	0	0	1	0	1	0	0	1	0
1	2	3	4	ЖЕН	07/07/57	165	58	404	0	1	0	0	1	0	1	0	0

Фрагмент ДФБД "Служба знакомства"

Таблица 3

		Основные характеристики					Управление транспортными средствами					
		Технич. хар-ки		Экономич. хар-ки		основные материалы	Гидродинамика		Средства водителя/пилота			погодные условия
		мощность	скорость	расх. топли.	вес		обтекаемость	сопр. возд.	гидроторм. усилен.	рулев. управ.	торм. сист.	
x_1	x_2	x_3	x_4	x_5	t_1	t_2	t_3	t_4	t_5	t_6		
д												
пас. сам.	d_1	1000	1500	600	150	алум.	1,0	1,0	1,0	1,0	0,9	0,8
самолеты	d_2	1600	900	500	180	алум.	0,9	1,0	1,0	1,0	0,9	0,6
груз. сам.	d_3	70	90	10	1	сталь	0,4	0,2	0,8	0,9	1,0	0,4
авто-пас. авт.	d_4	100	60	20	3	сталь	0,1	0,1	0,9	0,8	0,9	0,3
мо-груз авт.	d_5	25	70	7	0,2	сталь	0,3	0,3	0,1	0,6	1,0	0,4
мотоциклы												

Фрагмент ДФБД по проблемам управления транспортными средствами

КОЗЛОВА МАРИЯ ИВАНОВНА

ПОЛ = ЖЕН

ДАТА РОЖДЕНИЯ = 07/07/57

РОСТ = 165

ВЕС = 58

ХАРАКТЕР = ВЕСЕЛЫЙ

ТЕЛЕФОН = 404

Более наглядно возможности предлагаемой модели можно проиллюстрировать на примере вымышленной ДФБД по проблемам управления транспортными средствами, фрагмент которой представлен в таблице 3. Основное смысловое содержание документов, описывающих те или иные транспортные средства, выражается с помощью тезауруса дескрипторов, смысловыразительные значения которых могут находиться в диапазоне от 0 до 1. Можно представить себе запрос на поиск информации в такой ДФБД, например, в виде

FIND САМОЛЕТЫ AND ТЕХН-ХАР(МОЩН < 9000, СКОР > 800) AND ЭКОНОМ-ХАР AND ГИДРОДИНАМ(ОБТЕК > 0,8) AND СР-ВОДИТ(ТОРМ-СИСТ > 0,5); .

В результате такого запроса будут получены технические и экономические характеристики самолетов, мощность которых меньше 9000, скорость больше 800, а дескрипторы ОБТЕКАЕМОСТЬ и ТОРМОЗНАЯ СИСТЕМА несут в соответствующих документах смысловую нагрузку, большую, чем 0,8 и 0,5 соответственно.

Предлагаемая модель данных может найти применение для создания документально-фактографических АИС, сочетающих в себе достоинства СУБД и информационно-поисковых систем.

Литература

1. Donald R. Deutsch, Jesse M. Drapper. Data Motels: Keys to Understanding Data Base Management Systems, Advances in Data Base Management, Vol.2, 1984, 1-22.
2. Salton G., McGill J. Introduction to Modern Information Retrieval, McGraw-Hill, Book Company, 1983.
3. Salton G. Dynamic Information and Library Processing, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975. (Русский перевод: Солтон Дж. Динамические библиотечно-информационные системы. Мир, М., 1979).
4. Криницкий Н.А., Миронов Г.А., Фролов Г.Д. Автоматизированные информационные системы. Наука, М., 1982.

5. Шемякин Ю.И. Тезаурус в автоматизированных системах управления и обработки информации. Воениздат, М., 1974.
6. Антопольский А.Б., Болозеров В.Н., Гаврилова В.И. и др. О новом стандарте на информационно-поисковый тезаурус (ГОСТ 7.25-80). Научно-техническая информация, т. I, 1982, №8, с.25-29.
7. Информатика. Под ред. Тараканова К.В. Книга, М., 1986.
8. Цаленко М.Ш. Семантические и математические модели баз данных. Итоги науки и техники. ВИНТИ АН СССР, серия информатика, М., 1985, 9.

НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

Вы можете получить по почте перечисленные ниже книги, если они не были заказаны ранее.

Д9-82-664	Труды совещания по коллективным методам ускорения. Дубна, 1982.	3 р. 30 к.
Д3,4-82-704	Труды IV Международной школы по нейтронной физике. Дубна, 1982.	5 р. 00 к.
Д11-83-511	Труды совещания по системам и методам аналитических вычислений на ЭВМ и их применению в теоретической физике. Дубна, 1982.	2 р. 50 к.
Д7-83-644	Труды Международной школы-семинара по физике тяжелых ионов. Алушта, 1983.	6 р. 55 к.
Д2,13-83-689	Труды рабочего совещания по проблемам излучения и детектирования гравитационных волн. Дубна, 1983.	2 р. 00 к.
Д13-84-63	Труды XI Международного симпозиума по ядерной электронике. Братислава, Чехословакия, 1983.	4 р. 50 к.
Д2-84-366	Труды 7 Международного совещания по проблемам квантовой теории поля. Алушта, 1984.	4 р. 30 к.
Д1,2-84-599	Труды VII Международного семинара по проблемам физики высоких энергий. Дубна, 1984.	5 р. 50 к.
Д17-84-850	Труды III Международного симпозиума по избранным проблемам статистической механики. Дубна, 1984. /2 тома/	7 р. 75 к.
Д10,11-84-818	Труды V Международного совещания по проблемам математического моделирования, программированию и математическим методам решения физических задач. Дубна, 1983	3 р. 50 к.
	Труды IX Всесоюзного совещания по ускорителям заряженных частиц. Дубна, 1984 /2 тома/	13 р. 50 к.
Д4-85-851	Труды Международной школы по структуре ядра, Алушта, 1985.	3 р. 75 к.
Д11-85-791	Труды Международного совещания по аналитическим вычислениям на ЭВМ и их применению в теоретической физике. Дубна, 1985.	4 р.
Д13-85-793	Труды XП Международного симпозиума по ядерной электронике. Дубна, 1985.	4 р. 80 к.
Д3,4,17-86-747	Труды У Международной школы по нейтронной физике. Алушта, 1986.	4 р. 50 к.

Заказы на упомянутые книги могут быть направлены по адресу:
101000 Москва, Главпочтамт, п/я 79
Издательский отдел Объединенного института ядерных исследований

Рукопись поступила в издательский отдел
5 февраля 1987 года.

**ТЕМАТИЧЕСКИЕ КАТЕГОРИИ ПУБЛИКАЦИЙ
ОБЪЕДИНЕННОГО ИНСТИТУТА ЯДЕРНЫХ
ИССЛЕДОВАНИЙ**

Индекс	Тематика
1.	Экспериментальная физика высоких энергий
2.	Теоретическая физика высоких энергий
3.	Экспериментальная нейтронная физика
4.	Теоретическая физика низких энергий
5.	Математика
6.	Ядерная спектроскопия и радиохимия
7.	Физика тяжелых ионов
8.	Криогеника
9.	Ускорители
10.	Автоматизация обработки экспериментальных данных
11.	Вычислительная математика и техника
12.	Химия
13.	Техника физического эксперимента
14.	Исследования твердых тел и жидкостей ядерными методами
15.	Экспериментальная физика ядерных реакций при низких энергиях
16.	Дозиметрия и физика защиты
17.	Теория конденсированного состояния
18.	Использование результатов и методов фундаментальных физических исследований в смежных областях науки и техники
19.	Биофизика

Романенко А.Г., Столярский Ю.В.

P10-87-62

Документально-фактографическая модель данных

Наиболее важными типами автоматизированных информационных систем /АИС/ являются в настоящее время фактографические АИС, основанные на системах управления базами данных /СУБД/, и документальные /или информационно-поисковые/ АИС, основанные на программных системах. Реляционная, сетевая и иерархическая модели данных, используемые СУБД, не совместимы с семантическими моделями слабоструктурируемой информации, используемыми в информационно-поисковых системах. Потребность интеграции АИС обоих типов на основе создания обобщенной модели данных становится крайне актуальной. Представлена попытка разработки обобщенной модели представления документально-фактографической информации.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1987

Перевод О.С.Виноградовой

Romanenko A.G., Stolyarsky Yu.V.

P10-87-62

A Document/Fact Data Model

The most important computer-based automated information systems (AIS) today are the factual AIS based on the data base management systems (DBMS) and document or information retrieval (IR) AIS based on the program systems. Relational, network and hierarchical data models used by DBMS are not compatible with semantic models of weakly-structured information used by IR systems. The need for integrating both types of AIS on the base of creation of common data model becomes extremely actual nowadays. An attempt is presented to develop common model representing document/fact data.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1987