

ИНФОРМАЦИОННЫЕ СИСТЕМЫ

УДК 002.53:681.3.016

Д. В. Ландэ

ОРГАНИЗАЦИЯ ТЕЛЕДОСТУПА К БАЗАМ НАУЧНО-ТЕХНИЧЕСКИХ ДАННЫХ В СЕТИ ПЕРСОНАЛЬНЫХ КОМПЬЮТЕРОВ

Определены основные компоненты архитектуры информационно-поисковой системы (ИПС), предложена структура базы данных в среде реляционных СУБД. Описана реализация приведенной архитектуры, а также технология подключения подсистемы пользователя ИПС к стандартным средствам телекоммуникации, базирующимся на сетях персональных ЭВМ и коммутируемых линиях связи.

Базы научно-технических данных — важная составляющая потенциала государства, и организация доступа к ним — главная задача, решаемая подразделениями национальной системы НТИ. Вопрос организации доступа к информации на хост-ЭВМ типа ЕС, при смене технической парадигмы — переход к современным многопроцессорным системам и хост-станциям на базе локальных сетей требует пересмотра.

Основным инструментом, обеспечивающим доступ к базам данных, является информационно-поисковая система (ИПС). При рассмотрении архитектуры ИПС (Retrieval System Architecture — RSA) можно выделить следующие компоненты:

- 1) структура базы данных ИПС (Retrieval Systems Data Base — RSDB);
- 2) язык запросов ИПС (Retrieval Systems Query Language — RSQL);
- 3) интерфейс пользователя ИПС (Retrieval Systems User Access — RSUA);
- 4) средства коммуникации (Retrieval Systems Communication — RSC). Наиболее важные и наименее проработанные до сих пор компоненты — это RSDB и RSQL.

Интерфейс пользователя ИПС должен быть создан на основе CUA, а средства коммуникации на основе CCS — всемирно признанных стандартов фирмы IBM [1, 2].

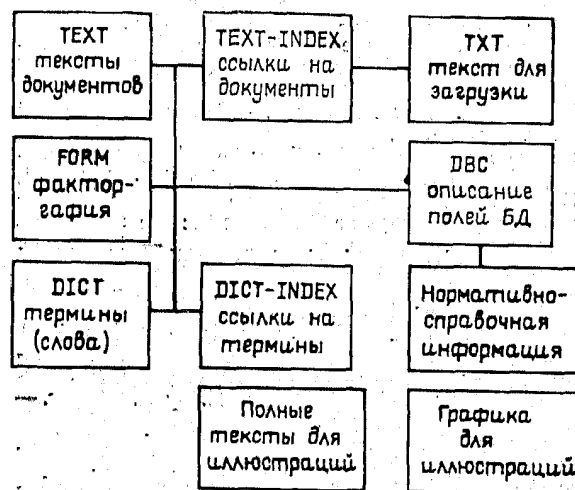
В качестве стандарта ИПС, длительное время (около 20 лет), признаваемого «де факто» в сфере НТИ рассматривалась архитектура системы STAIRS (Storage and Information Retrieval System), многие решения которой в настоящее время морально устарели и требуют улучшения. Это относится не только к интерфейсу пользователя, но и к способам загрузки и представления данных.

Следует учитывать, что архитектура базы данных ИПС STAIRS разрабатывалась без ориентации на СУБД и, следовательно, без учета свойств реляционности.

Структура, во многом сходная со структурой системы STAIRS, разработанная для реализации в среде СУБД [3, 4], может быть задана в виде файловой системы: файла описания базы данных и трех основных информационных файлов — текстового, словарного, фактографического и индексов к ним (рисунок).

Текстовый файл базы данных содержит тексты документов, примерно в том виде, в котором они отобра-

жаются на экране. Особенность представления документов в текстовом файле заключается в том, что каждый документ является логической конкатенацией некоторой совокупности «строк», выступающих в файле реляционной СУБД отдельными записями.



Структура базы данных информационно-поисковой системы в среде реляционной СУБД

Основное назначение словарного файла БД ИПС состоит в обеспечении быстрого поиска информации, представленной в текстовом файле. Он содержит термины, входящие в список полей, отмеченных пользователем при индексировании. В качестве терминов могут выступать как отдельные слова естественного языка, так и классификационные признаки, например, рубрики Рубрикатора ГАСНТИ. Словарный файл базы данных может быть использован для анализа лексики, входящей в базу научно-технических и экономических данных, кодов определенных диапазонов, частоты встречаемости определенных значений и т. п. На работу с этим файлом должны быть ориентированы процедуры поиска и получения статистики.

Фактографический файл содержит записи с фактографическими полями в файле базовой СУБД, на структуру которого не накладывается никаких ограничений.

В настоящее время разработано два пакета прикладных программ, реализующих приведенную архитектуру, современный интерфейс, функционально полный язык запросов, средства обучения и подсказки, для работы в локальной сети. Вместе с тем, два недостатка ограничивают сферу применения этих пакетов: практическая невозможность работы с базами данных в глобальных сетях в режиме on-line и относительная сложность освоения.

Наряду с этим, существует ряд программных средств, обеспечивающих дистанционный доступ к разнородной информации, обычно структурированной и характеризующейся небольшим объемом [5, 6]. К таким программам прежде всего относятся средства обеспечения электронных досок объявлений (BBS). В BBS имеется возможность подключения прикладных программ, в которых должна быть обеспечена возможность телеобработки. В качестве прикладной программы, в которой реализована идея гипертекста, работающей в среде BBS FrontDoor, можно назвать, например, Электронную библиотеку [6].

Вопрос же работы с базами данных в среде BBS до настоящего времени был открыт. Попытка реализации доступа к БД в режиме on-line в среде BBS может рассматриваться как принципиально новый подход к обеспечению доступа к БД в условиях децентрализованного информационного ресурса.

Для реализации эффективного доступа к БД, создаваемых на различных, в том числе и терминальных, уровнях государственной сети НТИ, информационно-поисковая система должна обладать двумя качествами — простотой и функциональностью. Первое качество должно минимизировать издержки при создании БД прямого доступа (сюда относятся затраты на обучение персонала, время создания и реорганизации БД и т. д.). Функциональность должна базироваться на возможностях работы в режиме on-line и на реализации функционально полной и логически несложной формы запросов, а также командного языка ИПС, позволяющего управлять процессом поиска и выдачи информации.

Для организации работы в глобальных сетях в режиме on-line с целью обеспечения возможности работы пользователей, обладающих самым разнообразным оборудованием (начиная от модемов V21 до работы с протоколом V32, MNP5) и любым телекоммуникационным пакетом (MTEZ, Comit, Telix, FD и др.), на сегодня имеет смысл отказаться от таких возможностей, как обработка графики и многооконных «меню». Можно также упростить язык запросов, чтобы даже неподготовленный пользователь мог в простой форме выражать свои информационные потребности.

На основе приведенной структуры была разработана информационно-поисковая система, функционально разделенная на две части — подсистему администрации БД и подсистему (программу) поиска. Процедура создания БД ИПС RePri вызывается путем запуска программы CRDB с параметром — именем создаваемой базы данных. Программа CRDB является меню-ориентированной оболочкой, позволяющей выполнять такие виды работ: инициализацию БД, определение полей БД, загрузку БД, индексирование (сортировку словаря). Загрузка БД RePri производится из внешнего ASCII файла, в формате, принятом в системе STAIRS и имеющем примерно следующий вид:

```
* * *DOCUMENT-1*
004 81033039000012204797337
007 МСПИ
008 W0900405A1 9003680
043 9003680
082 502
```

ISSN 0548—0027 . НТИ . СЕР. 2 . ИНФОРМ. ПРОЦЕССЫ И СИСТЕМЫ . 1994 . № 7

085 WIPO

086 WO

094 A1

139 Способ и устройство для управления преобразователем мощности

322 4H 02H 3/00 A; 4G 05F 1/46 B

905 УСТРОЙСТВО ИСПОЛЬЗУЕТСЯ ДЛЯ УПРАВЛЕНИЯ СТАТИЧЕСКИМ ПРЕОБРАЗОВАТЕЛЕМ МОЩНОСТИ 34 С ЦЕЛЮ ВЗАИМНОГО ОБМЕНА МОЩНОСТИ МЕЖДУ ПЕРВЫМ И ВТОРЫМ ИСТОЧНИКАМИ МОЩНОСТИ 30, 32. ПРЕДПОЧТИТЕЛЬНО ЕГО ПРИМЕНЕНИЕ, ЕСЛИ ОДИН ИЗ ИСТОЧНИКОВ МОЩНОСТИ ЯВЛЯЕТСЯ МОЩНЫМ ИСТОЧНИКОМ ПОСТОЯННОГО ТОКА 32, ТАКИМ КАК АККУМУЛЯТОРНЫЕ БАТАРЕИ, А ВТОРОЙ ИСТОЧНИК МОЩНОСТИ ЯВЛЯЕТСЯ СТАНДАРТНОЙ УНИВЕРСАЛЬНОЙ ШИНОЙ 30. УСТРОЙСТВО ИМЕЕТ КОНТРОЛЛЕР, РАБОТАЮЩИЙ В ЗАВИСИМОСТИ ОТ НАПРЯЖЕНИЯ НА ШИНЕ. СИГНАЛЫ КОМАНД УСТАНОВЛИВАЮТ ТРЕБУЕМЫЙ ВЗАИМНЫЙ ОБМЕН МОЩНОСТИ, А ТАКЖЕ ВЗАИМНЫЙ ОБМЕН РЕАКТИВНОЙ МОЩНОСТИ. ЦЕПЬ ОБРАТНОЙ СВЯЗИ В ЗАВИСИМОСТИ ОТ ВЗАИМНОГО ОБМЕНА ФАКТИЧЕСКИ ИМЕЮЩЕЙСЯ МОЩНОСТИ УСТАНОВЛИВАЕТ ВЗАИМНЫЙ ОБМЕН РЕАКТИВНОЙ МОЩНОСТИ ДЛЯ ГЕНЕРИРОВАНИЯ СИГНАЛОВ, УПРАВЛЯЮЩИХ СТАТИЧЕСКИМ ПРЕОБРАЗОВАТЕЛЕМ С ЦЕЛЮ ОБЕСПЕЧЕНИЯ ТРЕБУЕМОГО ВЗАИМНОГО ОБМЕНА МОЩНОСТИ.

* * * DOCUMENT-2

...

При этом один документ отделяется от другого специальной строкой (в данном случае начинающейся символами ***), каждому полю документа предшествует трехсимвольная числовая метка, позиция начала текста в каждой строке фиксирована, а длина текста в каждой строке не должна превышать 69 символов.

В систему RePri органически входит утилита EXPDBF, обеспечивающая экспорт данных из файлов системы dBASE в загрузочный формат для ИПС RePri. Эта утилита обеспечивает то, что подготовка данных для ИПС может производиться любыми имеющимися средствами, позволяющими создавать и модифицировать файлы типа dbf. К таким средствам могут относиться, например, утилита DBU системы Clipper, интерактивные среды СУБД dBASE, FoxBASE, FoxPro [7, 8, 9]. Формат входных данных для загрузки БД RePri аналогичен формату для загрузки в ИПС STAIRS и Диалог/М. В связи с этим существует отработанная технология конвертирования из обменных форматов (МЕКОФ, ISO 2709, MARC) в загрузочный формат АИПС RePri.

Процедура поиска запускается указанием имени программы поиска POISK с параметром — именем порта, через который осуществляется диалог с пользователем (например COM1 или COM2). Процедура поиска вызывается из среды стандартного телекоммуникационного программного обеспечения (например, FrontDoor).

Впервые упрощенная реализация структуры БД RePri (архитектуры RSDB) была отработана в АИПС «микроАСНТИ» (ЦАНТИ Минобразования Украины). Широкие возможности (в плане визуализации данных и сервиса) были предоставлены следующей реализацией — АИПС «ИНФАНТ», разработанной совместно специалистами УкрИНТЭИ и МП «АИСТ». Некоторая громоздкость этих систем, относительно большая сложность в их освоении (как следствие широких функциональных возможностей) и практическая невозможность их использования для работы в режиме диалога с удаленными пользователями на отечественных линиях связи привела к необходимости реализации варианта си-

стемы на языках программирования более низкого уровня, чем языки СУБД. Был создан набор инструментальных средств для работы с dbf-структурами, ориентированными на задачу построения АИПС.

В результате была разработана функционально полная и компактная система (360 Кбайт одна дискета, на которой размещаются исходные и загрузочные тексты программ, сервисные подсистемы и контрольный пример). Быстродействие подсистемы администратора RePgi (что важно при обработке больших объемов данных) в несколько раз больше, чем у ее предшественников.

Однопользовательская локальная эксплуатация базы данных RePgi реализуется путем запуска программы с параметром CON (консоль). Для обеспечения работы одного удаленного абонента необходим Hayes-модем, присоединенный к порту персональной ЭВМ, работающей под управлением MS-DOS и стандартной телекоммуникационной программы типа FrontDoor. Для этого при вызове программы поиска в качестве параметра указывается имя порта, например, COM2. Работу с несколькими модемами можно организовать на одном компьютере с помощью многозадачной ОС, например, DesqView. В этом случае каждая из ее подзадач управляет своим портом. Не представляет трудностей организация многопользовательской дистанционной работы при наличии локальной сети типа EtherNet. В этом случае БД должна размещаться на файл-сервере, доступ к которому возможен с рабочих мест, оснащенных модемами, обеспечивающими работу с удаленными абонентами.

СПИСОК ЛИТЕРАТУРЫ

- 1 Архитектура среды для разработки приложений.— Киев: «Крешатик», 1992,— 220 с.
- 2 Проектирование пользовательского интерфейса на персональных компьютерах Стандарт фирмы IBM. / Под ред. М. Дадашова.— Вильнюс: «DBS LTD», 1992,— 190 с.

3. Ландэ Д. В. Подход к формированию персональных баз данных научно-технической информации // Автоматизированные системы массового распространения компьютерной информации: Сборник ИПРИ АН Украины.— Киев: ИПРИ— 1991— С. 102—106.
4. Сороко В. Н., Ландэ Д. В. Использование технологии обработки документальных баз данных для организации обучающих режимов // Вестник КПИ. Выпуск 15.— Киев: «Лыбидь».— 1991.— С. 6—11.
5. Ландэ Д. В. Проблемы организации пользовательской информационной сети как составной части национальной системы НТИ Украины // Развитие национальных систем НТИ стран СНГ, центральной и восточной Европы в новых общественно-политических и социально-экономических условиях: Тезисы докл. и сообщений международной научно-практической конференции. 19—20 мая 1993 г., Киев, УкрИНТЭИ, ч. 2, стр. 15—16
6. А. Я. Аноприенко, А. А. Кухтин, С. С. Баданюк, В. А. Шевчук. Электронная библиотека как средство информационного сервиса // Развитие национальных систем НТИ стран СНГ, центральной и восточной Европы в новых общественно-политических и социально-экономических условиях: Тезисы докл. и сообщений международной научно-практической конференции. 19—20 мая 1993 г., Киев, УкрИНТЭИ, ч. 2.— С. 34—35
7. dBASE III Plus: Learning and Using.— Ashton Tate, 1986.— 817 p.
8. Clipper aBASE III Compiler — Nantucket Corporation, 1985.— 247 p.
9. Куправа Т. А. Создание и программирование баз данных средствами СУБД dBASE III Plus, FoxBase Plus, Clipper.— М.: Мир. 1991.— 110 с.
10. П. В. Кукруза. Turbo Pasca+Clipper // Компьютеры+Программы, № (2), 1993 г.— С. 50—58.

Материал поступил в редакцию 12.04.94.

Поправка к сб. «Научно-техническая информация», сер. 2, 1994, № 5

- На с. 6 в списке литературы пропущено:
8. Мирский Э. М. Международные исследования и дисциплинарная организация науки.— М.: Наука, 1980.— 304с.