Реферат

на тему:

«Методы поиска информации в сети интернет. Информационно-поисковые системы».

2010 г.

Содержание

Введение…………………………………………………………………...………3

1. Основные методы поиска информации в Интернете…………………......….4

1.1 Непосредственный поиск с использованием гипертекстовых ссылок….…4

1.2 Использование поисковых машин………………….......……………………4

1.3 Поиск с применением специальных средств…………….......…….………..5

1.4 Анализ новых ресурсов……………………………………….......…………..5

2. Поисковые системы………………………………………………….....………7

2.1 Использование поисковых систем на Западе……….......………….…….…7

2.2 История поисковых систем……………………………….......……………...8

3. Русскоязычные поисковые системы……………………………..……..……10

Заключение……………………………………………………………………….11

Список использованных источников…………………………………………...12

Введение

В наш ХХI век, владение информацией, говорит о том, что человек образован и респектабелен.

Более или менее серьезный подход к любой задаче начинается с анализа возможных методов ее решения.

Поиск информации в Интернете может быть произведен по нескольким методам, значительно различающимся как по эффективности и качеству поиска, так и по типу извлекаемой информации.

В ряде случаев приходится использовать весьма трудоемкие методы - результат того стоит.

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

В просторечии под поисковой системой подразумевают веб-сайт, на которой размещён интерфейс (фронт-энд) системы.

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

1. Основные методы поиска информации в Интернете

Можно выделить следующие основные методы поиска информации в Интернете, которые, в зависимости от целей и задач ищущего, используются по отдельности или в комбинации друг с другом:

1.1 Непосредственный поиск с использованием гипертекстовых ссылок

Поскольку все сайты в пространстве WWW фактически оказываются связанными между собой, поиск информации может быть произведен путем последовательного просмотра связанных страниц с помощью браузера.

Хотя этот полностью ручной метод поиска выглядит полным анахронизмом в Сети, содержащей более 60 млн. узлов, "ручной" просмотр Web-страниц часто оказывается единственно возможным на заключительных этапах информационного поиска, когда механическое "копание" уступает место более глубокому анализу.

Использование каталогов, классифицированных и тематических списков и всевозможных небольших справочников также относится к этому виду поиска.

1.2 Использование поисковых машин

Сегодня этот метод является одним из основных и фактически единственным при проведении предварительного поиска. Результатом последнего может являться список ресурсов Cети, подлежащих детальному рассмотрению.

Как правило, применение поисковых машин основано на использовании ключевых слов, которые передаются поисковым серверам в качестве аргументов поиска: что искать.

Если делать все правильно, то формирование списка ключевых слов требует предварительной работы по составлению тезауруса.

1.3 Поиск с применением специальных средств

Этот полностью автоматизированный метод может оказаться весьма эффективным для проведения первичного поиска.

Одна из технологий этого метода основана на применении специализированных программ - спайдеров, которые в автоматическом режиме просматривают Web-страницы, отыскивая на них искомую информацию.

Фактически это автоматизированный вариант просмотра с помощью гипертекстовых ссылок, описанный выше (поисковые машины для построения своих индексных таблиц используют похожие методы).

Нет нужды говорить, что результаты автоматического поиска обязательно требуют последующей обработки.

Применение данного метода целесообразно, если использование поисковых машин не может дать необходимых результатов (например, в силу нестандартности запроса, который не может быть адекватно задан существующими средствами поисковых машин).

В ряде случаев этот метод может быть очень эффективен.

Выбор между использованием спайдера или поисковых серверов являет собой вариант классического выбора между применением универсальных или специализированных средств.

1.4 Анализ новых ресурсов

Поиск по новообразованным ресурсам может оказаться необходимым при проведении повторных циклов поиска, поиска наиболее свежей информации или для анализа тенденций развития объекта исследования в динамике.

Другой возможной причиной может явиться то, что большинство поисковых машин обновляет свои индексы со значительной задержкой, вызванной гигантскими объемами обрабатываемых данных, и эта задержка обычно тем больше, чем менее популярна интересующая тема.

Это соображение может оказаться весьма существенным при проведении поиска в узкоспециальной предметной области.

2. Поисковые системы

Большинство поисковых систем ищут информацию на сайтах Всемирной паутины, но существуют также системы, способные искать файлы на ftp-серверах, товары в интернет-магазинах, а также информацию в группах новостей Usenet.

Основными критериями качества работы поисковой машины являются релевантность (степень соответствия запроса и найденного, то есть уместность результата), полнота базы, учёт морфологии языка.

Индексация информации осуществляется специальными поисковыми роботами.

В последнее время появился новый тип поисковых движков, основанных на технологии RSS, а также среди XML-данных разного типа.

Улучшение поиска — это одна из приоритетных задач сегодняшнего Интернета.

2.1 Использование поисковых систем на Западе

По данным компании Net Applications в декабре 2007 года использование поисковых систем на Западе распределялось следующим образом:

Google — 77,04 %

Yahoo — 12,46 %

MSN — 3,33 %

Microsoft Live Search — 2,57 %

AOL — 2,12 %

Ask — 1,38 %

AltaVista — 0,13 %

Excite — 0,07 %

Lycos — 0,02 %

All the Web — 0,02 %

В вышеприведенный отчёт не входят российские поисковики, такие как, например, Яндекс, Рамблер или Nigma.

По данным аналитической компании comScore все поисковые сайты в декабре 2007 года обработали 66 млрд 221 млн поисковых запросов.

Яндекс попал в статистику и находится на 9-ом месте.

2.2 История поисковых систем

Одним из первых инструментов поиска в интернете (до WWW) был Archie.

Первой поисковой системой для Всемирной паутины был «Wandex», уже не существующий индекс, который создавал «World Wide Web Wanderer» — бот, разработанный Мэтью Грэем (англ. Matthew Gray) из Массачусетского технологического института в 1993.

Также в 1993 году появилась поисковая система «Aliweb», работающая до сих пор.

Первой полнотекстовой, то есть индексирующей ресурсы при помощи робота, поисковой системой стала «WebCrawler», запущенная в 1994.

В отличие от своих предшественников, она позволяла пользователям искать по любым ключевым словам на любой веб-странице — с тех пор это стало стандартом во всех основных поисковых системах.

Кроме того, это был первый поисковик, о котором было известно в широких кругах.

В 1994 был запущен «Lycos», разработанный в университете Карнеги Мелона.

Вскоре появилось множество других конкурирующих поисковых машин, таких как «Excite», «Infoseek», «Inktomi», «Northern Light» и «AltaVista».

В некотором смысле они конкурировали с популярными интернет-каталогами, такими, как «Yahoo!».

Позже каталоги соединились или добавили к себе поисковые машины, чтобы увеличить функциональность.

В 1996 году русскоязычным пользователям интернета стало доступно морфологическое расширение к поисковой машине Altavista и оригинальные российские поисковые машины Rambler и Aport.

23 сентября 1997 была открыта поисковая машина Яндекс.

В последнее время завоёвывает всё большую популярность практика применения методов кластерного анализа и метапоиска.

Из международных машин такого плана наибольшую известность получила «Clusty» компании Vivísimo.

В 2005 году на российских просторах при поддержке МГУ запущен поисковик Nigma, поддерживающий автоматическую кластеризацию.

В 2006 году открылась российская метамашина Quintura, предлагающая визуальную кластеризацию в виде облака ключевых слов.

Nigma тоже экспериментировала с визуальной кластеризацией.

3. Русскоязычные поисковые системы

Большинство «русскоязычных» поисковых систем индексируют и ищут тексты на многих языках — украинском, белорусском, английском и др.

Отличаются же они от «всеязычных» систем, индексирующих все документы подряд, тем, что в основном индексируют ресурсы, расположенные в доменных зонах, где доминирует русский язык или другими способами ограничивают своих роботов русскоязычными сайтами.

Яндекс (46,3 % Рунета)

Mail.ru (8,9 % Рунета)

Rambler (3,3 % Рунета)

Nigma (0,5 % Рунета)

Генон (0,1 % Рунета)

Gogo.ru (<0,1 % Рунета)

Aport (<0,1 % Рунета)

Из перечисленных поисковых систем не все имеют собственный поисковый алгоритм — так QIP.ru используют поисковый механизм Яндекса, Mail.ru использует Google, а Nigma сочетает в себе как свой алгоритм, так и сборную выдачу от других поисковиков.

Заключение

Поиск, с применением поисковых машин, является самым распространенным и эффективным методом поиска чего-то конкретного в сети Интернет.

Хотя остальные методы ничем не хуже, только они применяются очень редко и только в том случае, если при помощи поисковой машины ничего нельзя найти.

Пользоваться поисковыми машинами удобно и легко.

Список использованных источников

1. http://www.ucheba.ru – УЧЕБА.RU

2. http://textbook.vadimstepanov.ru/index.html - ИНТЕРНЕТ В ПРОФЕССИОНАЛЬНОЙ ИНФОРМАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

3. http://www.proweb.ru/articles/oglavlenie/glava4/poisk - Группа компаний "Вебдом«

4. http://newsmir.ucoz.ru/load/1-1-0-9 - Лента новостей

5. http://simbiont-club.webzone.ru/search.htm - поисковые системы

6. http://www.seoded.ru/ssilki/poiskoviki.html - seoded.ru