

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ КОЗЬМЫ МИНИНА
(МИНИНСКИЙ УНИВЕРСИТЕТ)»

Кафедра страхования, финансов и кредита
Дисциплина «Интернет технологии»

Реферат
на тему «Поиск информации в Интернет»

Выполнил студент группы
ЭФЗС-18-2
Факультет «ФУиСТС»
Абдуллина Надия Ряшитовна

Нижний Новгород

2019

СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ.....	3
2. ПОИСК ИНФОРМАЦИИ.....	4
2.1 Основы поисковых служб.....	4
2.2 Основные поисковые системы.....	5
2.3 Задачи и стратегии поиска.....	10
2.4 Средства поиска.....	11
2.5 Технологии и результат поиска.....	12
3. ЗАКЛЮЧЕНИЕ.....	14

ВВЕДЕНИЕ

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 2006 г. оно достигло миллиарда 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.

В настоящее время Интернет объединяет сотни миллионов серверов, на которых размещены миллиарды различных сайтов и отдельных файлов, содержащих различного рода информацию. Это гигантское хранилище информации. Существуют различные приемы поиска информации в Интернет.

2. ПОИСК ИНФОРМАЦИИ

Сеть интернет растет очень быстрыми темпами, и найти нужную информацию среди миллиардов Web-страниц и файлов становится все сложнее. Для поиска информации используются специальные поисковые серверы, которые содержат более или менее полную и постоянно обновляемую информацию о Web-страницах, файлах и других документах, хранящихся на десятках миллионов серверов Интернета.

Различные поисковые сервера могут использовать различные механизмы поиска, хранения и предоставления пользователю информации. Поисковые серверы Интернета можно разделить на две группы:

поисковые системы общего назначения;

специализированные поисковые системы.

Современные поисковые системы часто являются информационными порталами, которые предоставляют пользователям не только возможности поиска документов в Интернете, но и доступ к другим информационным ресурсам (новостям, информации о погоде, о валютном курсе, интерактивным географическим картам и так далее).

2.1 Основы поисковых служб

Сегодня пользователь World Wide Web оказывается в той же ситуации, что читатель крупной библиотеки. Чем больше фонды библиотеки, тем труднее найти именно ту книгу, которая сейчас нужна. Для упрощения поиска в библиотеке существуют каталоги: систематический, алфавитный, предметный и другие. Существуют и специализированные каталоги, например каталоги новых поступлений.

В сети поиск обеспечивают специальные поисковые службы. Умение пользоваться ими, собственно говоря, и составляет умение пользоваться Интернетом. Все поисковые системы WWW основаны на гиперссылках. Обращаясь к поисковой службе, мы формулируем запрос, в котором формально описываем то, что хотим найти. Проведя нужные операции, служба формирует Web-документ, состоящий из гиперссылок, ведущих к ресурсам WWW, соответствующим нашему запросу. Какой из этих гиперссылок мы воспользуемся - дело наше. Несмотря на то, что результат всегда един (клиент получает список рекомендованных гиперссылок), принцип действия у разных поисковых служб может быть различным. Ниже мы рассмотрим, как классифицируются поисковые службы по типам предоставляемых услуг, а также по способам формирования своих ресурсов.

Последнее очень важно для эффективной работы не только пользователей, но и Web-мастеров. Пользователи должны учитывать способ формирования баз данных, чтобы знать, на что они могут рассчитывать при поиске информации, а Web-мастера должны это знать, чтобы добиться наилучшего представления своих ресурсов в основных поисковых службах мира.

2.2 Основные поисковые системы

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

Поисковые системы уже давно стали неотъемлемой частью российского Интернета. Поисковые системы сейчас - это огромные и сложные механизмы, представляющие собой не только инструмент поиска информации, но и заманчивые сферы для бизнеса.

Большинство пользователей поисковых систем никогда не задумывались (либо задумывались, но не нашли ответа) о принципе работы поисковых систем, о схеме обработки запросов пользователей, о том, из чего эти системы состоят и как функционируют.

Основные поисковые системы:

AltaVista (<http://www.altavista.com/>)

По количеству индексируемых Web-страниц Alta-Vista -- одна из крупнейших (но не самая крупная, как это иногда считается) поисковых систем мира. Огромный объем охвата Web-пространства и мощный набор поисковых команд делают эту систему излюбленным средством поиска для большинства пользователей. Система была запущена в эксплуатацию в декабре 1995 г. и долгое время считалась молодой, но бурно развивающейся. Для начинающих система имеет упрощенную службу, которая называется Ask Alta Vista. В своей работе эта служба привлекает другую поисковую службу -- Ask Jeeves (см. ниже). Несмотря на то, что сегодня Alta Vista уже не является самым удобным средством поиска и владеет не самым большим индексом, ее

популярность по-прежнему выше, чем у служб, которые превзошли ее по многим показателям, что связано с устоявшимися привычками миллионов пользователей.

Ask Jeeves (<http://www.askjeeves.com/>)

Эта система представляет образец человеческого (неавтоматизированного) подхода к индексации и хранению информации о Web-документах. Ее цель не в том, чтобы выдать в ответ на запрос как можно больше возможных ссылок, в том, чтобы они были наиболее точны. В идеале система должна вас направить именно к той Web-странице, которая вам как раз и нужна. Система была введена в эксплуатацию в июне 1997 г. Результаты поиска в этой системе используются и при размещении запроса в Alta Vista.

Direct Hit (<http://www.directhit.com/>)

Систему Direct Hit можно назвать системой «рафинирования» результатов поиска, выдаваемых другими поисковыми системами. Она контролирует, какими результатами поисков клиенты пользуются наиболее часто и, тем самым, создает некий рейтинг популярности Web-страниц. Этот рейтинг она предоставляет другим поисковым системам, и они учитывают его при выдаче результатов. Технология Direct Hit напрямую используется поисковой системой HotBot. Она также учитывается системами Lycos, LookSmart и MSN Search.

Excite (<http://www.excite.com/>)

Excite -- одна из самых популярных поисковых систем World Wide Web. Она имеет указатель среднего размера и кроме поиска Web-страниц предоставляет услуги по поиску других материалов, например сведений о компаниях или результатов спортивных соревнований. Система была создана в конце 1995 г., прошла этап бурного развития и в 1996 г. поглотила двух ближайших конкурентов: систему Magellan и WebCrawler, хотя те по-прежнему сохраняют свои торговые марки и предоставляют поисковые услуги как самостоятельные службы.

FAST Search (<http://www.alltheweb.com/>)

Эта система, принадлежащая норвежской компании, ранее называлась All TheWeb. Несмотря на то, что она была запущена совсем недавно (в мае 1999 г.), перед ней ставится задача обогнать конкурентов в полноте коллекций ссылкой на Web-страницы. Кажется, ей это удастся. Летом 1999 г. FAST Search первое перешагнула 200-миллионный рубеж проиндексированных Web-документов. Кроме системы FAST Search та же компания

контролирует подразделение системы Lycos, занимающееся поиском музыкальных файлов в формате MP3.

Go/Infoseek (<http://www.go.com/>)

Go/Infoseek -- это сочетание известной в прошлом поисковой системы Infoseek (была основана в 1995 г.) и современного портала Go (создан в 1999 г. и принадлежит небезызвестной Walt Disney Co). Кроме поисковых возможностей Infoseek портал предоставляет зарегистрированным пользователям возможность персональной настройки среды, бесплатную электронную почту и другие полезные сервисы. Пользователи этой системы отмечают высокое качество ссылок, возвращаемых системой, что объясняется специальным «фирменным» алгоритмом! поиска. Кроме автоматизированного указателя система имеет также обширный каталог Web-узлов, составляемый вручную.

GoTo (<http://www.goto.com/>)

Поисковая служба GoTo не имеет отношения к portalу Go , и известна тем, что формирует результаты поиска на коммерческой основе. Компании, владеющие Web-узлами, за определенную плату могут договориться, чтобы ссылки на них появлялись на более высоких местах в списках итоговых результатов. Насколько известно, более никто, кроме Alta Vista, не занимается такими махинациями. Впрочем, служба GoTo не считает это махинацией, а полагает, что платежеспособность Web-издателя напрямую связана с качеством поставляемой им информации, и потому для пользователя такой подход полезен. При формировании списка служба берет оплаченные ссылки из собственной базы данных, а неоплаченные черпает от поисковой системы Inktomi (см. ниже).

Служба была запущена в 1997 г. В ее основу был положен Web-краулер World Wide Web Worm, приобретенный у разработчика -- университета штата Колорадо. С его помощью собиралась первичная база указателя. В 1998 г. служба перешла на нынешнюю «коммерческую» модель обслуживания и начала использовать систему Inktomi в качестве генератора «некоммерческих» ссылок.

Google (<http://www.google.com/>)

Служба Google известна тем, что при генерации списка ссылок в первую очередь выдает ссылки на те Web-страницы, к которым из других документов ведет наибольшее количество ссылок. Это дает очень интересный эффект. На большинстве поисковых систем совершенно бесполезно искать информацию по таким тривиальным словам как Cars (Автомобили), Internet, WWW, Games (Игры). Поиск по этим словам вернет столько

ссылок, что среди них невозможно найти самые полезные. Служба Google при проведении поиска по тривиальным словам дает отличные результаты, поскольку руководствуется мнением других пользователей, ранее осуществлявших подобный поиск.

HotBot (<http://hotbot.lycos.com/>)

По популярности среди сетевой общественности HotBot приближается к Alta Vista. Как и Alta Vista, эта служба имеет огромный поисковый указатель и много полезных функций. При генерации первой страницы ссылок HotBot использует результаты, поступающие от системы Direct Hit (см. выше). При генерации последующих страниц используется система Inktomi (см. ниже), услугами которой пользуются и другие поисковые службы. Кроме поиска по ключевым словам служба предоставляет также возможность поиска по тематическому каталогу. Данные для каталога черпаются из проекта Open Directory (см. ниже).

Служба HotBot была создана в мае 1996 г. компанией Wired Digital, которую в 1998 г. перекупила поисковая служба Lycos. Несмотря на это, HotBot продолжает функционировать как самостоятельная поисковая служба, параллельно с Lycos.

Inktomi (<http://www.inktorm.com/>)

Ядро поисковой системы Inktomi было разработано в университете Калифорнии в Беркли. Несколько позже создатели системы учредили одноименную компанию и сформировали указатель ссылок Inktomi, который был использован в качестве основы службой HotBot. Сегодня этот указатель используют и некоторые другие поисковые службы. Несмотря на то, что при ответе на запрос пользователя все подключенные службы обращаются к одной и той же базе данных, результаты выдаются разные. Это связано с тем, что хотя Inktomi и обеспечивает партнерам доступ к общей базе указателей, система всетаки различает партнеров и применяет системы рейтингования и фильтрации результатов в соответствии с их идеологией.

LookSmart (<http://www.looksmart.com/>)

Служба LookSmart -- ближайший последователь концепции Yahoo! в том, что касается «человеческого фактора» при составлении каталогов Web-ресурсов. Она предоставляет доступ к своему каталогу службе Alta Vista и другим партнерам. В тех случаях, когда поиск по архивам LookSmart не дает результата, ей помогает своими указателями Alta Vista. Служба LookSmart была создана в 1996 г. при финансовой

поддержке концерна Reader's Digest, но примерно через год ее учредители откупили ' службу, и сегодня она независима.

Lycos (<http://www.lycos.com/>)

Первоначально служба Lycos была запущена как поисковая система, основанная на программе-роботе, собирающей информацию из WWW. Запущенная в эксплуатацию еще в мае 1994 г., сегодня она считается одной из старейших поисковых служб. Однако совсем недавно, в 1999 г. служба перешла на другую модель, похожую на Yahoo! и основанную на каталогизации Web-узлов. Свои основные каталоги служба получает от проекта Open Directory, а дополнительные возможности поиска опираются на мониторинг WWW с помощью программ-роботов. В качестве альтернативной услуги Lycos ведет еще один каталог -- каталог Web-узлов Lycos Community Guides.

Название компании образовано из латинского словосочетания, которое можно перевести примерно как волкопаяк. В 1998 г. служба Lycos выкупила акции своего ближайшего конкурента HotBot, который продолжает автономную работу под общим управлением Lycos.

MSN Search (<http://search.msn.com/>)

Служба MSN Search, принадлежащая компании Microsoft была создана в 1998 г. и является порталом, на котором представлены и другие поисковые службы. В основу результатов поиска, выдаваемых службой, положен указатель, поставляемый системой Inktomi.

Netscape Search (<http://search.netscape.com/>)

Поисковая служба Netscape Search получает первичные результаты поиска от проекта Open Directory и собственной базы Smart Browsing. Вторичные результаты Search поставляются службой Google. На портале Netscape NetCenter кроме этой собственной службы присутствуют ссылки и на другие поисковые службы.

Yahoo (<http://www.yahoo.com/>)

Yahoo'--одна из старейших и наиболее популярных поисковых служб. Ее каталог был запущен в эксплуатацию еще в 1994 г. Секрет успеха Yahoo' в человеческом факторе». На службу работают около 150 редакторов, неустанно улучшающих и пополняющих множество тематических разделов огромною каталога. В каталоге

содержатся данные по одному миллиону Web-узлов. Если поиск по собственному каталогу не дает результата, привлекаются ресурсы поискового указателя Inktomi.

Приведем пример поиска(Попробуем с помощью российского поискового сервера Яндекс найти сайт "Информатика и информационные технологии".)

В браузере открыть начальную страницу поискового сервера Яндекс. В поле поиска ввести ключевые слова, например "информатика учебники тесты CD-ROM".

В результате по запросу найдено 54 тыс. ответов, некоторые не совсем по теме, поэтому нужно точнее вводить запрос

интернет интерфейс информация поиск

2.2.1 Отечественные поисковые системы

С большим удовлетворением должна отметить, что российский сектор Интернета развивается очень бурно. Сегодня ситуация такова, что даже если западный сектор по каким-либо причинам «прикажет долго жить», российский сектор достаточно автономен, чтобы развиваться самостоятельно и достойно обслуживать отечественных клиентов.

Важную роль в становлении и развитии отечественного сектора Интернета сыграли отечественные поисковые службы (за что им огромное спасибо). Не знаю, смогли ли они к сегодняшнему дню подойти к самоокупаемости, но в первые годы без сомнения работали на энтузиазме, доброй воле своих сотрудников и вере в будущее. В России есть как универсальные, так и специализированные поисковые службы. Как и везде в мире, в первую очередь развиваются поисковые указатели и рейтинговые системы, но имеется и полноценный каталог.

2.3 Задачи и стратегии поиска

Прежде чем начать поиск следует четко сформулировать задачу. До тех пор, пока задача не будет сформулирована об успешности, а тем более оптимизации поиска, не может быть и речи. Объем информации в Интернете слишком велик, и как нигде здесь верна поговорка «Время - деньги». Поэтому в здесь нельзя искать по принципу «Пойди туда, не знаю куда. Возьми то, не знаю что». Иначе результат будет плачевный. Дать общий рецепт эффективной стратегии поиска информации в Интернет, пожалуй, невозможно. Есть лишь некоторые принципы, позволяющие тратить меньше времени. Вот

некоторые из них. Например, вам необходимо узнать, где обитает трёхлапый ленивец, то вряд ли вы пойдете в алфавитный каталог библиотеки. Может быть, вы найдете нужную литературу с помощью систематического каталога. С несколько большей вероятностью - с помощью предметного. Но, скорее всего, ни один из библиотечных каталогов вам не поможет. Но, зайдя в информационно-библиографический отдел крупной библиотеки, дежурный библиограф достанет библиографический указатель по животным или какую-то похожую книжицу, из которой вы и найдете ответ на свой вопрос. Подобную стратегию можно с успехом применять и в Интернет. В ИПС(информационно-поисковых сетях) общего назначения можно утонуть в тысячах ссылок, выданных вам на простой запрос. Целью использования универсальной ИПС общего назначения может быть поиск специализированной ИПС, посвященной тематике вашего поиска. Такая ИПС может быть распознана по наличию слов "информация (information)", "документ (document)" и т.п. Но часто специализированная ИПС может скрываться на сервере общественной, профессиональной или специализированной организации, издательства. Этот пример иллюстрирует еще один элемент стратегии: читайте найденные документы в поисках наиболее точных терминов и связей между терминами. Возможно, вы мыслите в совершенно не тех терминах, которые используют авторы искомых документов (ведь большую роль играют культурные различия!).

Третий элемент стратегии: используйте несколько ИПС. Если вы регулярно занимаетесь поиском информации по какой-то тематике, отметьте те ИПС. которые для вас наиболее эффективны.

2.4 Средства поиска

По принципу организации и использования средства поиска можно выделить следующие инструменты:

- поисковые машины -- являются ключевым инструментом поиска информации, поскольку содержат индексы большинства web-серверов Интернета. Однако именно это достоинство оборачивается их главным недостатком. На любой запрос они выдают обычно чрезмерно большое количество информации, среди которой только незначительная часть является полезной, после чего требуется значительный объем времени для ее извлечения и обработки;

- мета-средства поиска -- позволяют ускорить выполнение запроса путем передачи аргументов поиска, то есть ключевых слов, одновременно нескольким поисковым системам. При значительном ускорении процесса и увеличении охвата поиска, этот способ имеет ряд недостатков, связанных с необходимостью координации во времени поступления результатов обработки запроса от нескольких систем, а также тем, что они не позволяют использовать возможности языка запроса каждого из применяемых поисковых средств;

- специализированные средства поиска -- представляют собой «программы-пауки», которые в автоматическом режиме просматривают web-страницы, отыскивая на них нужную информацию. Механизм их работы близок к механизму, который используют поисковые системы для построения своих индексных таблиц. Выбор между первыми и вторыми представляет собой классический выбор между применением универсальных или специализированных средств;

- каталоги -- как и поисковые машины, используются посетителями Интернета для нахождения необходимой информации. Каталог представляет собой иерархически организованную структуру, в которую данные заносятся по инициативе пользователей. Как следствие, объем информации в них несколько ограничен по сравнению с поисковыми системами, но в то же время более упорядочен благодаря лежащей в их основе иерархической тематической структуре.

2.5 Технологии и результат поиска

Поисковые средства и технологии, используемые для реализации информационных потребностей, определяются типом и состоянием решаемой пользователем задачи основной деятельности: соотношением его знания и незнания об исследуемом объекте. Кроме того, процесс взаимодействия пользователя с системой определяется уровнем знания пользователем содержания ресурса (полноты представления, достоверности источника и т.д.) и функциональных возможностей системы как инструмента. В целом эти факторы обычно сводятся к понятию "профессионализма" - информационного (подготовленный/неподготовленный пользователь) и предметного (профессионал/непрофессионал) "профессионализма". Процесс поиска информации обычно носит эмпирический характер. Он представляет последовательность шагов, приводящих при посредстве системы к некоторому результату, позволяющих оценить его полноту. При этом поведение пользователя, как организующее начало управления

процессом поиска, мотивируется не только информационной потребностью, но и разнообразием стратегий, технологий и средств, предоставляемых системой. Обычно пользователь не имеет исчерпывающих знаний об информационном содержании ресурса, в котором проводит поиск, поэтому оценить адекватность выражения запроса, как и полноту получаемого результата, он может, отыскав дополнительные сведения, или организовав процесс так, чтобы часть результатов поиска могла использоваться для подтверждения или отрицания адекватности другой части. Операционными объектами, непосредственно участвующими во взаимодействии пользователей с поисковой системой являются поисковый образ документа (ПОД) и ПОЗ, соответствие которых устанавливается поисковым механизмом АИПС на формальном уровне. Адекватность образа действительному содержанию документа определяется качеством процесса свертки информации и уровнем знания субъектом средств отражения - концептуальной схемы предметной области и возможностей ИПЯ.

Я ввела в поиске (Где сделать маникюр дешево). Все ссылки, найденные по теме, представлены в перечне ссылок, однако, как видно, соответствие теме работы нестрогое. В основном это либо сайты салонов (с образцами, прайсами и др.) либо сайты близкие с моей темой (например наращивание). Много информации вообще не по теме. Пришлось вводить запросы, состоящие из нескольких слов, запросы в кавычках, вводить города. Эти специальные методы поиска уменьшили количество выдаваемой информации (поисковые системы реагировали по-разному) и значительно облегчили задачу поиска. Не могу выделить особо какую-либо поисковую систему. Результаты поиска примерно одинаковы. Наиболее эффективной, является поисковая система Апорт из-за быстродействия и максимальной результативности. Системы Rambler и Яндекс почти по всем запросам выдают большое количество результатов (в большинстве случаев не относящихся к теме). В общем случае, считаю, без разницы какой системой пользоваться. Мне кажется, что человек при выборе поисковой системы больше ориентируется на свои привычки либо на общественное мнение (или, гораздо реже, исходит из своих потребностей). Как вывод - наиболее эффективными для поиска нужной тематики являются наборы словосочетаний в кавычках. Эффективность поиска возрастает иногда при вводе запроса из двух-трёх слов, а иногда - из четырёх-пяти и более. Поиск в современном интернете - поиск крупных необходимой информации среди рекламы и прайс-листов. Мне кажется, что было бы неплохо хранить всю научную информацию где-то в одном месте, недоступном для других материалов, но это невыполнимо.

ЗАКЛЮЧЕНИЕ

В заключении хотелось бы определить базовый алгоритм поиска.

Прежде всего, определите, к какому разделу относится искомая информация.

Определитесь какую именно поисковую машину лучше использовать. Учитывайте её специфику. Выявите основные слова (или словосочетания), которые характерны для той информации, которую ищите. Попробуйте выявить ключевые слова. Задавая отдельные слова и фразы, учитывайте язык запросов, операторы поиска, логические связи искомых слов и т.п. элементы, которые делают поиск значительно более эффективным. Если вам известна точная фраза из искомых материалов, используйте её как цитату.

После проведения предварительного поиска (или первой ступени поиска), если выдано довольно большое количество документов, воспользуйтесь элементом уточнения: "Искать в избранном". При этом задание поиска следует обязательно уточнить, введя хотя бы одно новое слово или ограничение, которое по вашему мнению поможет провести правильный отбор требуемого материала. Грамотно располагайте материал и проводите его сортировку по удобным для вас признакам. Это заметно ускорит последующую обработку материала. Фиксируйте все интересные для вас найденные адреса документов в Интернете (например, с помощью программы Блокнот). Затем выберите среди них нужные для выполнения конкретной работы.

Современному человеку необходимо уметь проводить поиск в Интернет, чтобы «не утонуть» в изобилии информации.

В этой работе я изучила и поняла типы, средства поисковых систем. Я научилась правильно проводить поиск в интернете.