**Поиск информации с использованием компьютера**

Мишарева Елена Юрьевна

студентка 1 курса отделения «Строительство и эксплуатация зданий и сооружений»

ГБПОУ ВО «Калачеевский аграрный техникум»

научный руководитель преподаватель информатики Губанова Наталья Александровна.

**Содержание**

1. Введение 2
2. Информационные ресурсы сети Интернет 2
3. Типология методов поиска 3
4. Технология поиска с использованием поисковых машин 4
5. Владение пользовательскими инструментами и техникой 7
6. Заключение 8
7. Литература 8

**Введение**

Основная задача Интернета – предоставление необходимой информации. Интернет – это информационное пространство, в котором можно отыскать ответ практически на любой интересующий пользователя вопрос. Это огромная глобальная сеть, в которую как информационные ручейки стекаются потоки более мелких сетей. Любой пользователь, располагающий ПК и соответствующими программами, сможет подключиться к сети, используя её возможности для самых разных целей – проведения досуга, обучения, чтения научных работ, отправки электронной почты и т.д. По различным данным, в 2004 г. количество пользователей глобальной системы Интернет составило от 600 до 900 миллионов человек. Это число продолжает стремительно расти и уже в 2015 г. оно достигло 73, 8 миллиона человек. Сегодня глобальная сеть превращается в важный социальный и политический фактор современного информационного общества. С развитием Интернет-технологий появился новый гигантский источник информационных ресурсов, доступ к которым является не только относительно дешевым, но и очень быстрым.

**Информационные ресурсы сети Интернет**

Благодаря повсеместному развитию и применению компьютерных технологий в настоящее время в той или иной электронной форме находится информация всех областей человеческой деятельности: наука, производство, коммерция, литература, развлечения и т.д. Сеть Интернет имеет совместимость с различными электронными сетями и базами данных и позволяет получить удобный доступ практически к любому виду информации. Для миллионов людей разных профессий Интернет стал необходимым инструментом в работе и универсальным средством развлечения в часы досуга. В связи с этим возникает потребность в программных средствах, эффективно решающих проблемы поиска необходимой пользователю сети Интернет информации.

Информационные ресурсы, доступные через Интернет, огромны. Это десятки миллионов документов, представленных различными способами, число которых постоянно увеличивается. В зависимости от способа представления, вида и характера информации разнятся и методы доступа к ней, поэтому, прежде чем рассматривать методы поиска, рассмотрим классификацию информационных ресурсов.

По принципу организации и использования средства поиска можно разделить на каталоги (справочники, директории) и поисковые машины.

Каталоги

Каталоги являются справочниками, содержащими списки адресов Интернет, сгруппированные по определенным признакам. Как правило, они объединяются по тематике (наука, искусство, новости и т.д.), где каждая тема разветвляется на несколько подуровней.

Особенность этих средств поиска информации состоит в том, что создание структуры, базы данных и их постоянное обновление осуществляется "вручную", коллективом редакторов и программистов, и сам процесс поиска требует непосредственного участия пользователя, самостоятельно переходящего от ссылки к ссылке.

Поисковые машины

Действие поисковых машин заключается в постоянном последовательном исследовании всех узлов Интернет, доступных данной системе поиска, со всеми их связями и ответвлениями. В связи с постоянным обновлением информации машина поиска регулярно возвращается через определенный срок (порядка месяца) к уже изученным узлам, чтобы обнаружить и зарегистрировать изменения. Вся прочитанная информация индексируется, то есть создается специализированная база данных, в которой закодированы все исследованные системой страницы Интернет.

При поступлении запроса от пользователя машина поиска рассматривает всю индексированную информацию и выдает список документов, соответствующих задаче поиска. Найденные документы ранжируются в зависимости от местоположения ключевых слов (в заголовке, в начале текста, в первых параграфах) и частоты их появления в тексте.

Несмотря на схожий принцип работы, машины поиска различаются по языкам запроса, зонам поиска, глубине поиска внутри документа, методам ранжирования и приоритетов, поэтому применение разных поисковых машин дает различные результаты.

**Типология методов поиска**

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:

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.

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.

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.

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.

**Технология поиска с использованием поисковых машин**

Определение географических регионов поиска

Поскольку проведение информационного поиска преследует практические цели - практическая ценность информационного ресурса может зависеть от географического расположения соответствующего источника.

Составление тезауруса

Для эффективного использования поисковых серверов необходим список ключевых слов, организованный с учетом семантических отношений между ними, т.е. тезаурус. При составлении тезауруса необходимо предусмотреть обработку синонимов, омонимов и морфологических вариаций ключевых слов.

Использование законов Зипфа

Число, показывающее сколько раз встречается слово в тексте, называется частотой вхождения слова. Если расположить частоты по мере убывания и пронумеровать, то порядковый номер частоты называется ранг частоты. Вероятность обнаружения слова в тексте = частота вхождения слова / число слов в тексте. Зипф нашел, что если умножить вероятность обнаружения слова в тексте на ранг частоты, то получившаяся величина приблизительно постоянна для всех текстов на одном языке:

С = (частота вхождения слов X ранг частоты) / число слов

Это значит, что график зависимости ранга от частоты - равносторонняя гипербола.

Зипф также установил, что зависимость количества слов с данной частотой от частоты - также гипербола и постоянная для всех текстов в пределах одного языка.

Что можно извлечь из этих законов? Исследования вышеуказанных зависимостей для различных текстов показали, что наиболее значимые слова текста лежат в средней части диаграммы, так как слова с максимальной частотой как правило являются предлогами, частицами, местоимениями, в английском языке - артиклями (так называемые "стоп-слова"), а редко встречающиеся слова в большинстве случаев не имеют решающего значения. Основываясь на этой закономерности, можно предложить следующую методику.

Составление списка ключевых слов

Для составления оптимального набора ключевых слов используют процедуру, основанную на применении законов Зипфа, которая заключается в следующем: берут любой текст-источник, близкий к искомой теме, т.е. "образец", и анализируют его, выделяя значимые слова. В качестве текста-источника может служить книга, статья, Web-страница, любой другой документ. Анализ текста производится таким образом:

Удаление из текста стоп-слов.

Вычисление частоты вхождения каждого слова и составление списка, в котором слова расположены в порядке убывания их частоты.

Выбор диапазона частот, лежащего в середине списка, и отбор из этого диапазона слов, наиболее полно соответствующих смыслу текста.

Составление запроса к поисковой машине в форме перечисления отобранных таким образом ключевых слов, связанных логическим оператором ИЛИ (OR). Запрос в таком виде позволяет обнаружить тексты, в которых встречается хотя бы одно из перечисленных слов.

Число документов, полученных в результате поиска по этому запросу, может быть огромно. Однако, благодаря ранжированию документов (расположению их в порядке убывания частоты вхождения слов запроса в документ), применяемому в большинстве поисковых машин, на первых страницах списка практически все документы окажутся релевантными, причем документ-источник может находиться далеко от начала.

Более адекватной представляется структура тезауруса в виде так называемых семантических срезов, где для каждого основного термина отдельно строится таблица сопутствующих слов и слов шумовых (которые не должны встречаться в источнике), - некоторые поисковые машины (AltaVista) позволяют это использовать. Таким образом, вместо единой иерархической структуры терминов мы получаем пакет таблиц, которые могут расширяться и модифицироваться отдельно.

Отбор поисковых машин

Устанавливается последовательность использования поисковых машин в соответствии с убыванием ожидаемой эффективности поиска с применением каждой машины.

Всего известно около 180 поисковых серверов, различающихся по регионам охвата, принципам проведения поиска (а, следовательно, по входному языку и характеру воспринимаемых запросов), объему индексной базы, скорости обновления информации, способности искать "нестандартную" информацию и тому подобное. Основными критериями выбора поисковых серверов являются объем индексной базы сервера и степень развитости самой поисковой машины, то есть уровень сложности воспринимаемых ею запросов.

Составление и выполнение запросов к поисковым машинам

Это наиболее сложный и трудоемкий этап, связанный с обработкой большого количества информации (в основном шумовой). На основе тезауруса формируются запросы к выбранным поисковым серверам, после чего возможно уточнение запроса с целью отсечения очевидно нерелевантной информации. Затем производится отбор ресурсов, начиная с наиболее интересных, с точки зрения целей поиска. Данные с ресурсов, признанных релевантными, собираются для последующего анализа.

Формирование запросов

Как формат, так и семантика запросов варьируются в зависимости от применяемой поисковой машины и конкретной предметной области. Запросы составляются так, чтобы область поиска была максимально конкретизирована и сужена.

Предпочтение отдается использованию нескольких узких запросов по сравнению с одним расширенным. В общем случае для каждого основного понятия из тезауруса готовится отдельный пакет запросов. Так же производится пробная реализация запросов - как для уточнения и пополнения тезауруса, так и с целью отсечения шумовой информации.

Результат запроса (список ссылок) обрабатывается в два этапа. На первом этапе производится отсечение очевидно нерелевантных источников, попавших в выборку в силу несовершенства поисковой машины или недостаточной "интеллектуальности" запроса. Параллельно проводится семантический анализ, имеющий целью уточнение тезауруса для модификации последующих запросов. Дальнейшая обработка производится путем последовательного обращения на каждый из найденных ресурсов и анализа находящейся там информации.

Анализ ресурсов и сбор информации

Первичный анализ ресурсов основывается на аннотациях - в случае их наличия, и в необходимых случаях - на ознакомлении с информационным наполнением ресурса. Информация с отобранных источников извлекается с использованием соответствующих конкретному источнику методов, что может потребовать значительных коммуникационных, вычислительных и дисковых ресурсов.

В русскоязычной части Интернета в настоящее время доступен ряд ресурсов, предоставляющих вторичную информацию, как правило, в табулированной форме. Предоставление информации для публикации в подобных источниках является более дешевым вариантом для компаний, не имеющих собственного представительства в Интернете.

**Владение пользовательскими инструментами и техникой**

Следует не забывать о команде Find браузера. Если ваш браузер имеет в меню Файл, Правка или Вид команду Find (Найти), используйте ее для обнаружения трудноуловимых ключевых слов страницы. Чтобы отыскать слово, которое вам нужно, воспользуйтесь комбинацией клавиш CTRL+F в вашем браузере и введите искомое ключевое слово.

Необходимо так же фиксировать результаты ссылками и закладками:

1. Сохраняйте ссылки на важные и часто посещаемые страницы, используя команду Добавить в "Избранное". Используйте подобные механизмы для регистрации предварительных результатов поиска в процессе беглого отбора с целью дальнейшего подробного изучения.

2. Поддерживайте свою персональную коллекцию ссылок в рабочем состоянии: актуализируйте и систематизируйте ее, удаляйте устаревшие и ненужные.

3. Некоторые машины поиска, например AltaVista, позволяют отмечать закладкой успешный результат поиска. Благодаря этому позднее можно вернуться к данной подборке документов и просмотреть ее.

4. Существуют специализированные программные средства - интегрируемые модули расширения для стандартных браузеров, которые индексируют информационные ресурсы ранее посещаемых узлов. Они способны выполнять полнотекстовый поиск документов на любом узле, включая результаты обработки запросов.

Сохраняйте копии важных документов. Активно используйте команды контекстного меню Сохранить объект как… и Сохранить рисунок как… и команду меню Файл - Сохранить как… для создания копий необходимой вам информации из сети Интернет на локальном ресурсе вашего ПК. Этим вы решите некоторые проблемы:

1) изменчивости и недолговечности WWW (когда со своих адресов со временем исчезают страницы и целые сайты),

2) концентрации внимания исключительно на целях определенного этапа поиска (например отбор адресов или предварительный просмотр найденной информации),

3) экономии времени on-line подключения и перевод изучения отобранных материалов в режим off-line.

Отключайте вывод изображений. Если вы не желаете видеть многочисленные рекламные заставки и для вас важен не дизайн сайтов, а их содержательная информация, измените заданный режим просмотра в браузере на режим вывода только текста (исключив возможности мультимедиа). Это значительно увеличит скорость поиска. При этом, если возникнет необходимость быстрого просмотра важной графической информации, воспользуйтесь пунктом контекстного меню Показать рисунок, кликнув для его вызова правой кнопкой мыши в место не выведенного рисунка.

На сегодняшним момент самым распространенным браузером является Google Chrome

Google Chrome – это простой, быстрый и безопасный веб-браузер для современного Интернета.

Быстродействие

Google Chrome обеспечивает быструю работу в Интернете. Он легко открывается с рабочего стола, мгновенно загружает страницы и без промедления запускает сложные веб-приложения.

Удобство

Браузер Chrome имеет простой и рациональный интерфейс. Например, с помощью одного поля можно легко и быстро выполнять поиск и переходить на другие страницы, а также упорядочивать вкладки.

Безопасность

Браузер Chrome разрабатывался с целью повысить уровень безопасности при работе в Интернете. Это достигается за счет встроенной защиты от фишинга и вредоносного ПО, автоматических обновлений, гарантирующих своевременную установку исправлений для системы безопасности, а также многих других функций.

Конфиденциальность

Chrome помогает пользователям контролировать доступ к своим личным данным при работе в Интернете, а также защищает передаваемую информацию.

Персонализация

Существует множество способов настроить Chrome так, чтобы вам было удобно и приятно работать с ним. Это можно сделать с помощью специальных параметров, приложений, расширений и тем, доступных в Интернет-магазине Chrome.

Вход в систему

Войдя в Chrome, вы получите доступ к своим закладкам, истории веб-поиска и настройкам на любом компьютере, а также ко всем остальным сервисам Google.

**Заключение**

В сети Интернет хранится очень большой объем информации по различной тематике в виде статей в электронных газетах, отчетов, справочников, графических изображений, аудио- и видео-файлов и многого другого. Путешествуя по Интернет, можно найти любую информацию, не выходя из дома или офиса, иначе говоря, если какие-либо данные вводились когда-либо в компьютер, то вероятнее всего их можно найти где-то на необозримых просторах Интернет. Для этого нужен только непосредственно сам компьютер, подключенный к Интернет с установленной специальной программой – браузером, предназначенной для просмотра содержимого Web-страниц.

С появлением World Wide Web (WWW) - гипертекстовой информационной системы, опутавшей весь мир, связавшей сотни и сотни компьютеров, тысячи и тысячи страниц гипертекстовых документов - без высокоэффективных средств поиска просто не обойтись. Благодаря разнообразию поисковых систем, специально разработанным для рядового пользователя, каждый может без труда отсечь заведомо ненужный поток информации, лишь правильно сформулировав цель поиска.

Литература:

1. Информатика и ИКТ. Базовый уровень: Учебник для 10-11 классов / И.Г. Семакин, Е.К. Хеннер. - 6-е изд. – М.: БИНОМ. Лаборатория знаний, 2010. – 246 с.
2. bibliofond.ru
3. 5ballov.qip.ru