

Технологии информационного поиска в Интернет

1. Понятие об информационно-поисковых системах – ИПС.

Поиск информации – задача, которую человечество решает уже многие столетия. По мере роста объема информационных ресурсов, потенциально доступных одному человеку (например, посетителю библиотеки), вырабатывались все более совершенные поисковые средства и приемы, позволяющие найти необходимый документ. Технологии поиска информации совершенствовались в каталогах и информационных отделах крупных библиотек. В 70-е годы XX века появились библиотечные компьютерные базы данных, доступ к которым сначала обеспечивался через модемное подключение, а затем по протоколу telnet через Интернет. Стоимость работы с такими базами данных весьма велика. Например, одна минута работы с базой DIALOG (Ww.dialog.com) может стоить доллар, а вывод на экран одного элемента найденной записи (из, например, семидесяти) – 20 центов. Такая высокая стоимость поиска информации потребовала создания эффективных приемов поиска.

Все найденные за много лет средства и приемы информационного поиска стали доступны и эффективны и при поиске информации в Интернет.

Для поиска информации в библиотеках используются, в основном, три вида каталогов: алфавитные, систематические и предметные. Информационно-поисковые системы (ИПС) Интернет, при всем их внешнем разнообразии, также попадают в один из этих классов.

2. ИПС в библиотеках и сети Интернет

Классификационные информационно-поисковые системы

В классификационных ИПС используется иерархическая (древовидная) организация информации, которая называется КЛАССИФИКАТОРОМ. Разделы классификатора называются РУБРИКАМИ. Библиотечный аналог классификационной ИПС – систематический каталог. Классификатор разрабатывается и совершенствуется коллективом авторов. Затем его использует другой коллектив специалистов, называемых СИСТЕМАТИЗАТОРАМИ. Систематизаторы, зная классификатор, читают документы и приписывают им классификационные индексы, указывающие, каким разделам классификатора эти документы соответствуют.

Классический пример классификационной ИПС – Yahoo! (www.yahoo.com). Едва появившись, Yahoo! быстро завоевала признание качественной проработкой классификатора. Год назад в Yahoo! работали более 100 систематизаторов.

Классификационные ИПС обладают рядом специфических недостатков. Уже разработка классификатора связана с оценкой относительной важности различных областей человеческой деятельности. Например, сравнивая классификаторы многих ИПС в Интернет (таких, как Yahoo!), замечаем, что во многих из них нет раздела “Наука”. Если нет необходимого нам класса в ИПС, в поиске поможет человек, который знает и классификацию и различные области знаний, и связи между ними – «информационный брокер». В библиотеке это – библиограф. Библиографы, понимая, что читатели не всегда глубоко изучают классификации, положенные в основу систематических каталогов, выработали два приема, облегчающие жизнь читателям. Они называются «ОТСЫЛКА» и «ССЫЛКА»:

– ОТСЫЛКА используется тогда, когда создатели классификатора и систематизаторы в состоянии принять четкое решение, к какому разделу классификатора относится данный документ, а пользователь в поисках этого документа может обратиться к другому разделу. Тогда в этом «другом» разделе помещается ссылка «См» в тот раздел классификатора, где действительно находится данный документ. Например, информация о картах стран может быть размещена в разделах «Наука» - «География» - «Страна» или «Экономика» - «География» - «Страна» или «Справочники» - «Карты» - «Страна». Принимается решение, что карты стран помещаются во второй раздел «Экономика» - «География» - «Страна», а в

остальные два раздела помещаются отсылки на него.

– ССЫЛКА используется в ситуации, когда даже классификаторы и систематизаторы не в состоянии принять четкого решения об отнесении документа к определенному разделу классификатора. Ссылка может иметь самые разнообразные варианты, например, «Похожие документы» и т.п.

Словарные ИПС

Проблемы, связанные с использованием классификационных ИПС, привели к созданию информационно-поисковых систем словарного типа. Основная их идея – создать базу данных из слов, встречающихся в документах Интернет, в котором при каждом слове будет храниться список документов, в которых встречается это слово. Если поиск слов в таком словаре происходит быстро, то можно отказаться от услуг систематизаторов и классификаторов. Несмотря на обилие слов, большинство из них употребляется нечасто, из них исключают предлоги, союзы, артикли, т.е. слова, бесполезные при поиске информации. В результате словарь самой крупной ИПС AltaVista имеет объем всего несколько Гбайт. Между искомыми словами вставляют ЛОГИЧЕСКИЕ ОПЕРАТОРЫ ИПС (И, ИЛИ, НЕ) и поиск происходит за один проход.

Многие ИПС являются одновременно словарными и классификационными или объединяют свои усилия в разработке классификаторов.

3. Интернет – копилка информации. Какая информация есть в Интернет?

Интернет – это несколько десятков миллионов компьютеров, разбросанных по всему свету и соединенных *любыми средствами связи в единую Сеть*. В качестве линий связи могут выступать простые телефонные каналы, коаксиальные и оптоволоконные линии, каналы радиосвязи, радиорелейные и спутниковые линии.

Бурное развитие Интернет является самым значительным и волнующим событием в компьютерном мире после появления персональных компьютеров в 80-х годах XX столетия. Как в средние века изобретение печатного станка ускорило перемены в экономике, политике, культуре и общественных отношениях, так сегодня Интернет оказывает воздействие на все стороны человеческого сообщества. Причина в том, что глобальная компьютерная сеть стала не только средством оперативного обмена информацией, но и огромным, к тому же чрезвычайно мобильным хранилищем самой разнообразной информации. Объединение глобальных сетей в Интернет знаменует собой третью информационную революцию, когда практически вся накопленная человечеством информация оказалась переведенной на электронные носители. Мощные компьютеры и поисковые системы способны оперативно предоставлять эту информацию пользователю. Трудно представить жизнь образованного человека без Интернет, хотя используется у нас сейчас только для 3-х дел: доступа к информации, самовыражения и общения.

Доступ к информации:

В Интернет можно найти: информацию о спорте и отдыхе, компьютерах, языках программирования и разработках ПО; готовые рефераты и материалы для них; страницы, посвященные музыкальным группам, спортивным командам, кинозвездам, политическим деятелям и т.д. Например, на музыкальных сайтах можно найти историю любимых групп, дискографию, тексты песен с нотами, свежие клипы и расписание гастролей. Поклонники группы после каждого концерта пишут о нем отчеты и отправляют их на сервер групп по электронной почте. Так постепенно трудами многих людей умножается информационное пространство Интернета.

Интернет для самовыражения:

Предположим, что вам есть что сказать миру. Есть два способа самовыражения: публичное выступление и публикация. Для того чтобы заниматься печатными публикациями, необходимо решить много юридических, технических и финансовых проблем. Не всякий завод, не любая организация может позволить себе издание многотиражной газеты.

Эпоха Интернета меняет все. Любая организация и любой человек теперь могут стать сетевым издательством. В Интернете можете издавать хоть всемирный альманах и ни о каких разрешениях не спрашивать. Ничто не мешает вам иметь свою собственную страничку или собственный виртуальный сервер, посвященный себе, своей семье и своим друзьям. Тогда пожелтевшие фотографии из бабушкиных альбомов успешно перекоچуют на сервер, и в мире станет одним виртуальным сервером больше. Каждый посетитель такого сервера сможет увидеть, как ваша личная история и история вашей семьи связаны с историей народа, страны, человечества в целом

Интернет для общения

Часто у владельца компьютера возникает потребность в общении с другими людьми. И здесь Интернет откроет перед вами новые, не востребуемые до сих пор возможности. Он позволит получить доступ к десяткам миллионов компьютеров всего цивилизованного мира, где вы найдете информацию по любой интересующей вас теме. Вы не только сможете переписываться и разговаривать с огромным количеством людей, живущих на планете, но и увидите с ними, внести свой посильный вклад в развитие цивилизации, стать гражданином мира.

4. Поиск информации в Интернет

Когда говорят о поиске информации в Интернет, прежде всего имеют в виду службу WWW, так как это огромное хранилище распределенной информации. Выбираем и настраиваем на свой вкус программу-браузер (в переводе – посетитель выставки, обозреватель - специальная программы для путешествия по сети Интернет) и отправляемся в путь. Для того чтобы попасть куда-то, надо знать только одну вещь-адрес. При этом возможны три варианта:

- Адрес вам известен, но вы там еще никогда не были;
- Адрес вам известен, и вы там уже бывали;
- Адрес вам неизвестен, и вы, естественно, там никогда не были.

Если вы знаете адрес, но там не бывали: введите в строку АДРЕС: этот адрес с клавиатуры, например www.yandex.ru (адрес русскоязычного поискового сервера). Сделать это совсем нетрудно, хотя и не очень удобно. Дождитесь, пока установится соединение, затем пойдет прием информации, и только когда внизу в статусной строке появиться сообщение Document Done (Документ принят), можно двигаться дальше. Прибыв на место, не забудьте занести этот адрес в «Закладки» (добавить), чтобы в будущем не затрудняться с набором с клавиатуры.

Если вы здесь уже когда-то бывали и сохранили «закладку», то снова попасть в это же место можно очень просто, указав на адрес в «Закладках».

Если же вы хотите попасть куда-то, толком еще не зная куда именно, вам нужно воспользоваться специальным поисковым сервером. В мире есть такие серверы, где по ключевым словам, по темам, по обрывкам информации можно узнать адреса серверов, которые вам нужны, причем не просто узнать, а тут же, щелкнув мышкой на ссылке, отправиться в пункт назначения. Не будь подобных поисковых серверов, работать в Сети, в которой сегодня насчитываются десятки миллионов документов (культура искусство, медицина, отдых и развлечения и т. п.), было бы невозможно.

Составители справочников в ИПС Интернет собирают информацию сами и часто предлагают всем желающим присылать сведения о своих виртуальных серверах и домашних страницах в тематические поисковые серверы.

Поисковые серверы в Интернет

Что это такое? Это необычные узлы сети, которые занимаются поиском, сбором и упорядочением информации, имеющейся на всех прочих серверах Сети. На поисковых серверах установлены специальные программы поиска. Такие программы по вашему запросу перерывают огромное количество собранных и хранящихся на сервере данных обо

всех ресурсах Сети: от крупных серверов до самой маленькой и скромной домашней страницы какого-нибудь начинающего пользователя.

Для того чтобы поисковый сервер знал обо всем, что ежесекундно появляется в Сети, существуют различные способы. Чаще всего пользователи сами извещают поисковый сервер о появлении новых документов. Некоторые серверы пользуются специальными программами-роботами, «червяками», которые самостоятельно исследуют пространство Интернет в поисках новой, еще не отраженной в их базах информации с тем, чтобы восполнить этот пробел.

Тем или иным способом, но данные обо всех закоулках Сети «стекаются» на поисковые серверы для того, чтобы облегчить вам жизнь и упростить поиск нужной информации. Остается только научиться правильно использовать мощные ресурсы поисковых машин.

В мире существует несколько наиболее популярных поисковых серверов. Начните свою навигацию с одного из них. Мы очень рекомендуем сервер www.altavista.com, расположенный в Кремниевой Долине (США), этот сервер появился в Интернете не так уж давно, но, по-видимому, знает о ресурсах Сети больше, чем другие. При этом он не загружает вас ненужной информацией, не задает лишних вопросов, а скромно и очень быстро исполняет ваши желания. И при этом еще понимает символы кириллицы в русских словах. При входе на сервер дается возможность выбрать режим поиска: простой запрос или расширенный.

Для простого поиска на этом и любом другом сервере в окно поиска введите ключевое слово (что вы ищете) и нажмите «Найти» или «Сопоставить». Все серверы, на которых будут обнаружены ключевые слова, станут доступными для выбора. Символ <*> применяется для замены любого количества символов, (например, можно набрать образован*, чтобы найти все документы по образованию). Применяйте в поиске только строчные буквы, т.к. прописные воспринимаются «буквально».

Расширенный поиск позволяет создать запрос, используя логические операторы и, или, но, около (and, or, not, near) и скобки.

Например, ищем документы, посвященные воспитанию детей – школьников:

(дет* or школьн*) and воспит*

обучающие игры для детей:

(дет*or школьн*) and (игр* near обуч*)

Будет найдена информация не о каких-то играх, а об обучающих играх для детей или школьников.

Найденные документы, содержащие ключевые слова, должны быть представлены пользователю. В каком порядке? Поисковые серверы способны выдать списки документов, содержащие миллионы ссылок (кстати, это говорит о том, что запрос составлен плохо). Хорошо, чтобы найденные документы были представлены в порядке заданных критериев важности –ранжированы. Методы ранжирования в разных поисковиках разные. Altavista позволяет в режиме сложного поиска указать перечень терминов, которые повышают ранг найденного документа (т.е. перемещают его ближе к началу списка). Rambler и Yandex позволяют указать вес каждого из терминов, участвующих в запросе и точно настроить порядок следования документов.

5.Русскоязычные тематические поисковые серверы

Существуют также тематические поисковые серверы. Как в библиотеках существуют алфавитные, систематические и тематические каталоги, так и среди поисковых серверов есть такие, которые производят предварительную сортировку информации по тематическому принципу. Это полезно, когда интересуется широкий круг вопросов по узкой теме. Работа с тематическими серверами требует предварительных исследований для общего отбора информации. Владелец сервера не в силах сделать такой отбор и полагается на авторов статей, которые публикуя статью в Интернет, дополнительно заполняют специальные формы. Так за «бортом» тематических серверов остаются

огромные пласты информации, о которых никто не удосужился известить тематические серверы. Предлагаем следующие тематические поисковые серверы:

- www.yandex.ru - многоуровневый тематический русскоязычный. Создан в 1997 г и помимо www-серверов.ru, su просматривает содержание зарубежных русскоязычных web-узлов.
- www.rambler.ru - тематический русскоязычный. Первая профессиональная поисковая система, начала работать с конца 1996 года. Обеспечивает поиск на 2 млн. страниц России и ближнего зарубежья. Удобный интерфейс, легко составить запрос.
- www.yahoo.com - многоуровневый тематический.
- www.aport.ru – тематический.

Будем искать что-нибудь о дельфинах. Когда браузер загрузит первую страничку поискового сервера, вы видите рубрики и окно для ввода ключевого слова. Можно пойти двумя путями: выбрать рубрику (например, наука - зоология), войти в нее и начать поиск по ключевому слову. Но чтобы расширить поиск, можно сразу ввести ключевое слово и нажать слово ПОИСК (seach). Через некоторое время поисковик сообщит вам, что найдено несколько категорий, включающих это слово. Вы можете зайти в любую из них и продолжить поиск. Когда вы ищете информацию по какой-нибудь общей теме, лучше всего начать именно с каталога, хотя каталоги охватывают лишь малую часть ресурсов Интернет.

Имея дело с поисковыми серверами, не жалейте времени. Чем грамотнее вы составите запрос, тем лучше будут результаты. Вводите не только отдельные слова, но и словосочетания. Если поисковый сервер не выдал никаких результатов, попробуйте еще раз, используя другие слова и словосочетания. Не бойтесь повторить свой запрос на другом поисковом сервере. Зачастую он дает совершенно другие результаты. Поисковые серверы часто предоставляют на своих страницах возможность поиска в других поисковых серверах.

6. Вопросы:

- 1) Что такое ИПС? Какие ИПС в библиотеке вы знаете?
- 2) Как называются программы просмотра страниц всемирной паутины?
- 3) Какие виды запросов для поиска информации вы знаете?
- 4) Задания:
- 5) Посмотрите информацию об Озерске на сервере www.ozg.ru
- 6) Посмотрите, какую информацию представляет для вас Ленинградская областная детская библиотека на сервере www.deti.spb.ru
- 7) С помощью поискового сервера www.yandex.ru просмотрите странички двух школ.
- 8) С помощью поискового сервера www.rambler.ru выполните запрос ДОМАШНИЕ ЖИВОТНЫЕ СОБАКИ