

КОМПЬЮТЕРИЗАЦИЯ БИБЛИОТЕКИ

УДК 026/027:681.3

Вислый А.И., Мосягин В.В.

Поиск

(Занятие 4-е)¹

Рассказано о кодировке русского алфавита, различных процедурах поиска, о взаимодействии компьютера и пользователя.

На предыдущем занятии мы убедились, что электронный каталог (хороший) значительно богаче предметного, систематического и любых других мыслимых каталогов и картотек, вместе взятых. Теперь разговор о важных нюансах и принципиально новых возможностях поиска.

Русификация

Вы получили компьютер. Вариант первый: имеются русские буквы и на клавиатуре, и на экране.

Вариант второй: на клавиатуре русские буквы есть, а на экране они появляться не желают. Вариант третий: русских букв нет. Программист скажет: "Нет драйвера русификации!". И принесет Вам маленькую программку, и научит: "Нажимайте вот эти клавиши для переключения с одного алфавита на другой". Вроде бы все в порядке.

Но не всегда! Попробуем разобраться в существе вопроса. Каждую букву (символ) компьютер кодирует числом. Так, латинскую букву «А» — числом 65, русскую букву «А» — числом 128 и т. д. У себя в памяти он хранит числа. Кодировка латинских букв осуществляется автоматически. А русские буквы кодируются в зависимости от той программки, которую Вам принесли. Однако програми сейчас слишком много, и кодировка, к несчастью, не у всех совпадает.

Основная проблема возникает с маленькой русской буквой «р». Запомните код маленькой русской буквы «р»: 224. И спрашивайте у программистов, такой ли код дает их программа. Если они этого не знают, заставьте их

¹ Продолжение. Начало см. в № 10.

разобраться. Это очень важно. При несовпадении кодировок может случиться, что на компьютере, где Вы вводили данные, поиск идет успешно, а на соседнем, куда Вы данные переписали, все поисковые элементы с буквой «р» не ищутся.

Мы используем у себя программу gk.exe и никаких проблем с кодировкой у нас пока не было.

Большие и маленькие буквы Правильнее, конечно, прописные и строчные. Но чтобы Вам было легче читать, упростим: большие и маленькие.

С точки зрения компьютера варианты фамилии Иванов, иванов, ИВАНОВ, иВанов являются совершенно разными. Вспомните про кодировку. Буква «И» (большая) имеет код 136, а буква «и» (маленькая) — код 168. Второй и четвертый варианты просты, но вот первый и третий могут доставить много неприятностей.

Еще более проблематично обстоит дело с ключевыми словами. Как писать ключевые слова? С большой буквы, с маленькой или, может быть, трлко большими буквами?

Чтобы избавиться от этих проблем раз и навсегда, достаточно при внесении в поисковые таблицы сразу же перевести все маленькие буквы в большие. Это компьютер может сделать автоматически. Далее, при поиске информации, то же самое нужно сделать и с текстом запроса, т. е., если Вы вводите фамилию автора — Иванов, компьютер сначала переведет ее в ИВАНОВ и только после этого приступит к поиску информации.

В принципе, развивая метод перевода маленьких букв в большие, можно придумать и более сложные процедуры занесения информации в поисковые таблицы и ее дальнейший поиск. Например, при обработке сведений об авторах можно сначала все знаки препинания заменить пробелами, потом все повторяющиеся пробелы заменить одним пробелом, взять первое слово (фамилию) и переписать большими буквами, из других слов (имени и отчества) взять по первой букве и тоже записать большими буквами. Тогда из двух вариантов написания фамилии

Иванов Иван Иванович

Иванов И.И.

получится один и тот же поисковый элемент:

1 этап	Иванов, Иван Иванович	Иванов И И
2 этап	Иванов Иван Иванович	Иванов И И
3 этап	ИВАНОВ Иван Иванович	ИВАНОВ И И
4 этап	ИВАНОВ И И	ИВАНОВ И И

Так как подобных процедур можно придумать множество, то хорошо бы, чтобы программа позволяла легко настраиваться на произвольную обработку поисковой информации.

Поиск по началу слова

* Большинство систем позволяют производить поиск по началу слова (фразы). Это означает, что при запросе ИВАНО Вы получите книги с авторами: ИВАНОВ, ИВАНОВА, ИВАНОВСКАЯ, ИВАНОУСКАС и т. д. В каких-то случаях это хорошо, а в каких-то — не очень. Например, если Вы хотите получить сведения о всех книгах по компьютерам и подозреваете, что в одном случае было введено ключевое слово КОМПЬЮТЕР, в другом — КОМПЬЮТЕРЫ, в третьем — КОМПЬЮТЕРИЗАЦИЯ и т. д., то способ поиска по началу слова для Вас очень удобен. Вы вводите слово КОМПЬЮТЕР и получаете все варианты сразу. Но если Вам нужен конкретно ИВАНОВ, то поиск по началу слова Вам ни к чему. Следовательно, в системе должна быть предусмотрена возможность переключения от точного поиска к поиску по началу слова и наоборот. Иногда поиск по началу слова называют поиском «по усечению».

Проблема синонимов

Точнее, не синонимов, а разных по написанию слов с одинаковым смыслом при поиске. Ключевые слова КОМПЬЮТЕР и КОМПЬЮТЕРЫ с точки зрения компьютера абсолютно разные слова. Точно так же, как и ключевые слова КОМПЬЮТЕР и ЭВМ. Если Вы хотите обеспечить поиск книги по любому варианту, Вы должны указать в качестве ключевых слов все варианты:

КОМПЬЮТЕР

КОМПЬЮТЕРЫ

ЭВМ

ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНАЯ МАШИНА

ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ

Этот список из пяти ключевых слов можно сократить до трех методом усечения. Так, поиск по началу слова КОМПЬЮТЕР позволяет вместо первых двух слов писать только одно. Точно так же поиск по началу слова ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬН позволяет вместо последних двух слов писать одно. Таким образом список ключевых слов-синонимов сокращается!

КОМПЬЮТЕР

ЭВМ

ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНАЯ МАШИНА

Что делать с оставшимися тремя? Есть два варианта. В первом их просто необходимо вводить при вводе каждого нового издания соответствующей тематики. В этом случае все зависит от квалификации систематизатора. Если он догадается ввести достаточное количество синонимов, то все в порядке. Во втором варианте необходимо предварительное составление словаря ключевых слов или тезауруса. С точки зрения теории этот вариант предпочтительнее. Однако сам процесс составления таких словарей или тезаурусов трудоемкий и длительный. Особенно он сложен в библиотеках университетов, где тематический охват литературы очень широк. На практике чаще применяется первый подход. Одновременно с этим систематизаторам можно порекомендовать найти и использовать в работе словарь или тезаурус по соответствующей тематике.

Поиск по словарю При поиске всегда хочется знать, какие ключевые слова, авторы, заглавия и т. п. есть в электронном каталоге. Поэтому в системе (БИБЛИОТЕКА 3.02) предусмотрена возможность показать на экране полный список, например, тематических рубрик:

С л о в а р ь

- Абазинская литература, советская
- Абазины
- Аббревиатуры
 - Английский язык
 - Словари
- Абдуллаев, Абдумавлян Абдуллаевич (р.1930).
- Абоская академия
 - Научно-исслед. работа
- Абрагам, Анатолий Израилевич (р.1915).
- Абразивные материалы
- Абразивные инструменты
- Абсорбция и адсорбция
 - Анализ и исследование
- Абстракционизм в искусстве
- Абхазская ССР
 - Население
- Абхазы
- Аварская литература

Конечно, такой список может быть очень большим — сотни тысяч и даже миллионы записей. Поэтому система предусматривает процедуру поиска в таком словаре.

После того как нужное ключевое слово найдено в словаре, можно произвести поиск уже в электронном каталоге и отобрать необходимые издания.

Комбинированный поиск Все вышесказанное относится к поиску по единичному запросу. Компьютер позволяет произвести комбинированный поиск, т. е. найти, например, все издания с ключевым словом АЭРОДИНАМИКА и с автором ИВАНОВ. Возможны два подхода к организации такого поиска.

При последовательном поиске сначала отбирается информация по ключевому слову, а затем результаты уточняются по автору. Так, в приведенном выше примере по ключевому слову АЭРОДИНАМИКА может быть найдено 235 записей. После уточнения по автору ИВАНОВ (т. е. поиске среди 235 записей изданий с автором ИВАНОВ) в результате может остаться только 5 записей.

Последовательный поиск обеспечивает выполнение любых самых сложных запросов с помощью всего лишь трех операций:

- уточнения,
- расширения,
- уточнения по отрицанию.

Допустим, мы уже отобрали какое-то количество записей. Тогда уточнение означает: отобрать среди этих записей только те, которые удовлетворяют новому условию. Расширение означает: добавить к уже отобранным записям те, которые удовлетворяют новому условию. Уточнение по отрицанию означает: выбросить из отобранных записей те, которые удовлетворяют новому условию.

Приведем пример комбинированного запроса: "Найти все издания с ключевым словом КОМПЬЮТЕР и ЭВМ, но чтобы автор был не ИВАНОВ". Этот запрос можно выполнить в три этапа:

- поиск по ключевому слову КОМПЬЮТЕР,
- расширение запроса по ключевому слову ЭВМ,
- уточнение по отрицанию по автору ИВАНОВ.

Чем плох метод последовательного поиска? Тем, что каждый элемент запроса требует проведения отдельного поиска и, следовательно, участия в этом процессе пользователя. Чем хорош метод последовательного поиска? Тем, что на каждом его этапе пользователь контролирует ситуацию и сам решает, какое следующее действие произвести.

При поиске посредством булевой алгебры весь запрос формулируется сразу с использованием того или иного языка запросов. Например:

key = "АЭРОДИНАМИКА". and. au1 = "ИВАНОВ"

Получив такой запрос, компьютер сам разбирается, что ему делать. В данном примере он найдет все издания с ключевым словом АЭРОДИНАМИКА и автором ИВАНОВ.

Пример из предыдущего раздела будет выглядеть так:

(key = "КОМПЬЮТЕР". or. key = "ЭВМ"). and. au1 = "ИВАНОВ"

Поиск с использованием булевой алгебры удобен тем, что весь запрос можно сформулировать сразу и не думать, как он будет выполняться. Но учтите, для использования этого способа поиска Вам придется выучить правила формулировки запроса (язык запросов).

Сеть персональных компьютеров Теперь поговорим о том, где должен находиться электронный каталог и как (откуда) должен быть организован доступ к нему? Конечно, если у Вас всего один компьютер, то проблем нет. И электронный каталог находится на этом компьютере, и поиск Вы проводите на этом же компьютере. Но предположим, что у Вас накопилось достаточно информации (несколько десятков тысяч записей) и все уже готово для того, чтобы этой информацией пользовались читатели. Тогда Вам не обойтись без сети персональных компьютеров.

Сети персональных компьютеров бывают двух типов: локальные и с удаленным доступом. Локальные сети рассчитаны на расстояния в несколько сотен метров между компьютерами. Сети с удаленным доступом могут работать на любых расстояниях по телефонным или иным каналам передачи информации. Все, что будет сказано ниже, относится к локальным сетям.

Что собой представляет локальная сеть? Это несколько персональных компьютеров, соединенных проводами. Один из них (более мощный, чем остальные) называется файловым сервером. На этом компьютере и размещается электронный каталог. Остальные компьютеры могут пользоваться данными, лежащими на файловом сервере.

Для создания локальной сети нужен один "хороший" компьютер для файлового сервера (IBM AT 386 или IBM AT 486), несколько персональных компьютеров для локальных станций сети, сетевые адаптеры на каждый компьютер ("железка", которая вставляется в компьютер и подключается к проводам) и сетевая операционная система. Покупка всего этого оборудования сейчас — это вопрос денег. Мы рекомендуем сетевые адаптеры ETHERNET и операционную систему Novell 3.11.

Установка локальной сети решает проблему поиска информации с нескольких персональных компьютеров одновременно. Для этого, кроме перечисленного выше оборудования, необходимо иметь сетевую версию библиотечной системы. БИБЛИОТЕКА 3.02 является сетевой. Опыт работы с ней

показал, что при восьми компьютерах, подключенных к файловому серверу, времени на поиск затрачивается лишь на 4% больше, чем при одном компьютере. Эти результаты получены при файловом сервере IBM AT 386 (4МБ ОЗУ, SCSI 350 МБ). Время поиска одной записи менее 0,1 секунды.