**Всемирная паутина www: гипертекст, поисковые машины**

Реферат выполнила: студентка 238 группы Мельник О. Н.

Санкт-петербургский государственный университет экономики и финансов

Санкт – Петербург 2009 г.

Всемирная Паутина - это совокупность информационных ресурсов, связанных средствами телекоммуникаций и основанных на гипертекстовом представлении данных, разбросанных по всему миру. Всемирная Паутина также обозначается как WWW (World Wide Web), W3 или просто Web.

Единицей гипертекстовых данных Всемирной Паутины является Web-страница - минимальный фрагмент гипертекста, который можно загрузить и прочитать за один раз. Web- страница содержит текст, графику и ссылки на другие Web- страницы.

Для идентификации гипертекстовой страницы в Сети используется так называемый URL - Universal Resource Locator - Универсальный Локатор Ресурсов. В нем содержится информация о способе передачи данных между клиентом и сервером (протоколе), имени сервера, на котором находится страница, пути в дереве каталогов сервера, по которому находится страница и имени страницы.

Для представления Web-страниц используется специальный язык разметки гипертекста – HTML (Hypertext Markup Language). Конструкции этого языка – тэги – позволяют управлять шрифтом, цветом текста и фона, определять ссылки, вставлять графику, аудио и видео и т.п. Сами тэги при чтении страницы не видны, а видны только результаты их действия. Если сравнить, как выглядит страница в исходном тексте и она же на экране браузера, то порой удивительно, насколько велики различия.

Клиента для WWW называют браузером. Браузер – это программа, которая загружает и, интерпретируя тэги HTML, показывает в своем окне WWW-страницы. Браузер отображает картинки в теле страницы, проигрывает аудио и видео, загружает следующую страницу, когда пользователь щелкает мышью по ссылке и пр. и пр. для того, чтобы сделать путешествие по Всемирной Паутине как можно более комфортным.

**Гипертекст и web-страницы**

Гипертекст (Hypertext) - это документ (в первую очередь текстовый), содержащий гиперссылки. Гиперссылка - это связь слова или содержащегося в документе изображения с другим ресурсом, которым может быть как еще один документ, так и раздел текущего документа. Подобные "связанные" слова или картинки документа, как правило, выделяются по оформлению из общего текста. Общепринятой является практика подчеркивания слова или предложения, связанного гиперссылкой.

Сам термин "гипертекст" появился с возникновением электронных документов, задолго до появления службы WWW. Поскольку современные электронные документы содержат не только текст, но и мультимедиа-информацию, понятие гипертекста было расширено до понятия гипермедиа. Гипермедиа - это метод организации мультимедиа-информации на основе ссылок на разные типы данных. Гипермедиа-документы могут использоваться не только в WWW, но и, например, в электронной энциклопедии на CD-ROM, где ссылка в тексте часто ведет на аудио- или видеоролик.

Особенно продуктивной идея гипертекста оказалась применительно к объединению цифровой информации, распределенной на серверах во всем мире. Документ, доступный через Web, называют Web-страницей, а группы страниц, связанных общим именем, темой и объединенных навигационно, - Web-сайтами. Первую страницу, которую видит пользователь при обращении на тот или иной ресурс, называют стартовой, домашней или индексной страницей (home page). Система гиперссылок определяет структуру Web-сайта. Страницы на сайте могут иметь линейную древовидную структуру, но чаще на каждой странице имеется несколько ссылок, что и позволяет говорить о структуре "паутина" (см. рис. 1)

Рисунок 1

**Поиск информации в Internet**

Как правило, последовательность поиска информации в Internet выглядит следующим образом:

Конкретизируется область поиска, выделяются ключевые слова, характерные для интересующей области. Рекомендуется по возможности подбирать такие слова, которые не используются ни в каких других областях деятельности или знания.

На поисковый сервер посылается запрос, сформированный из ключевых слов. На многих поисковиках имеется возможность конкретизировать область поиска путем выбора соответствующих пунктов предлагаемых меню.

Так или иначе, от сервера приходит список ссылок на WWW- страницы, в которых обнаружены запрошенные слова с кратким описанием каждой из них или просто с небольшим начальным фрагментом страницы.

Далее следует исследование подходящих (судя по заголовку или описанию) страниц. Здесь начинается “свободное плавание”, то есть переход по ссылкам с текстом, который обещает нечто интересное.

Сохранение на диске интересующего материала.

Отыскать нужную информацию в Седьмом океане - океане информации было бы просто невозможно если бы не существовало поисковых серверов. Такие серверы предоставляют возможность поиска страниц по ключевому слову или в иерархическом дереве каталога. Существуют как чисто англоязычные и русскоязычные, так и смешанные серверы. Документов на английском языке в Internet гораздо больше, чем на русском, поэтому имеет смысл посетить также и западные поисковики.

Для поиска информации по ключевому слову на поисковый сервер посылается запрос. Запрос формируется автоматически - нужно только ввести слова в поле ввода, выбрать нужные опции поиска и нажать кнопку. Обратно приходит Web-страница с отчетом о результатах поиска. Так как список найденных документов может быть очень большим, он разбивается на фрагменты по 10-100 элементов (на каждом сервере по-разному), а в конце отчета приводятся ссылки на следующие фрагменты списка.

Каждый элемент отчета формируется следующим образом:

- в начале идет заголовок документа являющийся также и ссылкой на этот документ (если у документа нет заголовка - - может быть и такое - - то здесь помещается надпись 'No title');

- далее следует небольшой фрагмент самого документа по которому практически всегда можно понять, о чем идет речь в документе и, следовательно, подходит он вам или нет;

- и, наконец, список URL данного документа (один и тот же документ может быть продублирован по разным адресам).

Чтобы перейти к документу, достаточно щелкнуть либо по ссылке с текстом заголовка, либо по одной из ссылок с URL. Кстати, единственный поисковик, который выдает пользователю не фрагмент документа, а настоящую аннотацию (на английском языке), написанную людьми - это Yahoo.

Второй метод отыскания документов, предоставляемый поисковыми серверами - - поиск по каталогам. Практически на всех поисковых машинах есть такие каталоги (исключение - - Altavista, Rambler). В таких каталогах вся область знаний разбивается на несколько крупных подобластей, например, искусство, бизнес, компьютеры, образование, спорт, политика, путешествия, здоровье и т.д. и т.п. Дале каждая такая подобласть разбивается на более мелкие разделы и подразделы. Таким образом в процессе поиска вы на каждом шаге будете уточнять искомое понятие, сужая область поиска, пока не получите список ссылок на документы, уже более соответствующих нужной вам теме.

Наиболее популярными англоязычными поисковиками являются: Yahoo, Magellan, Infoseek, Excite, WebCrawler, русскоязычными: Rambler, Русская Машина Поиска, всеязычным - - Altavista, Altavista Telia.

Классификация поисковых машин

Первая таблица включает в себя ведущие мировые поисковые машины, которые были первыми поисковыми службами в сети Интернет. С их помощью пользователи могут вести поиск, используя ключевые слова или логические конструкции. Большинство служб использует специальные языки запросов. Кроме того, можно воспользоваться уже существующей иерархией каталогов. Например, если вы ищете сайт определенной газеты, то нужно следовать по каталогу Yahoo! следующим образом:

News and Media >> Newspapers > The New York Times

Altavista и Yahoo! лидируют в этом списке. На этих сайтах существует обширная база данных и хорошо структурированная иерархия. Euroseek незаменим при поиске отдельных европейских ресурсов.

|  |
| --- |
| Глобальные поисковые машины |
| Altavista - http://www.altavista.com |
| Yahoo! - http://www.yahoo.com |
| Lycos - http://www.lycos.com/ |
| HotBot - http://www.hotbot.com |
| EuroSeek - http://www.euroseek.net/page?ifl=uk |
| Excite - http://www.excite.com |
| Infoseek - http://www.infoseek.com/ |

Поскольку все вышеперечисленные поисковые системы используют различные базы данных, то используя одни и те же ключевые слова, можно получить совершенно разные результаты. Иногда лучше обратится к так называемым системам "мета-поиска". Такие поисковые машины посылают ваш запрос на огромное количество разных поисковых систем. Затем обрабатывают полученные результаты, удаляют повторяющиеся адреса ресурсов и представляют более широкий спектр того, что представлено в сети Интернет.

|  |
| --- |
| Системы мета-поиска |
| Inference Find - http://www.inference.com/infind/ |
| Meta Find - http://www.metafind.com/ |
| Metacrawler - http://www.metacrawler.com/ |
| Webcrawler - http://www.webcrawler.com/ |
| Dogpile - http://www.dogpile.com/ |

 Иногда по определенной тематике поисковая система выдает такое огромное количество ссылок, что пользователь просто не в состоянии просмотреть их всех и найти лучшее. Northern Light и About.com (раннее известный как The Mining Company) используют услуги экспертов, которые вносят в базу данных самые лучшие ресурсы. Эти сайты очень информативны и представляют очень ценную информацию.

|  |
| --- |
| Специальные поисковые системы |
| Northern Light - http://www.nlsearch.com/ |
| About.com - http://about.com |

 Иногда вам необходима самая общая информация. Для этого можно воспользоваться онлайновыми энциклопедиями. В некоторых случаях статьи энциклопедии содержат ссылки на необходимые ресурсы.

|  |
| --- |
| Энциклопедии |
| Энциклопедия Британника - http://www.britannica.com/ |
| Кирилл и Мефодий - http://www.km.ru |

Из-за невероятного увеличения информационных ресурсов Интернет в последние годы некоторые старые поисковые системы перестали быть эффективными. Поэтому появились такие поисковые машины, как Google и Direct Hit, которые сортируют ссылки по принципу популярности. Используя математический алгоритм, они выдают адреса тех ресурсов, на которые чаще всего указывают другие сайты.

|  |
| --- |
| Поисковые системы второго поколения |
| Google - http://www.google.com/ |
| Direct Hit - http://www.directhit.com/ |

**Заключение**

Интернет, прочно войдя в нашу жизнь, смог за очень короткое время существенно изменить её. Сейчас интернет – наиболее быстрое, надежное средство общения и обмена информацией между людьми. С его помощью появилась возможность нахождения любой интересующей информации, без учета её места расположения, появилось такое свойство информации, как общедоступность в мировом масштабе.

**Список литературы**

Галанин С. «WWW – всемирная паутина». / http://iatp.ulstu.ru/edu/internet/www/

Прохоров А.Н. «Всемирная паутина (www)». / http://www.intuit.ru/department/office/od/7/

Поисковые машины Интернет. / http://www.pskov.org.ru/engines.html

Поисковая система. Википедия. /

http://ru.wikipedia.org/wiki/%D0%9F%D0%BE%D0%B8%D1%81%D0%BA%D0%BE%D0%B2%D0%B0%D1%8F\_%D1%81%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0

Всемирная паутина. / http://dic.academic.ru/dic.nsf/ruwiki/18318