

image not found or type unknown



Введение:

Сеть Интернет похожа на огромную мировую библиотеку, имеющую только одно, но существенное отличие: для поиска книги в библиотеке есть каталог, в крайнем случае, можно обратиться к опытному библиотекарю. Полного каталога Интернета не существует. Но, тем не менее, поиск в глобальной компьютерной сети возможен, и это, пожалуй, является одной из наиболее важных его сторон. Для поиска данных в сети используются специальные серверы, информация на которых поддерживается и обновляется практически автоматически.

Интернет – это глобальная компьютерная сеть, которая связывает между собой как пользователей компьютерных сетей, так и пользователей ПК. Интернет медленно, но, верно, становится основным средством корпоративного общения, уступая пока телефону.

В Сети наличествует гигантское количество информационных ресурсов. По некоторым оценкам, число документов превысило 65 млн. и продолжает стремительно расти. Такой объем информации требует правильной организации процесса поиска и применения специальных технических средств, таких как поисковые машины. Простой поиск по достаточно распространенному ключевому слову дает обычно от десятков тысяч до нескольких миллионов ссылок. Очевидно, что работа с таким большим количеством документов практически невозможна, тем более что подавляющая их часть содержит информацию, не относящуюся к делу.

Источники информации в Интернете различаются по способу представления информации, а следовательно, и по методу доступа к ним.

1. Средства поиска файлов

Поиск файла вручную в сложной структуре каталогов ftp-сервера может занять достаточно много времени. Для упрощения и ускорения поиска была разработана поисковая служба Интернета Archie, представляющая собой специальные Archie-сервера, хранящие содержание каталогов анонимных ftp-серверов. При обращении с поисковым запросом на Archie-сервер результатом поиска является список адресов анонимных ftp-серверов, на которых имеется искомый файл.

Но возникает задача отыскать среди множества файлов этого сервера искомый, что достаточно сложно из-за маловыразительных и непонятных имен файлов и каталогов. Для решения этой проблемы используется система Gopher, позволяющая перемещаться по системе контекстных меню, показывающих содержимое файлов с использованием понятых обозначений. Существует очень много Gopher-серверов, которые содержат архивы данных в виде иерархически структурированных каталогов, упорядоченных по содержанию. Работа с ними очень проста и соответствует работе с обычным отображением файловой системы.

Существует расширение этой системы - Veronica, которое содержит в своей базе данных каталоги всех Gopher-серверов. После ввода поискового запроса Veronica автоматически просматривает все Gopher-каталоги на наличие искомой информации и тем самым избавляет от долгого поиска вручную по многим Gopher-серверам.

С таким способом навигации Gopher в определенной степени был предшественником WWW. В настоящее время применение Gopher уменьшается пропорционально росту использования WWW.

1. Основные требования к поиску

К результатам поиска предъявляются требования полноты охвата ресурсов, достоверности полученной информации, минимальных затрат времени и максимальная скорость поиска.

Требование полноты охвата ресурсов не нуждается в дополнительных пояснениях, за исключением необходимости использовать при поиске ресурсы не только WWW, но и других служб Интернета.

Достоверность информации, учитывая природу Интернета, становится чрезвычайно важным требованием. Оценка достоверности может производиться как традиционными методами (проверка легальности публикаций на бумажных носителях, получение сведений об организациях и авторах, выяснение действительности их электронных ресурсов и т.п.), так и с использованием возможностей Интернета (ознакомление с альтернативными источниками информации, сверка фактического материала, установление частоты его использования другими источниками; выяснение статуса документа и рейтинга источника средствами поисковых систем, получение информации о компетентности и статусе автора материала с помощью специальных поисковых сервисов Интернета; анализ отдельных элементов организации сайта с целью

оценки квалификации поддерживающих его специалистов и другое).

Время поиска, не считая затрат времени, связанных с техническими характеристиками подключения, в основном зависит от планирования поиска и навыков работы специалиста по поиску с ресурсом выбранного типа. Планирование поиска заключается в определении требуемых для разрешения поискового требования поисковых служб и порядка их применения. Кроме того, многое зависит от навыков и опыта конкретного специалиста по поиску.

Как уже отмечалось, информация в Интернете доступна из источников разного типа. Прежде всего — это WWW-ресурсы (гипертекстовая система, каталоги ресурсов, поисковые машины). Кроме того, это уже известные читателю электронная почта, почтовые роботы, Usenet и другие телеконференции, а также ftp-системы и архивы (с применением Gopher и Veronica). WWW позволяет производить поиск требуемых ресурсов на основе своих гиперсвойств, то есть имеющиеся поисковые системы работают с использованием гиперссылок в автоматическом режиме, не исключая возможности ручного просмотра. В WWW имеется целый ряд поисковых сервисов как общего, так и специализированного назначения.

Каталоги ресурсов представляют собой базы данных с адресами ресурсов Интернета и самыми разными тематиками. Обычно они имеют иерархическую структуру, привычную для пользователя, и некоторые средства поиска по ней. Эти каталоги в большей своей части обслуживаются специалистами по классификации, то есть предопределяется некоторый субъективный подход к отбору информации, который, с одной стороны, несколько гарантирует достоверность информации, но с другой - предопределяет возможность отсутствия (пропуска) части информации, а также ее запоздалое размещение в каталоге.

Поисковые машины — это механизм автоматического построения ссылок (индексов) на различные ресурсы. Поисковые машины могут быть ориентированы на глобальные, специализированные или локальные ресурсы. По сути они являются мощными ИПС, которые с помощью специальных программ-роботов (так называемых "пауков") постоянно осуществляют автоматический поиск требуемой информации в Интернете. Созданные на этой основе специализированные БД обеспечивают поиск информации по запросам пользователей на основе специальных ИПЯ. Правда, охват просматриваемой информации зависит от применяемых алгоритмов и даже для мощных поисковых машин оставляет желать лучшего.

Электронная почта применяется в Интернете и в WWW. Адреса при этом попадают в поисковые системы и доступны поисковым машинам.

Почтовые роботы - это специальные программы, способные отвечать определенными действиями на команды, поступающие им, но электронной почте. Их основное назначение - пересылка данных по запросу в случае, когда те недоступны иным способом, а также как альтернатива работы в режиме online с каким-либо из известных ресурсов, например ftp-архивами. Адрес почтового робота имеет формат электронной почты. При поиске почтовые роботы обычно используются лишь как посредники при получении информации. Иногда приходится сталкиваться с тем, что они оказываются единственным средством получения нужных сведений.

Usenet и другие региональные и специализированные телеконференции представляют собой электронные "доски объявлений", где пользователь размещает свою информацию в одной из тематических групп новостей, передаваемых подписчикам соответствующей тематики. Этот ресурс наиболее значим для быстрого накопления информации, но узкому вопросу, а при поиске - чаще для получения частной, неофициальной информации.

Ресурсы, доступные по telnet, в ряде случаев представляют собой совершенно уникальную информацию, прежде всего по библиотечным каталогам европейских и американских университетов, а также государственных учреждений.

Как уже отмечалось, система файловых архивов ftp имеет достаточно обширные ресурсы ценной информации, до сих пор не переведенной в WWW. Архивы ftp представляют собой в первую очередь источники получения программного обеспечения. Поиск в них может представлять определенный интерес при знании структуры архивов; построения файловых систем, имен файлов и каталогов, содержащих требуемые ресурсы.

1. Требования к инструментам поиска

Как отмечалось ранее, чертами, присущими профессиональному поиску, являются его полнота, достоверность и высокая скорость. Наиболее серьезным и нетривиальным фактором, определяющим быстроту достижения цели поиска, оказывается планирование поисковой процедуры. Это требует, с одной стороны, выбора типа ресурсов, которые потенциально способны нести информацию, релевантную поисковой задаче, а с другой - выбора инструментов поиска, обслуживающих соответствующее информационное поле, в зависимости от их

предполагаемой результативности. Если говорить о наиболее емком на сегодняшний день с точки зрения информационного наполнения WWW-пространстве, то относительное изобилие его поисковых средств делает решение большинства практических задач многовариантным. Построение оптимальной последовательности применения тех или иных инструментов на каждом этапе поиска и предопределяет его эффективность. Помочь решить проблему выбора может четкое представление о видах, назначении и особенностях работы информационно-поисковых систем (ИПС) Интернета.

1. Планирование поиска

Поиск и сбор информации в Интернете нуждаются в планировании. Ошибочная логика построения запроса, неоптимизированная последовательность применения инструментов поиска, попытки ускорить поиск - все это не просто затягивает получение результата, но может поставить под угрозу смысл поисковой работы.

Остановимся на нескольких важных моментах, связанных с планированием и первыми шагами таких работ.

Начинать необходимо со всестороннего лексического анализа искомой информации. Следует использовать любое, достаточно достоверное и подробное описание исследуемого вопроса для получения первичных сведений. Таким источником вполне может стать как узкоспециальный справочник, так и электронная энциклопедия общего профиля. На основе изученного материала необходимо сформировать максимально широкий набор ключевых слов в виде отдельных терминов, словосочетаний, профессиональной лексики, сленга, слов-клише и устойчивых словесных штампов, при необходимости на нескольких языках. Заранее следует определить и возможные уточнения поискового запроса- редкие слова, синонимы и антонимы, названия и фамилии, тесно связанные с искомым вопросом. Желательно также заранее предусмотреть возможные нерелевантные отклики на запросы, то есть возможные характеристики поискового шума. После накопления этих предварительных данных можно перейти к получению первичной информации из Интернета.

Основная задача этой стадии учесть особенности Интернета, который является не только носителем технологий, но и традиций, и собственной этики. Сетевая лексика, сленг и написание общеупотребительных слов здесь могут отличаться от принятых.

Сведения о наличии в Интернете необходимых данных лучше всего искать в ранее известном каталоге, поддерживающем поиск по ключевым словам. При решении, например, простых задач типа "Получить текст Конституции Республики Беларусь" или "В каких правовых актах употребляется название родного города" известный сайт или каталог может быть более быстрым способом получения информации, чем автоматический индекс, и обеспечит большую достоверность.

После лексического анализа информации наступает технологический этап. Выбор информационного поля Интернета и поисковых инструментов производится на основе вышеизложенных подходов.

Используются тестовые запросы из одного-двух ключевых слов или фразы, затем анализируется количественный отклик. Содержательный анализ данных позволяет корректировать запросы, но релевантности отклика. В результате тестирования выясняются наиболее представительные источники информации, после чего следует уточнить последовательность применения поисковых инструментов. На этом этап планирования завершается.

В заключение отметим, что при решении задачи сбора информации из Интернета значительную роль играют региональные и специализированные поисковые сервисы. Применение глобальных индексов не для прямого поиска нужных сведений, а для локализации этих поисковых инструментов нередко позволяет сократить сроки решения поставленной поисковой задачи.

Заключение:

Принимая во внимания все выше сказанное, можно попытаться одним словом определить суть Интернета: это – общение, общение между отдельными людьми и целыми нациями без вмешательства правительственных авторитетов. Эта новая технология с огромной скоростью изменяет облик цивилизации, коренным образом меняя представление человечества о мире и самом себе. Сеть Интернет уже вобрала в себя десятки миллионов человек, более сотни стран, она полностью изменила процессы распространения и восприятия информации. В наш век информационных технологий виртуальная реальность интернет, способствующая стиранию государственных границ, сокращению географических расстояний, ликвидирующая преграды между культурами, становится не менее явственной, чем окружающий нас материальный мир.

С развитием INTERNET появилась возможность быстрого и удобного поиска необходимой документальной информации. Теперь можно не заниматься подбором

и изучением огромного количества литературы в книжных магазинах и библиотеках. Информацию можно получить, не выходя из дома или офиса. Для этого нужен только непосредственно сам компьютер, подключенный к INTERNET с установленной специальной программой – браузером, предназначенной для просмотра содержимого Web-страниц.

Благодаря разнообразию поисковых систем, специально разработанным для рядового пользователя, каждый может без труда отсеять заведомо ненужный поток информации, лишь правильно сформулировав цель поиска.