

Лекция № 13

**ПОИСК ИНФОРМАЦИИ С
ИСПОЛЬЗОВАНИЕМ
КОМПЬЮТЕРА. ПРОГРАММНЫЕ
ПОИСКОВЫЕ СЕРВИСЫ**

Для получения качественного результата при проведении поиска необходимо соблюдать ряд условий.

Основными из них являются:

- контроль полноты охвата ресурсов;
- достоверности найденной информации.

Возможность нахождения той или иной информации в Сети определяется *полнотой охвата ее ресурсов*.

Часто проведение поиска требует задействования максимального объема возможных источников.

Необходимым условием успешного планирования и проведения поисковых работ становится знание основных существующих на сегодняшний день типов ресурсов Интернета, понимание технической и тематической специфики их информационного наполнения и особенностей доступа к ним.

Качество проводимого поиска определяется *достоверностью найденной информации.*

Контроль ее достоверности может производиться разными способами, в которые входит:

- нахождение и сверка с альтернативными источниками информации,
- установление частоты его использования другими источниками,
- выяснение статуса документа и сайта, на котором он находится,
- получение сведений о компетентности и положении автора материала и ряд других.

По способу организации и хранения информации ее источники в Интернете можно разделить на:

- *файловые серверы;*
- *web-сайты;*
- *телеконференции;*
- *базы данных.*

Файловые серверы

являются традиционным способом хранения данных и представляют собой компьютеры, часть дискового пространства которых доступна через Интернет.

Доступ к данным на таком сервере осуществляется с помощью специальных программ, поддерживающих протокол передачи файлов — FTP. Данный протокол в общем случае требует авторизации, то есть идентификации пользователя.

Для осуществления доступа к файлам со стороны произвольного пользователя Сети обычно используется так называемый анонимный вход под регистрационным именем anonymous, для которого пароль не требуется.

Web-сайты

являются сегодня основным и наиболее распространенным типом информационных ресурсов в Сети.

Сайт может содержать информацию, представленную в самой произвольной форме:

- графической,
- звуковой,
- видеоизображения и т. д.

Телеконференции

могут являться источником необходимой информации, как правило, носящей неофициальный характер.

Телеконференции представляют собой способ общения людей, имеющих доступ в Сеть, и предназначены для обсуждения каких-либо вопросов или распространения информации.

Они позволяют добиться обратной связи со множеством лиц и произвести детальное обсуждение какой-либо проблемы территориально разобщенными людьми.

Базы данных

- могут содержать самую произвольную информацию:
- публикации,
- справочную информацию,
- другие данные.

Наиболее широко распространен способ доступа к базам данных через стандартные браузеры, так как он обеспечивает максимальную потенциальную аудиторию потребителей информации.

Наряду с непосредственным извлечением информации из баз данных широко используется динамическое построение web-страниц в процессе исполнения пользовательских запросов.

По принципу организации и использования средства поиска можно выделить следующие *инструменты*:

- *поисковые машины;*
- *мега-средства поиска;*
- *специализированные средства поиска;*
- *каталоги.*

Поисковые машины

являются ключевым инструментом поиска информации, поскольку содержат индексы большинства web-серверов Интернета.

Однако именно это достоинство оборачивается их главным недостатком.

На любой запрос они выдают обычно чрезмерно большое количество информации, среди которой только незначительная часть является полезной, после чего требуется значительный объем времени для ее извлечения и обработки.

Мега-средства поиска

позволяют ускорить выполнение запроса путем передачи аргументов поиска, то есть ключевых слов, одновременно нескольким поисковым системам.

При значительном ускорении процесса и увеличении охвата поиска, этот способ имеет ряд недостатков, связанных с необходимостью координации во времени поступления результатов обработки запроса от нескольких систем, а также тем, что они не позволяют использовать возможности языка запроса каждого из применяемых поисковых средств.

Специализированные средства поиска

представляют собой «программы-пауки», которые в автоматическом режиме просматривают web-страницы, отыскивая на них нужную информацию.

Механизм их работы близок к механизму, который используют поисковые системы для построения своих индексных таблиц.

Выбор между ними представляет собой классический выбор между применением универсальных или специализированных средств.

Каталоги

как и поисковые машины, используются посетителями Интернета для нахождения необходимой информации.

Каталог представляет собой иерархически организованную структуру, в которую данные заносятся по инициативе пользователей.

Как следствие, объем информации в них несколько ограничен по сравнению с поисковыми системами, но в то же время более упорядочен благодаря лежащей в их основе иерархической тематической структуре.

Поиск информации в Интернете может быть произведен при помощи *двух основных методов*:

- *использование поисковых систем;*
- *поиск по гипертекстовым ссылкам.*

В зависимости от целей и задач поиска эти методы могут быть использованы по отдельности или в комбинации друг с другом.

Использование поисковых систем

сегодня является одним из основных при проведении предварительного поиска.

Его применение основано на ключевых словах, которые передаются системе в качестве аргумента поиска.

Результатом является список ресурсов Интернета, подлежащих детальному рассмотрению.

Получение наиболее релевантного результата требует проведения предварительной работы по составлению тезауруса.

Поиск по гипертекстовым ссылкам

Поиск информации может быть произведен путем последовательного просмотра с помощью браузера связанных ссылками web-страниц.

К этому виду поиска также относится использование каталогов, классифицированных и тематических списков и всевозможных небольших справочников.

Такой метод трудоемок, однако «ручной» просмотр web-страниц часто оказывается единственно возможным на заключительных этапах информационного поиска.

Он может быть также более эффективен при проведении повторных циклов или просмотре вновь образованных ресурсов.

Наиболее широко используемым, но и наиболее сложным является метод поиска с использованием поисковых систем.

- распространенность обусловлена тем, что поисковые системы содержат в себе индексы огромного количества сайтов и при правильно сформированном запросе можно сразу же получить ссылки на интересующие ресурсы;
- сложность метода состоит в том, что для того, чтобы результат был качественным, необходимо уметь выбрать наиболее подходящие поисковые системы, правильно формулировать запросы к ним, учитывать их особенности и функциональные возможности.

Для эффективного использования поисковых серверов необходимо:

- *составить тезаурус* (список ключевых слов, организованный с учетом семантических отношений между ними);
- *отобрать поисковую систему* (основными критериями выбора поисковых систем являются объем индексной базы сервера и степень развитости самой поисковой машины, то есть уровень сложности воспринимаемых ею запросов);

Для эффективного использования поисковых серверов необходимо:

- *составить и выполнить запрос к поисковым машинам* (на основе тезауруса формируются запросы к выбранным поисковым серверам. После получения первоначальных результатов возможно уточнение запросов с целью отсечения очевидно нерелевантной информации. Затем производится отбор ресурсов, начиная с наиболее интересных, с точки зрения целей поиска, и данные с ресурсов, признанных релевантными, собираются для последующего анализа.

Для эффективного использования поисковых серверов необходимо:

- *провести анализ ресурсов и сбор искомой информации.* Первичный анализ ресурсов может основываться на аннотациях, если они есть, а при их отсутствии — на ознакомлении с информационным наполнением ресурса. Далее информация извлекается с отобранных источников и используется в соответствующих поиску целях.

Программные поисковые сервисы

Поисковая система — программно-аппаратный комплекс с веб-интерфейсом, предоставляющий возможность поиска информации в Интернете.

Программной частью поисковой системы является *поисковая машина* (поисковый движок) - комплекс программ, обеспечивающий функциональность поисковой системы и являющийся коммерческой тайной компании-разработчика поисковой системы.

Популярные поисковые системы:

- *всеязычные:* Google, Bing, Yahoo! И принадлежащие этой компании поисковые машины: Inktomi, AltaVista, Alltheweb;
- *англоязычные и международные:* AskJeeves (механизм Teoma);
- *русскоязычные:* Яндекс, Mail.ru, Rambler, Nigma, Генон, Gogo.ru , Aport.

Необычные поисковые системы:

- Google (с его помощью ортодоксальные иудеи могут найти контент, признанный раввинами удовлетворяющим религиозным требованиям);
- Yauba (индийский поиск «для параноиков» — в отличие от привычных поисковиков, за пользователями не следят, а все записи о его действиях удаляются с сервера);
- TinEye (поисковая система, специализирующаяся на поиске изображений в Интернете);
- Генон (поисковая система, собирающая и создающая контент у себя на сайте).

Яндекс (www.yandex.ru)

Как на любых поисковых серверах, на Яндексе есть поле для ввода запроса.

Запрос – это набор ключевых слов, которые сообщают поисковой системе, что вы хотите найти.

Чем конкретнее запрос, тем точнее будет результат!

После ввода запроса, Яндекс произведет поиск по своей базе данных и покажет результаты в виде ссылок на подходящие страницы и краткие выдержки из текстов.

Яндекс (www.yandex.ru)

Часто число ссылок бывает очень большим - тысячи или даже миллионы страниц могут соответствовать запросу.

Загрузить сразу все найденные ссылки невозможно, Яндекс выдает результат по 10 ссылок на странице.

Ссылки отсортированы по релевантности, потому скорее всего, вы найдете то, что нужно, среди первого десятка ссылок.

Поиск по ключевым словам

Предположим, что мы хотим найти изображения новых автомобилей и их характеристики.

Если ввести поле ввода текст «новые машины», большинство полученных ссылок будет касаться Интернет-магазинов, где в продажу поступили новые машины, или советы по покупке новой машины. Это слишком общий и неконкретный вопрос.

Запрос нужно хорошо продумать и внести в него интересующие слова.

Необязательно, чтобы получилось осмысленное предложение или словосочетание, главное - указать нужное слово.

Не нужно ставить знаки препинания - они в поисковых системах имеют особое значение!

Разделяйте слова пробелами.

Если ввести более конкретный текст запроса: «новые модели машин». то страниц будет найдено меньше , но они будут больше соответствовать нашей цели. Из пары десятков выданных страниц уже самые первые из них будут полностью удовлетворять нашим целям - содержать фотографии и характеристики новых моделей автомобилей.

Управляющие символы в запросах

Знаки препинания в запросах позволяют более точно описать, что мы хотим найти.

С помощью символов «+» и «-» показывают значимость того или иного слова.

Символ «+» означает, что слово обязательно должно встречаться на странице.

Символ «-» означает, что данное слово не должно встречаться вообще.

Эти символы ставятся вплотную к слову (без пробела).

Когда вводится набор слов, то находятся страницы, которые содержат эти слова, причем необязательно, что слова будут стоять рядом.

Если же поместить это словосочетание в кавычки, то будут найдены только те страницы, на которых эти слова стоят рядом.

По умолчанию поиск учитывает все формы заданного слова согласно правилам русского языка.

Однако существует возможность поиска по точной словоформе, для этого перед словоформой ставится восклицательный знак «!».

Так по запросу !информатика будут найдены все документы, содержащие словоформу «информатика».

Между словами можно поставить знак «|», чтобы найти документы, содержащие любое из этих слов.

Удобно при поиске синонимов.

Запрос

фото | фотография | фотоснимок | снимок
| фотоизображение

задает поиск документов, содержащих хотя бы одно из перечисленных слов.

Еще один знак, тильда «~» позволит найти документы с предложением, содержащим первое слово, но не содержащее второе.

По запросу
билет ~ студенческий
будут найдены все документы,
содержащие слово «билет», рядом с
которыми (в пределах предложения) нет
слова «студенческий».

Поиск картинок

В системе Яндекс для поиска графической информации имеется специальная ссылка Картинки, расположенная под полем ввода запроса. Щелкнув по ней, переключимся в режим поиска картинок.

Сам поиск осуществляется точно так же - вводим запрос и нажимаем Enter. Можно выбрать размер картинок.

Вопросы

1. Чем определяется возможность нахождения и качество поиска информации в сети Интернет?
2. Перечислите инструменты поиска информации.
3. Каковы основные методы поиска информации в Интернете?
4. Каков алгоритм эффективного использования поисковых серверов?
5. Что такое поисковая система?
6. Какие популярные поисковые системы вы знаете?
7. Назовите необычные поисковые системы.
8. Как правильно составить запрос для поиска информации?
9. Какую роль играют управляющие символы в запросах?
10. Как осуществить поиск картинок в системе Яндекс?