## § 1. БАНКИ ИНФОРМАЦИИ

### 1.1. БАНКИ ДАННЫХ

Банк данных - наиболее характерный пример информационной системы. В банке данных хранится достаточно универсальная, необходимая для решения разнообразных прикладных задач, информация об определенной предметной области в специальном представлении, чаще всего предполагающем хранение и обработку с помощью компьютеров. При этом сами данные образуют базу данных, а банк, наряду с базой, содержит программные средства обработки данных и реализации запросов, т.е. систему управления базой данных (СУБД). Как правило, банки данных являются системами коллективного пользования. К информации, хранимой в них, часто можно получить доступ по телекоммуникационным сетям.

В современном мире существует огромное число банков данных. В них содержатся сведения коммерческого характера, данные по библиотечным фондам, системам здравоохранения, транспорта и т.д. Быстро развиваются банки, содержащие сведения о системах образования - национальных, региональных (в России они называются «банки педагогической информации»).

Итак, основа банка - база данных. Определение базы данных, основные понятия, связанные с различными моделями данных, уже обсуждались в гл. 2 и это обсуждение будет продолжено ниже в этой главе.

Остановимся на классификации банков информации. Эта классификация может быть проведена с разных точек зрения. По **назначению** можно выделить следующие классы банков информации:

• *информационно-справочные* системы (общего назначения и специализированные);

• банки данных в *автоматизированных системах управления* (предприятий и организаций, технологическими процессами и т.д.);

• банки данных в *системах автоматизации научных исследований.*

Однако такая классификация является не вполне строгой и завершенной. По **режиму** **функционирования** можно рассматривать банки информации *пакетного, диалогового* и *смешанного* типов, В связи с широким распространением персональных компьютеров, локальных и глобальных сетей ЭВМ подавляющее распространение получили диалоговые системы.

По **архитектуре** вычислительной среды различают *централизованные* и *распределенные* банки информации.

К настоящему времени сложились следующие три основных типа банков информации: банки документов, банки данных и банки знаний.

Исторически первым типом банков информации явились банки документов или документальные информационно-поисковые системы. Документальные информационно-поисковые системы бурно развивались в 60-е годы, они широко используются в качестве справочного инструмента пользователей научно-технической информацией, в информационном обслуживании управленческих работников, специалистов и др. В настоящее время интерес к этим системам возобновился в связи с развитием глобальных информационных сетей *(****Internet****)* и появлением гипертекстовых серверов типа WWW, ***Gopher*** и т.д.. которые вместе с соответствующими поисковыми системами *{****Archie****,* ***Whatis*** и т.д.) можно отнести к распределенным банкам документов.

Объектом хранения в таких системах является документ (научная статья, монография, приказ, циркуляр, письмо и т.д.) или факты, извлеченные из документов. Для обеспечения поиска и доступа к таким документам необходима их предварительная семантическая обработка - индексация. Индексация до настоящего времени остается неавтоматизированной процедурой и выполняется специалистами -людьми, индексирующими документы и запросы.

### 1.2. БАНКИ ДОКУМЕНТОВ

В документальной информационно-поисковой системе выделяют следующие компоненты:

• массив документов (текстов) или фактов, выступающих в качестве объектов хранения и поиска;

• информационно-поисковый язык, предназначенный для отображения содержания документов и операций над ними, в том числе и запросов для поиска документов;

• правила, алгоритмы, методы индексирования и поиска документов, позволяющие описывать документы и операция над ними на информационно-поисковом языке;

• комплекс программных и аппаратных средств, с помощью которых реализуются процессы накопления, хранения и поиска документов;

• обслуживающий персонал, включающий администраторабанка документов, системных аналитиков, программистов и индексаторов. Банки документов работают обычно в двух режимах:

1) избирательного распределения информации, обеспечивающего регулярное текущее информирование пользователей банка о новых поступлениях документов;

2) ретроспективного поиска информации по разовым запросам во всем массиве документов.

Важнейший этап обработки нового документа при поступлении его в хранилище документальной информационно-поисковой системы - индексирование документа -слагается из следующих действий:

1) выявления основного смыслового содержания документа (с учетом точки зрения автора документа и информационных потребностей пользователя системы);

2) описания смыслового содержания документа на информационно-поисковом языке (ИПЯ) и получения соответствующего поискового образа документа (ПОД).

При выполнении ретроспективного поиска производится

1) выявление смыслового содержания запроса;

2) получение поискового обзора запроса (ПОЗ) на информационно-поисковом языке системы.

Результатами индексирования документов и запросов являются их поисковые образы (ПОД и ПОЗ). Поиск документов по запросу означает сопоставление ПОД и ПОЗ. Качество поиска зависит от критериев смыслового соответствия документа запросу (критериев выдачи).

Различают

• теоретико-множественные критерии, основанные на оценкестепени совпадениялексических единиц ИПЯ (слов), используемых в ПОД и в ПОЗ;

• критерии, учитывающие с помощью «весовых коэффициентов» относительную информационную значимость отдельных лексических единиц, входящих в ПОЗ;

• логические критерии, основанныена использовании логических операций (дизъюнкции, конъюнкции, импликации и др.).

Информационно-поисковые языки (ИПЯ), которые используются в настоящее время, можно разделить на три большие группы:

• классификационные языки;

•дескрипторные;

• комбинированные.

Языки классификационного типа, в свою очередь, делятся по структуре:

• ИПЯ иерархической структуры;

• ИПЯ фасетной структуры;

• эмпирические (неиерархические) языки.

**Классификационные системы.** В *иерархических* классификационных системах лексические единицы (термины) находятся между собой в отношениях включения. При записи они располагаются в порядке постепенного перехода от общих к более частным. Существуют иерархические системы, в которых рубрики включают по 20 и более подклассов, рубрик и подрубрик в нисходящем порядке. Примером такой иерархической классификационной системы является универсальная десятичная классификация (УДК), широко используемая в библиотечном деле и документальных поисковых системах. УДК охватывает весь спектр знаний.

Шифры УДК, которые можно увидеть на обороте титульного листа всех книг, перед заголовками статей в журналах и сборниках, имеют более чем столетнюю историю. В 1905 г. в Брюсселе на французском языке вышло первое сводное издание таблиц десятичной классификации. Эти таблицы были созданы на основе таблиц десятичной классификации Мельвиля Дьюи, впервые изданных в 1876 г.

Каждый класс (первая ступень деления) в УДК содержит группу более или менее близких наук, например, класс 5 - математику и естественные науки, класс 6 -прикладные науки: технику, в том числе информатику, сельское хозяйство, медицину. Каждая последующая присоединяемая цифра не меняет значения предыдущих, а лишь уточняет их, обозначая более частное понятие. УДК настоящего текста: 681.3:62-52.

В основе *фасетной* классификации лежит многоаспектное распределение понятий какой-либо отрасли науки или техники по однородным взаимно исключающим друг друга фасетам.

Примером *эмпирической* (неиерархической) классификации может быть алфавитно-предметная классификация. Словарный запас такой классификации состоит из упорядоченного по алфавиту множества слов, словосочетаний и фраз естественного языка, обозначающих предметы какой-либо отрасли науки или практической деятельности. Каждому предмету или вопросу при этом отводится только один индекс, собирающий всю информацию относительно данного предмета или вопроса независимо от аспекта рассмотрения. В предметных классификациях используются следующие термины:

• предметная рубрика (заголовок) - слово или фраза естественного языка, используемая для обозначения основного предмета документа (или запроса);

• предметный подзаголовок - слово или фраза, обозначающая аспект рассмотрения предмета, указанного в предметном заголовке или в подклассе предметов, входящих в класс, обозначенный предметным заголовком;

•предметный словник - упорядоченное по алфавиту множество предметных заголовков, используемых для построения какого-либо каталога или указателя.

Система предметных заголовков и подзаголовков и более мелких разделов создает сложную предметную классификацию по аспектам рассмотрения предмета, т.е. имеет некоторые черты фасетных классификаций.

**Дескрипторные информационно-поисковые языки.** Дескрипторные информационно-поисковые языки основаны на методе координатного индексирования, сущность которого сводится к тому, что смысловое содержание документа может быть с достаточной точностью и полнотой выражено списком ключевых слов, содержащихся в тексте.

Ключевое слово - это лексическая единица информационно-поискового языка, являющаяся существительным, прилагательным, глаголом, наречием, числительным или местоимением естественного языка или словосочетанием. Основной критерий отбора ключевых слов из текста - степеньих полезности для индексирования документа или запроса.

Координатное индексирование выполняется с помощью ключевых слов и логических операции конъюнкции и дизъюнкции. Близкие по смыслу ключевые слова образуют классы условной эквивалентности, имена которых также являются единицами поискового языка и называются дескрипторами. Дескрипторы вместе с набором ключевых слов языка и семантических связей между ними образуют тезаурус - систематизированный набор данных об области знания, позволяющий в ней ориентироваться.

В дескрипторной статье тезауруса обычно устанавливаются следующие виды семантических отношений:

• отношение синонимии;

• отношение подчинения;

• отношение ассоциации.

Такие связи служат для увеличения полноты представления содержания документов и позволяют формировать запрос не обязательно в ключевых словах документа.

Пример дескрипторной статьи:

языки алгоритмические синонимы: алгоритмические языки

машинно-ориентированные языки проблемно-ориентированные языки вышестоящие: программное обеспечение

языки формальные нижестоящие: автокоды

алгол паскаль фортран си бейсик ассоциации: алгоритмы

программирование

Целесообразность применения того или иного языка во многом зависит от назначения информационной системы, степени ее автоматизации. Для описания документов в библиотеках, общих и технических архивах обычно применяют классификационные языки. В автоматизированных информационно-поисковых системах используются, главным образом, дескрипторные языки.

### 1.3. БАНК ПЕДАГОГИЧЕСКОЙ ИНФОРМАЦИИ

В качестве примера реально существующего и достаточно широко используемого банка данных рассмотрим банк педагогической информации (БПИ), созданный в Республиканском институте повышения квалификации работников образования под руководством В.И.Журавлева. На основе этого банка в ряде регионов России созданы и успешно функционируют региональные банки педагогической информации.

БПИ предназначен для хранения всех видов информации, циркулирующей в системе образования страны, и удовлетворения информационных потребностей пользователей.

С банком работают лица следующих категорий:

• работники образования, получающие с разными целями информацию из банка (к ним относятся учителя, методисты, работники управленческих структур всех уровней, ученые педагоги, преподаватели вузов, студенты и учащиеся, родители и др.);

• поставщики информации, т.е. авторы учебной, учебно-методической и научно-методической литературы, разработчики нормативно-правовой документации, относящейся к системе образования и т.д.;

•работники информационной системы, функция которых состоит в накоплении информационного фонда вторичных документов - информационных модулей, разработанных в соответствии с концепцией БПИ (информационные (информационно-педагогические) модули готовятся на основе первичных материалов, представленных поставщиками информации; такая работа требует специальной подготовки);

• непосредственные работники информационной службы, ведущие лингвистическую часть информационной системы: администратор банка, эксперты.

В перспективе в нашей стране должна быть создана единая сеть банков педагогической информации, объединяющая как центральный республиканский банк, так и региональные БПИ. В последних наряду с информацией общего назначения сосредоточивается информация регионального характера, отражающая опыт учителей и методистов региона, деятельность региональных учреждений повышения квалификации и местных вузов, связанную с региональной системой образования, региональных органов управления образованием и другую информацию педагогической направленности. Принципиально важной частью такой сети является такая система обмена данными между различными банками, при которой пользователь любого из них может получать информацию из любого банка и отправлять ее туда.

**Содержание и структура банка.** Охарактеризуем содержание БПИ приведя перечень и краткое описание нескольких его разделов.

*Учебная литература, авторские курсы.* Вариативность отечественного образования - характерное его свойство, возникшее в последнее десятилетие. Почти по всем дисциплинам есть несколько вариантов изложения, подкрепленных учебниками, сборниками упражнений, тестов и т.д. Учитель не может иметь все эти материалы, да в этом и нет необходимости, так как он обычно работает по одному из вариантов. Однако, время от времени возникает потребность ознакомиться и с другими вариантами, что можно сделать с помощью банка.

Особенно эта проблема актуальна при необходимости ознакомиться с разработками по специальным (авторским) курсам, к которым зачастую практически нет другого доступа.

*Методы обучения.* В конце XX века в образовании во всем мире активизировались поиски новых форм и методов обучения. Это связано со стремительными переменами в общественном устройстве (и, в частности, с информационной революцией). Ежегодно публикуется множество материалов на эту тему, аккумулируемых в БПИ.

*Аспекты педагогической науки.* Новые разделы теории воспитания и описание практических приемов реализации педагогических концепций - важная часть педагогической информации.

*Диагностика педагогического профессионализма.* Проблемы диагностики успехов обучения были актуальны всегда, но стали еще более актуальными в связи с внедрением государственных образовательных стандартов. Все чаще используются тестовые методики контроля, сравнительно новые для нашей школы; на эту тему постоянно возникают новые материалы.

*Зарубежная педагогическая информация.* Без знания происходящего в этой сфере в мире невозможно правильно ориентироваться в проблемах образования собственной страны. Прямой доступ к зарубежным материалам для большей части работников отечественной системы образования затруднен по финансовым, языковым, терминологическим и прочим причинам. Аккумулирование в банке обзорных (и оригинальных) материалов позволяет существенно продвинуться в решении этой проблемы.

*Инновации в образовании.* Многочисленные находки учителей, методистов, работников органов управления часто остаются невостребованными просто из-за невозможности ознакомления с ними широкой педагогической общественности. Даже по отношению к широко известным педагогам-новаторам существуют проблемы доступа к конкретным материалам.

*Историко-педагогическое наследие,* как отечественное, так и зарубежное, весьма велико, и соответствующие материалы необходимо иметь в БПИ.

*Повышение квалификации работников образования -* необходимое условие функционирования системы образования. Наряду с давно действующими в этой сфере институтами повышения квалификации, указанной деятельностью занимаются вузы, коммерческие организации и др. Информация на эту тему в БПИ необходима и поможет сделать правильный выбор.

*Опыт управленческой работы* для системы образования не менее важен чем чисто педагогический опыт. Приемы организации управления учебными заведениями, учебно-воспитательным процессом систематизируются в БПИ.

*Правовая основа педагогической деятельности* важна для каждого участника педагогического процесса - учителя, ученика, родителей, администраторов. В системе образования существует большое количество нормативно-правовых актов, накапливаемых в БПИ.

В БПИ содержатся также сведения о рынке образовательных услуг, неформальных объединениях молодежи и других аспектах деятельности системы образования.

Как и во всяком «электронном» банке информации, является важным вопрос о структурных единицах хранения информации. Основной такой единицей является информационно-педагогический модуль (ИПМ). ИПМ - формализованное отражение информации, обладающее следующими свойствами:

• логической законченностью (изложенная в одном ИПМ информация соответствует одному и только одному типу);

• информативностью (объем и качество информации в ИПМ достаточны для ее использования в науке или практике без обращения к источнику);

• популярностью изложения (язык изложения доступен педагогу-практику);

• краткостью изложения (объем ИПМ не превышает 5 кбайт).

ИПМ записываются в специальных форматах. Структура и форма записи в них выбрана таким образом, чтобы информацию мог формализовать не только специально подготовленный человек, но и сам автор, что оправдано экономически и способствует формированию информационной культуры работников образования. Поскольку оформление содержательной (т.е. основной) части информации требует знаний в достаточно узких предметных подобластях образования, никто лучше автора этого сделать не может; администратор же банка (или методист) могут при необходимости оказать автору помощь в некоторых формальных моментах. В то же время, создание ИПМ по источникам, авторы которых не доступны для прямого общения (например, когда ИПМ создается по статьям в зарубежных журналах), могут осуществлять специалисты банка - референты.

Отметим, что подобная практика существует и в научных реферативных журналах, куда может быть помещен либо реферат статьи, написанный референтом, либо автореферат, написанный автором статьи.

ИПМ представляет собой текстовый файл, состоящий из двух полей: поля поискового образа документа и поля записи данных. Каждое из полей распадается на подполя - элементы данных. Рассмотрим структуру и содержание записи отдельно для каждого поля.

*Поле поискового образа документа.* В этом поле записываются элементы данных, необходимые для поиска информации, ее классификации, анализа наличия или отсутствия, экспертизы по различным параметрам, осуществления коммуникаций как внутри системы, так и с другими автоматизированными информационными системами. Поскольку параметры информации в таких системах жестко стандартизированы, а эти стандарты (ГОСТы) недостаточно полно отражают специфику педагогических исследований, то допускается описание элементов данных, не имеющих аналога в ГОСТах; они помечаются символом «звездочка» (\*). Элементы данных, запись в которых (или ее часть) может повторяться в других подполях, помечены символом (1). Максимальная длина записи в элементах данного поля составляет 80 байт (длина строки экрана). В тех случаях, когда элемент данных имеет другое ограничение, оно указано в конце строки в скобках.

Структура записи элементов данных:

1) название рубрикатора (\*);

2) указатель десятичной классификации (УДК);

3) название тезауруса (\*);

4) информационное поле (\*);

5) фасет (\*);

6) дескриптор тезауруса (1);

7) дата ввода записи;

8) организация - создатель записи;

9) автор, авторский коллектив (\*)(1);

10) основное заглавие (250);

11) место работы автора (страна);

12) место работы автора (область);

13) составитель текста записи (1);

14) редактор текста записи (1);

15) переводчик (1);

16) тип информации (\*);

17) ключевые слова (через запятую) (1);

18) вид исходного документа;

19) дополнительные данные об исходном документе;

20) библиографическая ссылка (250)(\*)(1).

Для формирования пунктов 1-3 необходимо пользоваться справочниками по УДК, пунктов 18 и 19 - таблицами кодов «вид документа» и «код дополнительных данных», имеющимися в документации к банку.

Поясним использованное выше понятие «фасет». Поскольку информационное поле охватывает очень широкий объем понятий, то используется фасетный метод классификации: термины внутри поля располагаются по фасетам - небольшим параллельным группам. Внутри фасеты - дескрипторы, т.е. еще более мелкое дробление. Все это делается для организации эффективного и удобного для пользователя и для системы поиска информации - по классу, по автору, по региону, по ключевым словам. Детальное описание фасетированного рубрикатора находится в документации банка и частично заведено в оболочку пользователя.

*Поле записи данных.* В этом поле записывается содержательная часть информации, заносимой в ИПМ. В случае, если информация распадается на ряд единичных типов информации (т.е. образует информационный комплекс), поле записи данных распадается на отдельные подполя, в каждом из которых записывается отдельный ИПМ со своим названием. Записи ИПМ в этом случае должны образовывать логическую последовательность в дедуктивном порядке. Например, уместно рассматривать данный учебник как совокупность глав по относительно самостоятельным разделам информатики и при составлении соответствующего ИПМ в общее поле записи данных поместить ИПМ по каждой из глав.

**Пример информационно-педагогического** **модуля**. В качестве прообраза, для которого создается ИПМ, рассматривается данный учебник.

*Поле поискового образа.*

1 Народное образование. Информатика.

2 681.3

3 Тезаурус ЮНЕСКО-МБП по образованию

4 Обучение

5 Обучение

6 Предметное обучение

708.02.1999

8 Авторскийколлектив

9 Авторский коллектив в составе: А.В.Могилев, Н.И.Пак, Е.К.Хеннер

10 Информатика

11 Российская Федерация

12 Воронежская область, Красноярский край, Пермская область

13 Е.К.Хеннер

14 И. П. Сидоров

15

16 Учебник

17 Информатика, теоретические основы информатики, программное обеспечение, языки программирования, архитектура ЭВМ, компьютерные сети, информационные системы, компьютерное моделирование

18 Рукопись

19 Учебное пособие для подготовки бакалавров по профилю «Информатика» иучителей информатики

20 Информатика. / Под ред. Е.К.Хеннера. - М.: Издательский Центр «Академия», 1999

*Поле записи данных*

«Учебное пособие «Информатика» предназначено для подготовки бакалавров по направлению 540106 «Естествознание», профилю «Информатика», магистров по программе «Информатика в образовании» и учителей информатики (специальность 030100 «Информатика»), а также подготовки по педагогическим специальностям, для которых предусмотрено получение дополнительной квалификации «Учитель информатики». Пособие соответствует государственным стандартам по той части предметных блоков, которая касается информатики и вычислительной техники. В пособие включены материалы по следующим разделам: «Теоретические основы информатики», «Программное обеспечение ЭВМ», «Языки и методы программирования», «Вычислительная техника», «Компьютерные сети и телекоммуникации», «Информационные системы», «Компьютерное моделирование». Материал, включенный в пособие, может обеспечить пяти -шести семестровый курс информатики, изучаемый в педагогических вузах при профильной подготовке по информатике бакалавров, магистров и специалистов -учителей информатики. Пособие может также быть полезным для студентов различных вузов, обучающихся по специальностям, связанным с информатикой. Кроме того, пособие можно использовать при повышении квалификации и переподготовке учителей».

**Поиск** **информации в банке**. Здесь речь пойдет лишь о внешней стороне поиска, т.е. его пользовательском интерфейсе. Скрытые от пользователя механизмы поиска обсуждались в гл. 3 (раздел «сортировка и поиск»).

Работа пользователя с банком начинается с меню, изображенном в верхней строчке рис. 6.1.



*Рис. 6.1.* Вид меню БПИ в режиме «Просмотр»

В режиме «Пользователь» доступны только два поля меню: «Просмотр» и «Выход». Выбрав пункт меню «Просмотр», получаем подпункты, изображенные на рисунке 6.1. Таким образом, любой материал, имеющийся в банке, можно найти тремя различными путями.

Путь «По атрибутам»открывает наиболее широкие возможности поиска информации, отраженные на рисунке 6.2.



*Рис. 6.2.* Поиск информации по атрибутам

При поиске по Тезаурусу имеется семь полей, на которые поделено все образовательное пространство:

• вещи (педагогические средства);

• контекст (общество и образование);

• люди (учащиеся и учителя);

• обучение;

• развитие и учение;

• содержание образования и воспитания;

• управление и исследования.

Поиск по фасетам и дескрипторам может происходить: по классу информации, по автору, по региону, по ключевым словам.

Аналогично происходит поиск информации и «По дополнительным реквизитам». При выборе такого пути возникает меню, в котором можно выбирать пункты «Организация», «Составитель», «Редактор», «Дополнительные сведения», «Дата» и продолжать поиск.

Третий путь организации поиска - «По каталогам». Ступив на него, получаем приведенное на рис. 6.3 меню. Выбрав нужный каталог, можем попасть в подкаталоги и т.д., пока не найдем нужного ИПМ или не убедимся в его отсутствии.



*Рис. 6.3.* Поиск информации по каталогам

**Функции** **администратора банка**. Здесь мы опишем лишь те функции администратора банка, которые обусловлены программной оболочкой. Они отражены в следующем меню:



По существу операции означают следующее:

• просмотр и корректировка информационных материалов;

• дополнительный ввод новой информации;

• работа со справочниками (корректировка, доввод);

• создание новой базы с выборкой материалов;

• изменение пароля;

• коррекцию справочников.

После вхождения в пункт меню «Доввод» на экране появляется таблица (рис. 6.4), в которой следует последовательно заполнить позиции; результаты отражаются в колонке «Результаты ввода».

При этом система позволяет воспользоваться рядом встроенных в нее справочников, а также корректировать сами справочники.

Доввод новой информации по позициям заканчивается вводом текста, который должен быть заранее подготовлен в текстовом редакторе (рис. 6.5); при этом необходимо указать диск, на котором находится нужный текст, каталог и название файла.

Данная оболочка позволяет автоматически (раздел меню ***export****)* создавать новую базу с выбранными материалами или производить обмен текстами между аналогичными базами. Детальное выполнение этих манипуляций можно произвести следуя указаниям последовательно появляющихся меню.

После описанных выше операций необходимо произвести корректировку базы с помощью операции «Чистка». Эта чисто техническая операция описана в документации.

Для предотвращения недозволенного выполнения пользователями функций по управлению банком операции, доступные лишь администратору, защищены паролем, который следует периодически обновлять.



*Рис. 6.4.* Операции в меню «Доввод»



*Рис. 6.5.* Меню для доввода текста