

Нижегородский государственный педагогический университет имени Козьмы
Минина, факультет управления и социально-технических сервисов

Контрольная работа

Тема: «Организация эффективного поиска информации в Интернет»

Выполнил

Студент 1-го курса

Дувакин Артем Денисович

руководитель проекта

Круподерова Елена

Петровна

город Нижний Новгород

2023 год

ВВЕДЕНИЕ

В современном информационном обществе доступ к информации является ключевым фактором для успешного принятия решений, обучения и развития. Однако, объем информации, доступной в интернете, постоянно растет, и пользователи сталкиваются с проблемой эффективного поиска. Часто пользователи тратят много времени, не получая нужных результатов или заблудившись среди информационного мусора.. Между тем, есть простые приемы, которые позволяют искать нужные сведения в интернете быстрее и эффективнее. И даже если вы не начнете пользоваться всеми теми, о которых мы сегодня расскажем, все равно вам будет гораздо проще и легче заниматься интернет-серфингом – а ведь именно так и называется поиск информации во всемирной сети.

В наш XXI 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 В ИНТЕРНЕТЕ.

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.

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:

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 несколько десятков миллионов 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.

Использование поисковых машин. Сегодня этот метод является одним из основных и фактически единственным при проведении предварительного поиска. Результатом последнего может являться список ресурсов Интернет, 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 (словаря).

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 является собой вариант классического выбора между применением универсальных или специализированных средств.

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

2.АЛГОРИТМЫ ПОИСКА ИНФОРМАЦИИ

При разработке эффективного поискового алгоритма существует несколько подходов, которые могут быть использованы для оптимизации процесса нахождения нужной информации. Рассмотрим некоторые из них:

Линейный поиск

Линейный поиск — это простой алгоритм, который последовательно перебирает все элементы в наборе данных и сравнивает их с целевым значением. Если значение найдено, алгоритм возвращает его индекс. Если значение не найдено, алгоритм возвращает специальное значение, чтобы указать на отсутствие.

Бинарный поиск

Бинарный поиск — это алгоритм, который применяется к отсортированному набору данных. Алгоритм разделяет набор данных на две половины и сравнивает целевое значение с средним элементом. Если целевое значение равно среднему элементу, алгоритм возвращает его индекс. Если целевое значение меньше среднего элемента, алгоритм повторяет поиск в первой половине данных. Если целевое значение больше среднего элемента, алгоритм повторяет поиск во второй половине данных. Процесс продолжается до тех пор, пока целевое значение не будет найдено или не останется элементов для поиска.

Индексированный поиск

Индексированный поиск — это подход, который предполагает предварительное создание индекса всех элементов набора данных. Индекс представляет собой структуру данных, которая содержит связь между значением элемента и его расположением. При поиске алгоритм использует индекс для быстрого доступа к нужным элементам данных, минуя несущественные. Этот подход особенно полезен при поиске информации в больших наборах данных.

Алгоритмы поиска на основе хэш-таблиц

Алгоритмы поиска на основе хэш-таблиц используют структуру данных, известную как хэш-таблица. Хэш-таблица представляет собой специальную структуру данных, которая связывает ключи и значения. При поиске алгоритм использует хэш-функцию для вычисления хэш-значения ключа, которое идентифицирует индекс в хэш-таблице. Если ключ найден, алгоритм возвращает связанное значение. Алгоритмы поиска на основе хэш-таблиц обеспечивают быстрый доступ к информации, но требуют дополнительной памяти для хранения хэш-таблицы.

3. НАСТРОЙКА ПОИСКОВОЙ ИНДЕКСАЦИИ

Для эффективной работы поиска и быстрого нахождения нужных результатов необходимо правильно настроить поисковую индексацию. Рассмотрим основные моменты, которые помогут вам достичь оптимальных результатов.

1. Исключение ненужных страниц

Первым шагом в настройке поисковой индексации является исключение из поиска ненужных страниц. Например, вы можете исключить страницы с

формами обратной связи или страницы с личной информацией пользователей.

2. Настройка приоритета страниц

Далее, необходимо настроить приоритет страниц в поисковой индексации. Вы можете указать, какие страницы более важны для вашего сайта и должны отображаться в результатах поиска в первую очередь. Для этого можно использовать различные теги, такие как title или meta description.

3. Использование ключевых слов

Одним из важных аспектов в настройке поисковой индексации является использование ключевых слов. Вы можете анализировать популярные запросы пользователей и оптимизировать свои страницы под эти запросы. Важно использовать ключевые слова в заголовках, описаниях и тексте страницы.

4. Структурирование контента

Для улучшения работы поисковой индексации необходимо структурировать контент на вашем сайте. Используйте заголовки разных уровней (h1, h2, h3 и т.д.), маркированные и нумерованные списки (ul, ol) и другие элементы, которые помогут поисковым системам лучше понять содержание страницы.

5. Удобная навигация по сайту

Не менее важным аспектом в настройке поисковой индексации является удобная навигация по сайту. Обеспечьте четкую структуру внутренних ссылок на вашем сайте. Таким образом, поисковые системы смогут эффективно обходить и индексировать весь контент, а пользователи смогут легко перемещаться по сайту и находить нужную информацию.

4. 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.

5. 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 Грэм (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.

6. 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).

Лучшие русскоязычные поисковые системы

- Google - русская страница (<http://www.google.ru/>). Лучшая русскоязычная поисковая система. Ведёт поиск текста и внутри PDF-файлов. В режиме расширенного поиска может искать файлы заданного формата. Позволяет искать не только текстовые документы, но и картинки (файлы графических форматов) и карты местности (переключиться по соответствующей гиперссылке в верхней части страницы).
- Яндекс (<http://www.yandex.ru/>) - основной конкурент Google среди русскоязычных поисковых систем. Позволяет искать не только текстовые документы, но и картинки (файлы графических форматов) и карты городов.
- Yahoo! (<http://ru.yahoo.com/>) - лучшая в прошлом поисковая система общего назначения.
- Поиск@Mail.ru (<http://go.mail.ru/>) - поисковая система mail.ru. В последнее время становится реальным конкурентом перечисленным выше поисковым системам
- bing (<http://www.bing.com/>) - поисковая система Microsoft
- Nigma.ru (<http://nigma.ru/>)- научный проект "Интеллектуальная поисковая система Nigma.ru" МГУ им.М.В.Ломоносова, факультеты ВМиК и психологии - начат в 2005 году. Обеспечивает достаточно приличный поиск с интересными возможностями управления поиском.

- Рамблер (<http://www.rambler.ru/>) - несмотря на положительные отзывы в прессе автор ни разу не смог найти с помощью этой системы что-либо полезное. Это в основном каталог, а не поисковая система.

СПИСОК ЛИТЕРАТУРЫ

1. Иванова С.М. Система поиска и анализа достоверной информации в сети Интернет.
<https://cyberleninka.ru/article/n/sistema-poiska-i-analiza-dostovernoy-informatsii-v-seti-internet> 2016 год
2. Рошин С. Как быстро найти нужную информацию в Интернете.
<https://roschinsm.ru/books/src/book6.doc> 2010 год
3. Иванова С.М. Оценка достоверности информации, найденной в сети интернет.
<https://cyberleninka.ru/article/n/otsenka-dostovernosti-informatsii-naydennoy-v-seti-internet> 2015 год
4. Кутовенко, А.А. Профессиональный поиск в Интернете / А.А. Кутовенко. – Санкт-Петербург : Питер, 2011. – С. 12–34, 43–56
5. Шокин Ю.И. Проблемы поиска информации. Новосибирск:Наука, 2010. – 198 с