Содержание:

Введение

Современный мир характеризуется использованием новых информационных технологий во всех сферах жизнедеятельности человека. Информация становится определяющим фактором развития общества. Все информационное пространство, в котором человек существует, все больше углубляется в Internet. С появлением глобальной информационной компьютерной сети появилась возможность оперативно получать информацию из любой точки земного шара. Самым распространенным средством информационных компьютерных технологий являются поисковые системы. Первые поисковые системы появились в сети Интернет более двадцати лет назад. В то время они реализовывали лишь функцию – поиска ссылок к недавно созданным страницам. На начальном этапе появления интернета, число пользователей сети было ограниченным, а количество информации относительно небольшим. Сегодня же поисковые системы превратились в многофункциональный сервис со своими службами. Они позволяют пользователям искать в сети Интернет самую разнообразную информацию, благодаря чему пользуются колоссальным спросом.

Проблема поиска и сбора сведений - одна из важных проблем поисковых систем. В двадцатом столетии, с зарождением века информационных технологий, проблема поиска информации приобрела новый облик. Сейчас она заключается не в том, что количества информации недостаточно и поэтому ее сложно отыскать, а в том, что теперь в обществе наблюдается ее переизбыток, с каждым днем, объем данных растет с геометрической прогрессией, и поэтому найти ответ на интересующий вопрос может оказаться совсем непростой задачей.

Проблема поиска информации существенно усложняется при использовании виртуальных источников. Здесь используется технология онлайновых каталогов, впоследствии использования которой, пользователь имеет право выполнять поиск в каталогах сразу двух или более библиотек, Тем самым, еще больше усложняет себе задачу, но, с другой стороны, увеличивает вероятностью ее решения.

Иными словами, в современном мире невозможно представить жизнь без Интернета, с его помощью мы приобретаем разнообразные продукты пользования,

общаемся, работаем, проводим с пользой свободное время. Возможности Всемирной Паутины безграничны, роль надежных гидов в виртуальных лабиринтах играют поисковые системы. Нет ничего проще, чем написать в строке поисковика нужный запрос, и поисковая система выдаст огромное количество предложений по внесенным словам или фразе. Еще сравнительно недавно, о чем-то подобном даже не догадывались.

Таким образом, актуальность проблемы обусловливается противоречием между большими потоками информации, циркулирующими в современном мире и неумением быстрого и качественного ее поиска в сети Интернет.

Актуальность определила тему курсовой работы – «Сравнение возможностей популярных информационно-поисковых систем».

Объект исследования –процесс поиска информации в современных поисковых системах сети Internet.

Цель исследования –определить сущность и значимость информационно-поисковых систем в современном обществе и выявить наиболее совершенную с точки зрения интерфейса и алгоритма поиска систему для пользователя.

В соответствии с поставленной целью были определены следующие задачи исследования**:**

- рассмотреть теоретические основы автоматизированного информационного поиска;
- описать классификации и разновидности современных поисковых систем;
- выявить преимущества и недостатки поисковых систем;
- провести сравнительный анализ современных поисковых систем.

Глава 1 Теоретические аспекты поисковых систем

1.1 Понятие информационно-поисковая система

Информационно-поисковая система - программно-аппаратный комплекс с веб-интерфейсом, предоставляющий возможность поиска информации в Интернете. Под поисковой системой обычно подразумевается сайт, на котором размещен интерфейс системы. Программной частью поисковой системы является поисковая

машина - комплекс программ, обеспечивающий функциональность поисковой системы и обычно являющийся коммерческой тайной компании - разработчика поисковой системы. Наиболее крупные международные поисковые системы: «Google», «Yahoo», «MSN», «Яндекс», «Рамблер».

Рассмотрим подробнее понятие поискового запроса. Для примера возьмем поисковую систему «Google» (рис. 1.1). Поисковый запрос нужно сформулировать пользователем в соответствии с тем, что он хочет найти, максимально кратко и просто. Допустим, мы хотим найти информацию в «Google» о том, как выбрать ноутбук. Для этого открываем главную страницу «Google» и вводим текст поискового запроса «как выбрать ноутбук». Однако мы можем и не найти нужную нам информацию. В таких случаях нужно перефразировать свой запрос, так как в базе поисковой системы может не оказаться информации по нашему запросу (такое может быть при задании очень «узких» запросов, как, например, «как выбрать ноутбук в Таласе»).

Главная задача поисковой системы - предоставлять людям именно ту информацию, которую они ищут. А научить пользователей делать «правильные» запросы к системе, т.е. запросы, соответствующие принципам работы поисковых систем, невозможно. Поэтому разработчики создают такие алгоритмы и принципы работы поисковых систем, которые бы позволяли находить пользователям искомую ими информацию.

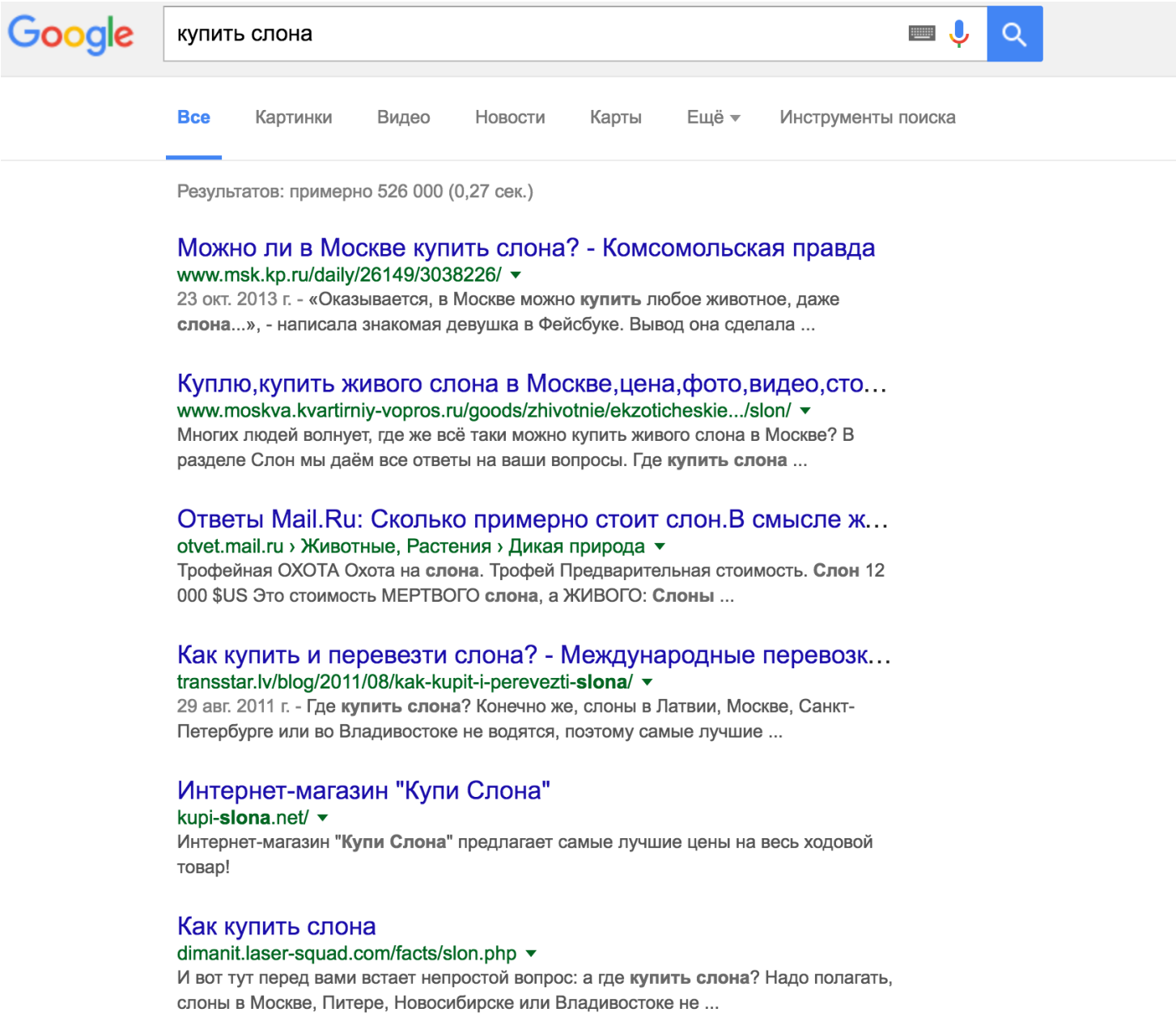


Рис. 1.1 - Поиск информации в «Google.ru»

Улучшение поиска - это одна из приоритетных задач современного Интернета. Разработчики поисковых систем постоянно совершенствуют алгоритмы и принципы поиска, добавляют новые функции и возможности, всячески пытаются ускорить работу системы /4/.

В начальный период развития Интернета число его пользователей было невелико, а объем доступной информации сравнительно небольшим. В большинстве своем доступ к сети Интернет имели лишь сотрудники научно-исследовательской сферы. В это время задача поиска информации в Интернете не была столь актуальной, как

в настоящее время.

Практически все крупные поисковые системы имеют свою собственную структуру, отличную от других. Однако можно выделить общие для всех поисковых машин основные компоненты. Различия в структуре могут быть лишь в виде реализации механизмов взаимодействия этих компонентов.

Поисковые системы (ПС) уже приличное время являются обязательной частью интернета. Сегодня они громадные и сложнейшие механизмы, которые представляют собой не только инструмент для нахождения любой необходимой информации, но и довольно увлекательные сферы для бизнеса.

Многие пользователи поиска никогда не думали о принципах их работы, о способах обработки пользовательских запросов, о том, как построены и функционируют данные системы. Данный материал поможет людям, которые занимаются оптимизацией ип родвижение своих сайтов, понять устройство и основные функции поисковых машин.

Поисковая система– это аппаратно-программный комплекс, который предназначен для осуществления функции поиска в интернете, и реагирующий на пользовательский запрос который обычно задают в виде какой-либо текстовой фразы (или точнее поискового запроса), выдачей ссылочного списка на информационные источники, осуществляющейся по релевантности. Самые распространенные и крупные системы поиска: Google, Bing, Yahoo, Baidu. В Рунете – Яндекс, Mail.Ru, Рамблер.

Рассмотрим поподробнее само значение запроса для поиска, взяв для примера систему Яндекс.

Запрос обязан быть сформулирован пользователем в полном соответствии с предметом его поиска, максимально просто и кратко. К примеру, мы желаем найти информацию в данном поисковике: «как выбрать автомобиль для себя». Чтобы сделать это, открываем главную страницу и вводим запрос для поиска «как выбрать авто». Потом наши функции сводятся к тому, чтобы зайти по предоставленным ссылкам на информационные источники в сети.

Но даже действуя таким образом, можно и не получить необходимую нам информацию. Если мы получили подобный отрицательный результат, нужно просто переформировать свой запрос, или же в базе поиска действительно нет никакой полезной информации по данному виду запроса (такое вполне возможно при

заданных «узких» параметров запроса, как, к примеру, «как выбрать автомобиль в Анадыри»).

Самая основная задача каждой поисковой системы – доставить людям именно тот вид информации, который им нужен. А приучить пользователей создавать «правильный» вид запросов к поисковым системам, то есть фразы, которые будут соответствовать их принципам работы, практически, невозможно.

Именно поэтому специалисты-разработчики поисковиков делают такие принципы и алгоритмы их работы, которые бы давали пользователям находить интересующие их сведения. Это означает, что система, должна «думать» так же, как мыслит человек при поиске необходимой информации в интернете.

Когда он вводит свой запрос в поисковую машину, он желает найти то, что ему надо, как можно проще и быстрее. Получив результат, пользователь составляет свою оценку работе системы, руководствуясь несколькими критериями. Получилось ли у него найти нужную информацию? Если нет, то сколько раз ему пришлось переформатировать текст запроса, чтобы найти ее? Насколько актуальная информация была им получена? Как быстро поисковая система обработала его запрос? Насколько удобно были предоставлены поисковые результаты? Был ли нужный результат первым, или находился на 30-ом месте? Сколько «мусора» (ненужной информации) было найдено вместе с полезными сведениями? Найдется ли актуальная для него информация, при использовании ПС, через неделю, либо через месяц?

Для того чтобы получить правильные ответы на подобные вопросы, разработчики поиска постоянно улучшают принципы ранжирования и его алгоритмы, добавляют им новые возможности и функции и любыми средствами пытаются сделать быстрее работу системы.

1.2 Архитектура современных ИПС для Интернета

Рассмотрим типовую схему информационно-поисковых систем Web (рис. 1.2).

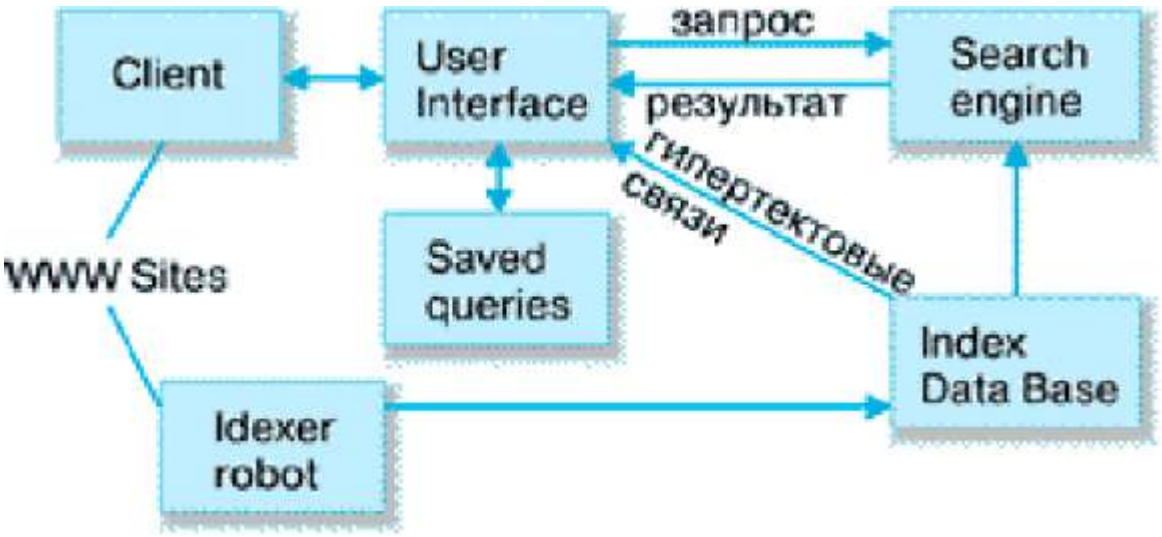


Рис. 1.2 - Типовая схема информационно-поисковой системы

Client (клиент) на этой схеме - это программа просмотра конкретного информационного ресурса (браузеры). В свою очередь, все эти информационные ресурсы являются объектом поиска информационно-поисковой системы.

User interface (пользовательский интерфейс) - это не просто программа просмотра. В случае информационно-поисковой системы под этим словосочетанием понимают также способ общения пользователя с поисковым аппаратом: системой формирования запросов и просмотров результатов поиска.

Search engine (поисковая машина) служит для трансляции запроса на информационно-поисковом языке (ИПЯ) в формальный запрос системы, поиска ссылок на информационные ресурсы сети и выдачи результатов этого поиска пользователю.

Index database (индекс базы данных) - индекс, который является основным массивом данных ИПС и служит для поиска адреса информационного ресурса. Архитектура индекса устроена таким образом, чтобы поиск проходил максимально быстро и при этом можно было бы оценить ценность каждого из найденных информационных ресурсов сети.

Queries (запросы пользователя) сохраняются в его (пользователя) личной базе данных. На отладку каждого запроса уходит достаточно много времени, и поэтому чрезвычайно важно запоминать запросы, на которые система дает нужные ответы.

Index robot (робот - индексирующий) - служит для сканирования Интернета и поддержания базы данных индекса в актуальном состоянии. Эта программа

является основным источником информации о состоянии информационных ресурсов сети.

WWW sites - это весь Интернет или, точнее, - информационные ресурсы, просмотр которых обеспечивается программами просмотра.

Рассмотрим назначение и принципы построения каждого из этих компонентов более подробно и определим, в чем отличие данной системы от традиционной ИПС локального типа.

1.3 Индекс поисковой системы

Индекс поисковой системы - это хранящаяся на поисковом сервере база данных, по которой осуществляется поиск запрошенной пользователем информации. Как правило, содержит ссылки на проиндексированные ресурсы и сжатые копии веб-страниц.

Копия страницы в индексе представляет собой инвертированный файл, где для каждого слова, имеющегося в исходном документе, перечислены позиции, в которых оно встречается. Индекс пополняется поисковым роботом во время периодических обходов Интернета.

Цель использования индекса - в повышении скорости поиска релевантных документов по поисковому запросу. Без индекса поисковая машина должна была бы сканировать каждый документ в корпусе, что потребовало бы большого количества времени и вычислительной мощности. Например, в то время, как индекс 10 000 документов может быть опрошен в пределах миллисекунд, последовательный просмотр каждого слова в 10 000 больших документов мог бы занять часы. Дополнительное хранилище, требуемое для хранения индекса, а также значительное увеличение времени, требуемого для его обновления, являются компромиссом за экономию времени при поиске информации. Эффективность поиска в каждой конкретной ИПС определяется исключительно архитектурой индекса.

Важным фактором является вид представления информации в программе-интерфейсе. При этом различают два типа интерфейсных страниц: страницы запросов и страницы результатов поиска.

При составлении запроса к системе используют либо меню-ориентированный подход, либо командную строку. Меню-ориентированный подход позволяет ввести список терминов, обычно через пробел, и выбрать тип логической связи между ними. Логическая связь распространяется на все термины. В большинстве систем это просто фраза на ИПЯ, которую можно расширить за счет добавления новых терминов и логических операторов. Но это только один тип использования сохраненных запросов. В традиционных системах это называется расширением или уточнением запроса, в зависимости от того, что получаем в результате преобразования запроса: увеличение размера выборки или ее сокращение. При этом традиционная система хранит не запрос как таковой, а результат поиска, т.е. список идентификаторов документов, который объединяется, пересекается со списком, полученным при поиске документов по новым терминам. К сожалению, сохранение списка идентификаторов найденных документов в Интернете не практикуется. Вызвано это особенностью протоколов взаимодействия программы-клиента и сервера системы, которые не поддерживают сеансовый режим работы.

Информационно-поисковый язык (ИПЯ) - искусственный язык, предназначенный для выражения семантических аспектов информационных источников и запросов в форме, пригодной для осуществления поиска информации. По своим знаковым системам и правилам синтаксиса ИПЯ различаются.

Процесс поиска информации предусматривает взаимодействие в режиме «запрос - ответ» пользователя и информационно-поисковой системы через посредство заранее согласованного ИПЯ. Таким образом, предпосылками для проведения информационного поиска являются:

а) предварительное индексирование информационного массива, т.е. создания поискового образа каждого информационного источника в массиве;

б) перевод информационного запроса пользователя определенного ИПЯ.

Информационно-поисковые языки делятся на два основных типа:

1. ИПЯ классификационного типа.

К языкам этого типа относятся иерархические, алфавитно-предметные и фасетные классификации.

1. ИПЯ дескрипторного типа

Словарь такого языка состоит из фиксированного набора слов и словосочетаний одной или нескольких естественных языков. Таким образом, индексирование информационного источника предполагает создание его поискового образа как определенного набора слов и словосочетаний, которые характеризуют его ключевые содержательные признаки. Методы полнотекстового поиска информации, в основном, предусматривают использование ИПЯ дескрипторного типа.

Глава 2 Анализ поисковых систем

2.1 История развития поисковых систем

Когда интернет только начал развиваться, число его постоянных пользователей было небольшим, и объем информации для доступа был сравнительно невеликим. В основном доступ к этой сети имели лишь специалисты научно-исследовательских сфер. В то время, задача нахождения информации не была столь актуальна как сейчас.

Одним из самых первых методов организации широкого доступа к ресурсам информации стало создание каталогов сайтов, причем ссылки на них начали группировать по тематике. Таким первым проектом стал ресурс Yahoo.com, который открылся весной 1994-ого года. Впоследствии когда количество сайтов в Yahoo-каталоге существенно увеличилось, была добавлена опция поиска необходимых сведений по каталогу. Это еще не было в полной мере поисковой системой, так как область такого поиска была ограничена только сайтами, входящими в данный каталог, а не абсолютно всеми ресурсами в интернете. Каталоги ссылок весьма широко использовались раньше, однако в настоящее время, практически в полной мере утратили свою популярность.

Ведь даже сегодняшние, громадные по своим объемам каталоги имеют информацию о незначительно части сайтов в интернете. Самый известный и большой каталог в мире DMOZ имеет информацию о пяти миллионах сайтов, когда база Google содержит информацию о более чем 25 миллиардов страниц.

Самой первой настоящей поисковой системой стала WebCrawler, возникшая еще в 1994-ом году.

В следующем году появились AltaVista и Lycos. Причем первая была лидером по поиску информации очень длительное время.

В 1997-ом году Сергей Брин вместе с Ларри Пейджем создал машину поисковую Google как исследовательский проект в Стэндфордском университете. Сегодня именно Google, самая востребованная и популярная поисковая система в мире.

В сентябре 1997-ом году была анонсирована (официально) ПС Yandex, которая в настоящий момент является самой популярной системой поиска в Рунете.

2.2 Основные характеристики поисковых систем

Полнота является одной из главнейших характеристик поиска, она представляет собой отношение цифры найденных по запросу информационных документов к их общему числу в интернете, относящихся к данному запросу. Например, в сети есть 100 страниц имеющих словосочетание «как выбрать авто», а по такому же запросу было отобрано всего 60 из общего количества, то в данном случае полнота поиска составит 0,6. Понятно, что чем полнее сам поиск, тем больше вероятность, что пользователь найдет именно тот документ, который ему необходим, конечно, если он вообще существует.

Еще одна основная функция поисковой системы – точность. Она определяет степень соответствия запросу пользователя найденных страниц в Сети. К примеру, если по ключевой фразе «как выбрать автомобиль» найдется сотня документов, в половине из них содержится данное словосочетание, а в остальных просто есть в наличии такие слова (как грамотно выбрать автомагнитолу, и установить ее в автомобиль»), то поисковая точность равна 50/100 = 0,5.

Чем поиск точнее, тем скорее пользователь найдет необходимую ему информацию, тем меньше разнообразного «мусора» будет встречаться среди результатов, тем меньше найденных документов будут не соответствовать смыслу запроса.

Актуальность - это значимая составляющая поиска, которую характеризует время, проходящее с момента опубликования информации в интернете до занесения ее в индексную базу поисковика.

К примеру, на следующий день после возникновения информации о выходе нового iPad, множество пользователей обратилась к поиску с соответствующими видами запросов. В большинстве случаев информация об этой новости уже доступна в

поиске, хотя времени с момента ее появления прошло очень мало. Это происходит благодаря наличию у крупных поисковых систем «быстрой базы», которая обновляется несколько раз за день.

Скорость поиска - такая функция как скорость поиска теснейшим образом связана с так называемой «устойчивостью к нагрузкам». Ежесекундно к поиску обращается огромное количество людей, подобная загруженность требует значительного сокращения времени для обработки одного запроса. Тут интересы, как поисковой системы, так и пользователя целиком совпадают: посетитель хочет получить результаты как можно быстрее, а поисковая система должна отработать его запрос тоже максимально быстро, чтобы не притормозить обработку последующих запросов.

Наглядное представление результатов является важнейшим элементом удобства поиска. По множеству запросов поисковая система находит тысячи, а в некоторых случаях и миллионы разных документов. Вследствие нечеткости составления ключевых фраз для поиска или его не точности, даже самые первые результаты запроса не всегда имеют только нужные сведения.

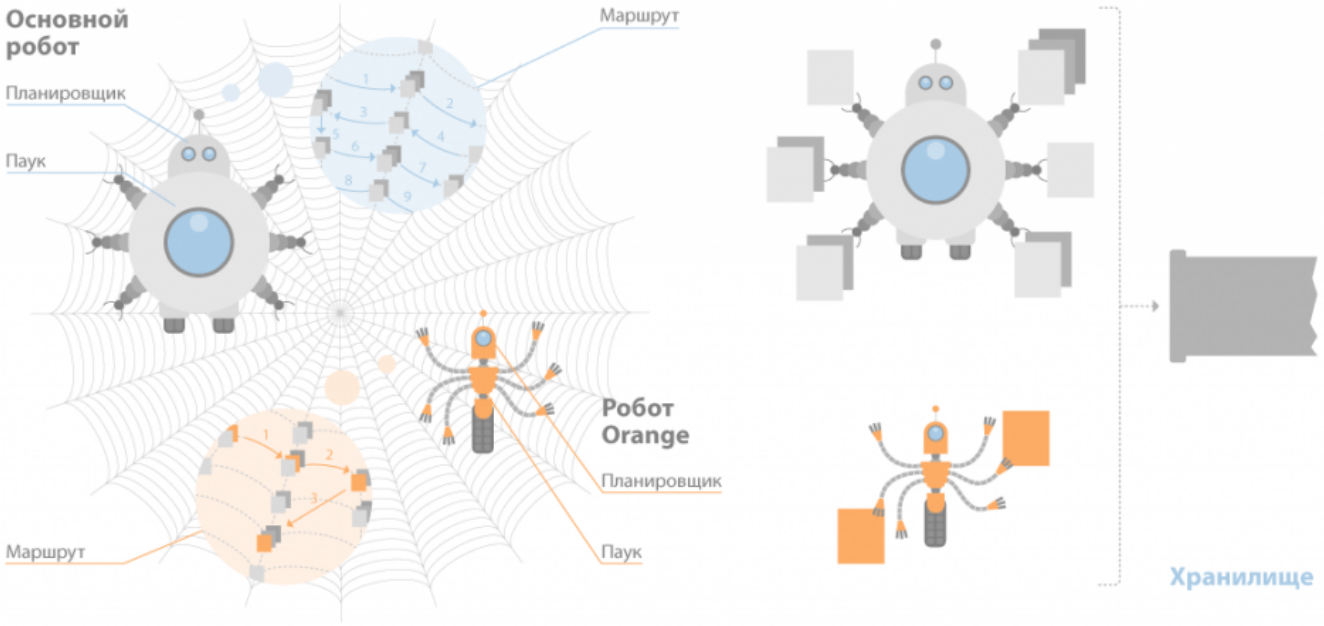
Это значит, что человеку часто приходится осуществлять собственный поиск среди предоставленных результатов. Разнообразные компоненты страниц выдачи ПС помогают ориентироваться в поисковых результатах.

2.3 Принципы работы поисковой системы

В России главной системой поиска является Яндекс, затем Google, а потом Поиск@Mail.ru. Все большие системы поиска имеют свою структуру, которая весьма отличается от других. Но все-таки можно выделить общие для всех поисковиков основные элементы.

Данный компонент состоит из трех программ-роботов:

Spider(по англ. паук) – программа которая предназначена для того чтобы скачивать веб-страницы. «Паук» скачивает определенную страницу, одновременно извлекая из нее все ссылки. Скачивается код html практически с каждой страницы. Для этого роботы используют HTTP-протоколы.


Основной робот
Планировщик
Паук
Маршрут
Робот Orange
Планировщик
Паук
Маршрут
Хранилище

Поисковые системы (ПС) уже приличное время являются обязательной частью интернета. Сегодня они громадные и сложнейшие механизмы, которые представляют собой не только инструмент для нахождения любой необходимой информации, но и довольно увлекательные сферы для бизнеса.

Многие пользователи поиска никогда не думали о принципах их работы, о способах обработки пользовательских запросов, о том, как построены и функционируют данные системы. Данный материал поможет людям, которые занимаются оптимизацией ип родвижение своих сайтов, понять устройство и основные функции поисковых машин.

Поисковая система– это аппаратно-программный комплекс, который предназначен для осуществления функции поиска в интернете, и реагирующий на пользовательский запрос который обычно задают в виде какой-либо текстовой фразы (или точнее поискового запроса), выдачей ссылочного списка на информационные источники, осуществляющейся по релевантности.

Рассмотрим поподробнее само значение запроса для поиска, взяв для примера систему Яндекс.

Запрос обязан быть сформулирован пользователем в полном соответствии с предметом его поиска, максимально просто и кратко. К примеру, мы желаем найти информацию в данном поисковике: «как выбрать автомобиль для себя». Чтобы сделать это, открываем главную страницу и вводим запрос для поиска «как выбрать авто». Потом наши функции сводятся к тому, чтобы зайти по предоставленным ссылкам на информационные источники в сети.

Но даже действуя таким образом, можно и не получить необходимую нам информацию. Если мы получили подобный отрицательный результат, нужно просто переформировать свой запрос, или же в базе поиска действительно нет никакой полезной информации по данному виду запроса (такое вполне возможно при заданных «узких» параметров запроса, как, к примеру, «как выбрать автомобиль в Анадыри»).

Самая основная задача каждой поисковой системы – доставить людям именно тот вид информации, который им нужен. А приучить пользователей создавать «правильный» вид запросов к поисковым системам, то есть фразы, которые будут соответствовать их принципам работы, практически, невозможно.

Именно поэтому специалисты-разработчики поисковиков делают такие принципы и алгоритмы их работы, которые бы давали пользователям находить интересующие их сведения. Это означает, что система, должна «думать» так же, как мыслит человек при поиске необходимой информации в интернете.

Рис. 2.1 – Схема работы

«Паук» функционирует следующим образом. Робот передает запрос на сервер “get/path/document” и иные команды запроса HTTP. В ответ программа-робот получает поток текста, который содержит информацию служебного вида и, естественно, сам документ.

Извлекаются все ссылки из тэгов. Вместе с ними обрабатывают редиректы. Любая скачанная страница сохраняется в таком формате:

- URL скаченной страницы;
- дата, когда осуществлялось скачивание страницы;
- заголовок http-ответа сервера;
- html-код, «тела» страницы.

Crawler(«путешествующий» паук). Данная программа автоматически заходит на все ссылки, которые найдены на странице, а также выделяет их. Его задача – определиться, куда в дальнейшем должен заходить паук, основываясь на этих ссылках или исходя из заданного списка адресов.

Crawler, исследуя найденные ссылки, ищет новые документы, еще не ставшие известными поисковой системе.

Indexer(робот-индексатор) – это программа, анализирующая страницы, которые скачали пауки.

Рис. 2.2 – Схема работы поискового робота

Индексатор полностью разбирает страницу на составные элементы и проводит их анализ, применяя свои морфологические и лексические виды алгоритмов.

Анализ проводится над разнообразными частями страницы, такими как заголовки, текст, ссылки, стилевые и структурные особенности, теги html и др.

Таким образом, модуль индексирования дает возможность проходить по ссылкам заданного количества ресурсов, скачивать страницы, извлекать ссылочную массу на новые страницы из полученных документов и делать подробный их анализ.

База данных (или индекс поисковика) - комплекс хранения данных, массив информации в котором сохраняются определенным образом переделанные параметры каждого обработанного модулем индексации и скачанного документа.

Поисковой сервер - это самый важный элемент всей системы, потому что от алгоритмов, лежащих в основе ее функциональности, прямо зависит скорость и, конечно же, качество поиска.

Поисковый сервер работает следующим образом:

- Запрос, который идет от пользователя подвергается морфологическому анализу. Информационное окружение любого документа, имеющегося в базе, генерируется (оно и будет в дальнейшем отображаться как сниппет, т.е. информационное поле текста соответствующего данному запросу).
- Полученные данные передают как входные параметры специализированному модулю ранжирования. Они обрабатываются по всем документам, и в итоге для каждого такого документа рассчитывается свой рейтинг, который характеризует релевантность такого документа запросу пользователя, и иных составляющих.
- В зависимости от условий заданных пользователем этот рейтинг вполне может быть подкорректирован дополнительными.
- Затем генерируется сам сниппет, т.е. для любого найденного документа из соответствующей таблицы извлекают заголовок, аннотацию, наиболее отвечающую запросу, и ссылка на этот документ, при этом найденные словоформы и слова подсвечивают.
- Результаты полученного поиска передаются осуществившему его человеку в виде страницы, на которую выдают поисковые результаты (SERP).

Все эти элементы тесно связаны между собой и функционируют, взаимодействуя, образовывая отчетливый, но достаточно непростой механизм функционирования ПС, требующий громадных затрат ресурсов.

Глава 3 Сравнительный анализ поисковых систем

3.1 Обзор популярных мировых и российских информационно-поисковых систем

Рейтинг мировых и российских информационно-поисковых систем, поможет нам выявить наиболее популярные поисковые системы, которые в дальнейшем мы будем рассматривать.

Google первая по популярности поисковая машина в мире обрабатывающая более 40 миллиардов запросов в месяц (доля рынка 83,4 %), и индексирует более 8 миллиардов веб-страниц. Google может находить информацию на 191 языке (на 15 октября 2012) [15]. Второе место (с большим отрывом) у поисковой системыYahoo! – 6,32% рынка. Третье место занимает крупнейший китайский поисковик Baidu.com – 4,96% рынка[20]. Уверенные позиции последнего связаны с тем, что на территории Китая заблокированы и Google, и Yahoo. Четвертое место занимает Bing(MSN),она является относительно молодой поисковой системой от Microsoft, её успех главным образом определяетсяогромным массивом статистических данных, который накопился у компании за годы существования браузера InternetExplorer, который в дальнейшем позволил ее инженерам создать поисковой алгоритм, дающий пользователямрелевантную выдачу.[27]

Лидер поисковых машин Интернета, Google занимает более 70% мирового рынка, а значит, семь из десяти находящихся в сети людей обращаются к его странице в поисках информации в Интернете. Сейчас регистрирует ежедневно около 50 миллионов поисковых запросов и индексирует более 8 миллиардов веб-страниц [9].

Информационно-поисковая система Google быларазработана в 1998 выпускниками Стэндфордского университета Сергеем Брином и Лари Пейджем, которые применили для ранжирования документов технологию PageRank, где одним из ключевых моментов является определение «авторитетности» конкретного документа на основе информации о документах, ссылающихся на него. Говоря

общими словами, чем больше документов ссылается на данный документ, и чем они авторитетнее, тем авторитетнее становитсяданный документ. Количественное значение авторитетности документа (другими словами, взвешенное количество ссылок или PageRank) относится к так называемым статическим факторам (т.е. независящим от конкретного запроса) и учитывается при определении релевантности документа конкретному запросу как весовой коэффициент. Наряду с этим Google применил для определения релевантности документа не только текст самого документа, но и текст ссылок на него. Эта технология позволила ему обеспечить выдачу довольно релевантных результатов на фоне других поисковиков. Довольно быстро Google стал лидировать в различных опросах по такому показателю, как удовлетворенность пользователей результатами поиска. Google осуществляет поиск по документам на более чем 35 языках, в том числе русском. В настоящее время многие порталы и специализированные сайты предоставляют услуги поиска информации в Интернете на базе Google, что делает задачу успешного позиционирования сайтов в Google еще более важной. Google проводит переиндексацию своей поисковой базы примерно раз в четыре недели. Во время этого усовершенствования, неофициально называемого Googledance, происходит обновление базы на основе информации, собранной роботами за время, прошедшее с предыдущего усовершенствования, и перерасчет значений PageRank документов [15].

Также существует определенное количество документов с достаточно большим значением PageRank, информация о которых в поисковой базе обновляется ежедневно, однако значение PageRank пересчитывается только во время Googledance. Нормированное значение PageRank для конкретного документа, загруженного в браузер, можно узнать, скачав и установив GoogleToolBar - специальную панель инструментов для работы с этим поисковиком. Не смотря на то, что в поисковике имеется форма для бесплатного добавления страницы в базу, Google предпочитает сам находить новые документы по ссылкам с уже известныхстраниц,и не будет индексировать добавленную через форму страницу, если в его базе не найдется ни одной страницы, ссылающейся на нее.

Так же на страницах результатов поиска Google отображаются платные (payperclick) рекламные объявления конкурирующих компаний, которые основывают рекламные объявления на брендах. «В то время как сервис мог бы помочь увеличить трафик, некоторые пользователи «сливаются», так как Google использует известность брендов для продажи рекламных объявлений, как правило, конкурирующим компаниям». Чтобы сгладить этот конфликт Google предложил

отключать эту возможность для желающих компаний.

Поисковая технология, позволяющая пользователю настраивать результаты выдачи по поисковым запросам. Пользователь может удалять результаты из списка и поднимать вверх списка. Технология была запущена компанией Google весной 2009 года и проработала до осени. В настоящий момент (4 мая 2013 года), в настройках поиска осталась настройка для включения «Википоиска», но в выдаче соответствующие элементы управления отсутствуют. Другие поисковые системы подобной функциональности пока не предоставляли.

22 сентября 2010 года компания запустила голосовой поиск в России. Чтобы осуществить поиск, необходимо нажать в телефоне кнопку рядом со строкой поиска и произнести свой запрос, телефон отправит ваш голос на сервер и браузер, будет выдавать строку с распознанным вашим запросом и результатами поиска по нему.

По случаю праздника или круглой даты какой-нибудь широко известной личности, стандартный логотип Google у региональных доменов может меняться на праздничный, имеющий определённую тематику, смысл. Например, по случаю дня рождения Наполеона Орды 11 февраля 2010 года на логотипе белорусского домена Google появились акварели этого известного художника, 6 июля поздравляли со 121 - летием Марка Шагала (логотип был в виде коллажа из фрагментов его работ). После десятилетнего ожидания 22 марта 2011 года Google выиграл патент на "GoogleDoodle".

Поисковая система Yahoo —одна из самых первых (создана Дэвидом Фило и Джерри Янгом в апреле 1994года) по сей день остается и самой популярной из них, традиционно сочетая поиск, как по ключевым словам, так и с помощью иерархического дерева разделов [6].

Нынешнее развитие Yahoo можно определить как движение в онлайн, интерактивность. Yahoo быстро осваивает эту область Интернет-услуг, но возникает одна проблема: ядро Yahoo! не было на это рассчитано. Не была в 1994 году заложено в него "онлайновая" составляющая, ее "приклеил" Тим Кугл несколькими годами позже. Естественно возникает угроза хакерских атак через эту незащищенную область.

Одно из новшеств поисковой системы Yahoo - панель задач для браузера Firefox,. Этот инструмент помогает пользоваться поиском Yahoo, не заходя на официальный сайт, а лишь используя функциональные кнопки панели.

1 сентября 2005 года поисковик Yahoo, которому принадлежит более 200 миллионов адресов электронной почты по всему миру, анонсировал запуск новой системы поиска текстов, фотографий и других документов, содержащихся в письмах.

Необходимость такого нововведения возникла вслед за увеличением объёма хранимых данных, ведь некоторые пользователи создают целые почтовые архивы. Подгоняемый конкурентом Google и его почтовым сервисом Gmail, Yahoo для хранения почты предлагает отныне 1 гигабайт бесплатного места, или 2 гигабайта по годовому абонементу. «Как только вы получаете возможность хранить больше информации, вам необходимы и расширенные поисковые возможности», – объясняет Эрик Петерсон, аналитик компании JupiterResearch.

Пользователи поисковой системы Yahoo, в свою очередь, смогут теперь использовать возможности детализированного поиска слов в названии или непосредственно в тексте письма, а также в присоединенных документах, не открывая их. Результат поиска отражается в трёх строках с указанием всех атрибутов. На панели справа отображаются все похожие документы. Найденные фотографии выводятся на экран в уменьшенном виде, что значительно облегчает поиск. Система также учитывает орфографические ошибки, позволяя искать слова лишь по первым буквам.

Yahoo планирует предложить новую систему небольшому числу американских пользователей, а затем распространить её по всему миру. Со стороны клиентов это не потребует никаких дополнительных усилий. «Когда услуга станет, доступна, в левом верхнем углу страницы вашего почтового ящика появится соответствующий баннер», – обещает компания Yahoo.

По данным comScoreMediaMetrix на июль этого года, домену Yahoo принадлежит 219 миллионов адресов электронной почты, что составляет 31,5% мирового рынка, уступая лишь Microsoft с 221 миллионом пользователей сервиса Hotmail (35,5% рынка) [6].

Baidu – лидер среди китайских поисковых систем. По количеству обрабатываемых запросов поисковый сайт Байду стоит на 3 месте в мире (3 миллиарда 428 миллионов; с долей в глобальном поиске 5,2 %). Уже в конце года в Китае свыше 170 млн. пользователей займутся поиском информации в Интернете. Аналитик J.P. Морган Дик Вей исходит в своем актуальном анализе из того, что это число вырастет в течение следующих трех, четырех лет до 100 млн. пользователей.

Гигантский рынок с высокими доходами для Baidu, сравнивают только прибыль, которую Google достигает в США с очень похожей бизнес-моделью [18].

Теперь опишем наиболее популярные поисковые системы российского рынка информационных ресурсов.

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. На сегодняшний день самой популярной русскоязычной поисковой системой является Яндекс – 54% всех поисковых запросов.

Основное отличие русскоязычных поисковых систем от иностранных одно – то, что глобальные поисковые системы, поддерживающие поиск на русском языке, не поддерживают русскую морфологию. В русскоязычной части сети Интернет работают около двух десятков поисковых систем, но подавляющие большинство пользователей работает лишь с несколькими, подробно остановимся на самых крупных.

Яндекс – на сегодня наиболее популярная русскоязычная поисковая система, ежемесячно к ней обращаются более 35 миллионов пользователей сети Internet. Начала свою работу во второй половине 1997 года учитывая морфологию русского языка. История компании «Яндекс» началась в 1990 году с разработки поискового программного обеспечения в компании «Аркадия». За два года работ были созданы две информационно-поисковые системы – Международная Классификация Изобретений, 4 и 5 редакция, а также Классификатор Товаров и Услуг. Обе системы работали локально под DOS и позволяли проводить поиск, выбирая слова из заданного словаря, с использованием стандартных логических операторов. В1993 году «Аркадия» стала подразделением компании CompTek. В 1993-1994 годы программные технологии были существенно усовершенствованы благодаря сотрудничеству с лабораторией Ю. Д. Апресяна (Институт Проблем Передачи Информации РАН). В частности, словарь, обеспечивающий поиск с учетом морфологии русского языка, занимал всего 300Кб, то есть целиком грузился в оперативную память и работал очень быстро. С этого момента пользователь мог задавать в запросе любые формы слов [18].

Слово Яндекс придумал за несколько лет до этого один из основных и старейших разработчиков поискового механизма. «Яndex» означает «Языковой index», или, если по-английски, «Yandex» - «YetAnotherindexer». За 4 года публичного существования Яndex возникли и другие толкования. Например, если в слове «Index» перевести с английского первую букву ("I" – «Я»), получится «Яndex».

В начале 1996 года был разработан алгоритм построения гипотез. Отныне морфологический разбор перестал быть привязан к словарю - если какого-либо слова в словаре нет, то находятся наиболее похожие на него словарные слова и по ним строится модель словоизменения. В это время Интернет в России только начинался. Еще через полгода стало очевидно, что ничто не отделяет CompTek от создания собственной глобальной поисковой машины. Объем Рунета составлял тогда всего несколько гигабайт. Осенью 1997 года был открыт Yandex.Ru.

Помимо поисковой системы, сегодня Яндекс – огромный портал с целым набором широко используемых сервисов, такими как каталог, Яндекс. деньги, и другие. Официально поисковая машина Yandex.Ru была анонсирована 23 сентября 1997 года на выставке Softool. Основными отличительными чертами Yandex.Ru на тот момент были проверка уникальности документов (исключение копий в разных кодировках), а также ключевые свойства поискового ядра Яндекс, а именно: учет морфологии русского языка (в том числе и поиск по точной словоформе), поиск с учетом расстояния (в том числе в пределах абзаца, точное словосочетание), и тщательно разработанный алгоритм оценки релевантности (соответствия ответа запросу), учитывающий не только количество слов запроса, найденных в тексте, но и «контрастность» слова (его относительную частоту для данного документа), расстояние между словами, и положение слова в документе [19].

Гибкий язык запросов, позволяет производить поиск по самым различным критериям. Так, например, для операции исключения можно указать область действия: запрос А ~~ В найдёт документы (страницы), в которых присутствует А, но не присутствует В, а запрос А ~ Б - документы, где слово Б не присутствует со словом А в одном предложении. Аналогично, оператор & ищет сочетания ключевых слов в предложении, а && - во всём документе.

По умолчанию Яндекс выводит по 10 ссылок на каждой странице выдачи результатов, в настройках результатов поиска можно увеличить размер страницы до 20, 30 или 50 найденных документов. Иногда порядок сайтов на этих страницах может отличаться, так как обновление баз для этих результатов происходит не одновременно.

Время от времени алгоритмы Яндекса, отвечающие за релевантность выдачи, меняются, что приводит к изменениям в результатах поисковых запросов. Такие изменения, официально объявленные, происходили, например, в марте 2004 года, августе 2005 года и январе 2007 года; по неофициальным сведениям, их значительно больше (например, в августе-сентябре 2007 года). Последнее такое изменение произошло в ноябре 2009 года, когда была выложена обновленная версия поисковой программы «Снежинск».

В частности, эти изменения направлены против поискового спама, приводящего к нерелевантным результатам по некоторым запросам (реже - по целым семействам запросов).

Rambler– старейшая поисковая система российского Интернет, запущена в 1996 году, на сегодня вторая по популярности с обращением более 25 миллионов посетителей в месяц. Помимо поисковой системы, сегодня Рамблер один из крупнейших порталов Русскоязычной части Интернета с большим набором широко известных сервисов, таких как каталог Рамблер, Рамблер-почта, Рамблер-ICQ или Рамблер-ТВ. По сути сегодня Рамблер - больше, чем просто поисковая система и набор сервисов, это крупная медиагруппа. Поисковая машина «Рамблер» начала работу в октябре 1996 года, на стартовом этапе содержала всего 100 тысяч документов. «Рамблер» не был первой отечественной поисковой системой, однако в первый год своего существования (когда весь русский веб с приемлемой степенью правдоподобия индексировался «Рамблером», «Апортом», «Русской поисковой машиной», а также шведской и калифорнийской AltaVista) вынес основной груз поисковых запросов [18].

Вторая версия «Рамблера» начала разрабатываться летом 2000 года, в марте нынешнего года приняла достаточно законченные очертания. В нее были введены функции, давно уже имевшиеся в конкурирующих системах. Она учитывает координаты слов, обучена строгой и нечеткой морфологии, связывает поиск с каталогом, в качестве которого используется Top100 (http://top100.rambler.ru/), группирует результаты поиска по сайтам, ищет по числам. Достаточно удачная архитектура продукта позволяет «Рамблер» иметь для поисковика количество серверов в 2 раза меньшее, чем у «Яндекса», и в 3 раза меньшее, чем у «Апорта» [27].

Апорт– третья попопулярности на сегодня поисковая система с обращением более 16 миллионов посетителей в месяц. Апорт позволяет пользователям осуществлять полнотекстовый поиск документов с учетом морфологии русского языка в

запросах. Поисковая система построена на основании новейших достижений в области информационного поиска и использует уникальные алгоритмы сортировки найденных результатов. Разнообразные специализированные поиски (Знакомства, Товары, Новости, Рефераты, MP3 и др.) дают пользователям дополнительные возможности находить различную информацию в Сети. В поисковую машину интегрирован один из крупнейших в Русскоязычной части Интернет каталогов Интернет-ресурсов «Апорт-каталог».

Поисковая машина «Апорт» была впервые продемонстрирована в феврале 1996 года на пресс-конференции «Агамы» по поводу открытия «Русского клуба». Тогда она искала только по сайту russia.agama.com. Потом она начала искать по четырем, потом по шести серверам. В итоге, день рождения и фактический старт системы сильно «размазались» по времени, а официальная презентация «Апорта» состоялась только 11 ноября 1997 года. К тому времени в его базе был проиндексирован первый миллион документов, расположенных на 10 тысячах серверов. Создателем системы выступила компания «Агама» - разработчик программного обеспечения для платформы Windows, главным из которых являлся корректор орфографии «Пропись». Лингвистические разработки «Агамы» использовались при создании поисковой машины, в которой, скажем, в отличие от «Рамблер», изначально учитывалась морфология слов и осуществлялась по желанию клиента проверка орфографии запроса.

Важнейшими свойствами первой версии «Апорта» являлся перевод запроса и результатов поиска на английский язык и обратно, а также реконструкция всех проиндексированных страниц из собственной базы (что означает возможность просмотра страниц, уже несуществующих в оригинале).

«Апорт 2000» стал первой российской поисковой машиной, практически реализовавший две базовых технологии американской поисковой машины Google. Первая – учет «ранга страницы» (PageRank), который характеризует ее популярность (вычисляется по количеству ссылок на ресурс из внешнего Интернета: вес ссылки с популярного сайта выше, чем вес ссылки с менее популярного; ссылки, включающие слова запроса, имеют больший вес, чем, скажем, слово «здесь»). Вторая – обработка запроса, ориентируясь на HTML-код страницы. В «Апорт 2000» учитывается также вхождение слов запроса в URL. Среди недокументированных особенностей - больший приоритет сайтам, получившим высшую и элитную лигу в каталоге AtRus [18].

Можно отметить и то, что «Апорт» первым устроил поиск по новостным лентам (какие бы ложные сведения о приоритете «Яндекса» в этом сервисе не распускал в свое время Internet.ru). И, наконец, еще одно первенство «Апорта»– использование платной нулевой строки в выдаче. Однако в «Апорте» нельзя купить не нулевое, а просто более высокое место для своего сайта в результатах поиска.

Организация масштабируемости в архитектуре «Апорт 2000» такова, что можно дробить поисковую базу «Апорта» на несколько отдельных баз, каждый маленький «Апорт» работает на своем компьютере. «Апорт 2000» считает, что весь Интернет поделен на фрагменты. После проведения поиска по этим фрагментам, пользователю интегрируется и выдается общий ответ. Добавлять новые маленькие "апортики" можно путем не очень сложной процедуры. В случаях аварий отдельных машин выдаются несколько отличные от штатных интегральные результаты, что мы можем время от времени наблюдать.

В данном параграфе были рассмотрены мировые и русскоязычные поисковые системы. По результатам рейтинговых данных были выявлены наиболее популярные системы поиска. Таковыми являются Google, среди мировых ИПС, и Яндекс, среди русскоязычных систем. Критериями выбора именно этих систем являются удобство поиска информации, а именно: высокое качество алгоритма сортировки результатов, гибкий язык запросов, релевантность. Кроме этого были рассмотрены свойства большинства систем, и были определены некоторые особенности каждой из них. Таким образом, удалось выявить, что каждая система по-своему удовлетворяет критериям поиска и вполне может конкурировать с другими поисковыми системами.

3.2 Сравнительный анализ современных информационно-поисковых систем

Теперь обратимся к положительным и отрицательным сторонам ранее рассмотренных наиболее популярных поисковых систем, тем самым продемонстрировав особенности, которыми должна обладать наиболее удобная система поиска

Таблица 3.1 – Преимущества и недостатки поисковых систем

Поисковая система	Преимущества	Недостатки
Яндекс	1)Непрерывное развитие системы. 2)Качество выдачи растет, все больше удобных сервисов предлагает компания: каталог, карты, новости, прогноз погоды, почта. 3) глубокий морфологический анализ обрабатываемых терминов. 4) обладает хорошим механизмом распознавания одного документа в нескольких кодировках или на зеркальных серверах. 5) оригинально сконструированный механизм выдачи результатов. 6) огромная индексная база.	1) Разница в выдаче при наборе слова с большой (маленькой) буквы (иногда выдача меняется, иногда нет). 2) Частое выпадение секторов поисковой базы - когда исчезают части сайтов из выдачи и восстанавливаются через 2-5 дней. 3) Обновление индексов поисковой базы происходит недостаточно часто и регулярно.

Rambler	1) Система работает с большой скоростью поиска. 2) Обновление поискового индекса происходит несколько раз в день. 3) Поисковик всегда находит самые свежие документы и последние новости. 4) Обладает близким к оптимальному выводом результатов поиска. 5) производит ранжирование результатов в зависимости от частоты употребления и местоположения искомых терминов. 6) Один и тот же документ в различных кодировках показывается только один раз, а его конкретные адреса. суммируются в списке, идущим за резюме.	1) На величину индекса релевантности влияет время существования сайта в сети. Эта особенность позволяет пользователям находить ресурсы, которые давно существуют, успешно развиваются, а не сайты-однодневки. Но такой подход значительно затрудняет попадание в выдачу новых сайтов, информация на которых подчас оказывается актуальной и, возможно, более важной для пользователя. 2) невозможность осуществления поиска по целой фразе указывая в запросах предельное расстояние искомых терминов друг от друга.

Aport	1)содержит довольно удобный в пользовании каталог. 2)широкие возможности составления запроса. 3) автоматический перевод запроса с русского на английский язык и наоборот. 4) Реконструкция проиндексированных страниц происходит из собственной базы. Это дает возможность просмотра уже несуществующих страниц.	1)не всегда быстро находит то, что от него просишь. 2) каталог не обновлялся уже очень давно. 3) способен выделять один и тот же документ в различных кодировках и выдавать ссылку на него лишь один раз, перечисляя конкретные адреса в списке URL. 4) не всегда корректная обработка названий страниц, из-за чего в результатах поиска часто указывается «документ без названия», в то время как метки title на большинстве таких страниц содержат важные данные.

Google	1) Очень мощная поисковая система, которая находится в постоянном развитии. 2) База индексов этой системы обновляется раз в два дня, качество выдачи очень высокое, найти необходимый документ или информацию довольно легко. 3) Система ориентирована в основном на ссылки, причем учитываются как входящие, так и исходящие ссылки с ресурса. 4) Способна выдавать результаты на запросы по семантике языка программирования (исходный код поиска).	1) Нередко встречаются ссылки на сайты с уже устаревшей информацией. 2) Случается, что ссылки, которые находятся в результатах поиска, ведут на сайт, находящийся в стадии разработки. 3) На запрос «фильм» и «фильмы» результаты поиска будут отличаться. 4) отсутствие возможности указать конкретную грамматическую форму слова, либо ударение также значительно усложняет процесс поиска информации.

Yahoo!	1) Содержит ссылки, которые наиболее полно отвечают указанной в запросе тематике. 2) Имеются интеллектуальные средства «отсечения» пустых, находящихся в разработке или чисто рекламных сайтов, далеких от искомой тематики. 3) всегда легко определить, в каком разделе находится нужная информация. 4) В случае если на Yahoo нет результатов, сразу выводятся результаты с AltaVista.	1) Возможна проблема с отсутствующими страницами, поскольку веб-мастера обычно забывают удалить свои сайты с поисковых систем, а на Yahoo нет механизма автоматического обновления. 2) Чисто русские ресурсы не добавляются, потому что их просто некому смотреть и оценивать содержимое. 2)Нет собственной поисковой машины. 3) Ищет слова, заданные в критерии поиска только в названии и описании страницы
Baidu	К концу 2002 года количество китайских сайтов, индексируемых Baidu, было на 50% больше, чем у любого конкурента.	Число заблокированных результатов поиска у Baidu на 30% больше, чем у Google Google оставила Baidu далеко позади, поскольку предлагает рекламодателям выход на международные рынки.

MSN(Bing)	1)Предоставляет пользователям возможность сортировать результаты поиска: по дате, по алфавиту, по релевантности. 2) При осуществлении поиск по ключевому слову, команда специалистов компании отслеживает наиболее релевантные на их взгляд сайты, вручную отбирают и классифицируют их, и вносят в определенные рубрики директории. 3) ранжирования узлов по популярности и сезонным изменениям. 4) Помощь со стороны человека-редактора.	1) Поисковая система полна спамом. 2) Использует внешние данные для обработки поисковых запросов, поэтому на релевантность влияют: расположение ключевых слов, популярность ресурса и текст ведущих на сайт, и ведущих с сайта ссылок.

Главный недостаток современных поисковых систем – это их централизация. А централизация означает, что вся информация хранится в одном месте, все работы и расчёты производятся в одном месте, все решения (результаты выдачи) принимаются в одном месте.

Итак, почему это недостаток, здесь несколько причин:

1) Полная централизация требует колоссальных ресурсов – это огромные базы данных, множество компьютеров и т.д. Учитывая темпы роста Интернета в ближайшем будущем придется применять просто невероятные мощности.

2) Только при управлении в одном центре можно достичь полной конфиденциальности. А так как по нашей концепции поисковая система должна быть открытой, то и необходимость в централизации отпадает полностью.

3) Поисковая система не всегда может правильно оценить конкретный ресурс. Правильнее самому обладателю сайта поручить выполнение ранжирования документов внутри сайта. И теперь, самое главное как уйти от централизации и устранить все эти минусы - это внедрение в каждый сайт своей мини-поисковой системы. Эта мини-поисковая система будет индексировать содержимое сайта по правилам самого обладателя сайта. Только вебмастер будет решать, какие страницы его сайта по каким запросам более релевантные. А потом свои индексы уже будет отправлять на сервер поисковой системы.

Ещё одной из основных проблем при создании новой поисковой системы является учет мнения пользователей.

Попытка непосредственного выявления представлений пользователей об идеальной поисковой системе обычно не приводит к нужному результату: пользователи перечисляют все, что когда-либо видели или использовали в существующих системах. Не стоит ждать от пользователей навыков проектирования – они вряд ли смогут быстро описать, как должна выглядеть идеальная поисковая система.

Более продуктивным подходом к решению этой проблемы является анализ идеальной модели поисковой системы, которой оперируют пользователи. Идеальная модель – это совокупность представлений пользователя о целях, функциях, структуре, способах контроля и управления, возможных действиях с системой, которые определяют его деятельность. Такой подход – от анализа представлений пользователей и построения идеальной модели к проектированию интерфейсов продукта - снижает риск того, что продукт не понравится пользователям, не будет принят и востребован ими.

В идеальной модели должны присутствовать следующие компоненты:

Primarynouns (электронное письмо, товар в Интернет-магазине, картинка, доступная для просмотра в Интернете) – это основные элементы, с которыми пользователь производит действия или манипуляции при работе с системой.

Сценарий использования - это описание представлений пользователей о взаимодействии с системой, разбитое на элементарные шаги. Сценарий использования иллюстрирует поведение пользователя при решении определенной задачи с помощью поисковой системы.

Диаграмма задач является графическим отображением представлений пользователей о перечне решаемых в системе задач.

Диаграмма навигации демонстрирует представления пользователей о порядке смены экранов, с которыми они сталкиваются при работе с системой, и содержании этих экранов. Диаграмма построена на основе сценариев использования системы и используется в процессе проектирования интерфейсов.

Проблема 1: Оптимизаторы не могут ясно понять, каким должен быть, «хороший» сайт в понимании поисковика и как сделать его таким, чтобы поисковик считал его наиболее релевантным по запросам.

Решение этой проблемы хорошо реализовано в поисковой системе MSNSearch. В системе ранжированием занимается не только поисковик, но ему также помогает человек-редактор. Благодаря этому, при осуществлении поиск по ключевому слову, команда специалистов компании отслеживает наиболее частые запросы, вводимые в поисковую форму, и подбирает сайты, наиболее релевантные тематике запроса, а так же вручную отбирают и классифицируют их, и вносят в определенные рубрики директории. Что, например, в сравнении с самой популярной поисковой системой мира – Google, которая сама определяет релевантность Интернет-страниц (страница, на которую ссылаются чаще, более релевантна и значит более популярна) помогает избежать этой проблемы.

Проблема 2: Наличие доступных и понятно изложенных правил по специальному синтаксису каждой отдельной поисковой системы.

Изложение доступных и понятно изложенных правил по специальному синтаксису присутствует в следующих поисковых системах:

Яndex;

Google;

Апорт;

Проблема 3: Высокий уровень релевантности выдаваемой информации.

Используя опыт, полученный в ходе выполнения курсовой работы, и опыт использования поисковых систем в жизни в целом, представляю список поисковых систем (начиная с той, у которой более релеванты результаты поставленным запросам), поисковые системы, не соответствующие, по моему мнению, критерию

«релевантность выдаваемой информации» не войдут в представленный ниже список:

Яndex;

Google;

Апорт;

Проблема 4: Спрос на поисковые системы, которые больше напоминают Интернет-портал, где можно завести почтовый ящик, узнавать курс валют и прогноз погоды, читать блоги и форумы.

Этому критерию пользователей отвечают:

Таблица 3.2 – Критерии пользователей

Поисковые системы	Почтовый ящик	Курс валют	Прогноз погоды	Блоги	Форумы
Яndex					
Google					
Rambler					
Апорт					
MSN					
Yahoo					

Получили, что всем необходимым критериям не соответствует ни одна и рассмотренных нами поисковых систем. Ближе всего к идеалу находятся поисковые системы Яndex, Rambler, Апорт. За ними следуют Google и MSN, и заключает шестерку ведущих поисковых систем – Yahoo.

Заключение

Пользователи сети Internet имеют широкие возможности для получения экономической, социальной, научной, технологической и разнообразной текущей информации.

Для исследовательской работы была сформулирована главная цель –определить сущность и значимость информационно-поисковых систем в современном обществе и выявить наиболее совершенную с точки зрения интерфейса и алгоритма поиска систему для пользователя.

В соответствии с поставленной целью втеоретической части курсовой роботы были рассмотрены основные элементы и понятия информационного поиска, показанаструктура, работа и компоненты информационно-поисковых систем.Также были определены основные показатели оценки работы поисковых систем.

Очень часто приходится искать информацию в сети, незная даже приблизительно адрес страницы, на которой она может располагаться. В таких случаях на помощь приходит поисковая машина.

Поисковые машины – это роботизированные системы. Специальная программа-робот, которую называют паук или ползун, постоянно обходит Сеть в поисках новой информации, которую она вносит в базу данных.

При поиске в Интернете важны две составляющие – полнота (ничего не потеряно) и точность (не найдено ничего лишнего). Обычно это все называют одним словом – релевантность, то есть соответствие ответа вопросу. Важными показателями являются охват и глубина поисковой машины, скоростью обхода и актуальностью ссылок (скорость обновления информации в этой базе данных), качеством поиска (чем ближе к началу списка оказывается нужный вам документ, тем лучше работает релевантность).

При решении практической задачи части исследовательской курсовой был проведен сравнительный анализ самых популярных поисковых системна мировом и российском рынке информационных ресурсов. Были выявлены их преимущества и недостатки.

С помощью анализа выяснилась еще одна из проблем: при создании новой поисковой системы учитывается мнение пользователей.

Попытка непосредственного выявления представлений пользователей об идеальной поисковой системе обычно не приводит к нужному результату: пользователи перечисляют все, что когда-либо видели или использовали в существующих системах.

В ходе работы выяснилось, что на настоящий момент времени не существует «идеальных» поисковых систем, однако, по данным произведенного анализа мы выяснили, что поисковая система Яndex больше всех приближена к модели «идеальной» поисковой системы. А такие поисковики как Google и Апорт поочерёдно делят то 2, то 3 места.

Стоит также обратить внимание на то, что каждая поисковая система будь то российская или зарубежная предоставляет различные возможности для поиска информации, поэтому нельзя однозначно определить какая из систем является наилучшей. Исходя из этого, для удобства поиска и полноты информации мы рекомендуем использовать несколько поисковых систем.

Также в рамках данной работы были рассмотрены приемы расширенного поиска, позволяющие в разы увеличить эффективность поиска и быстро найти необходимую информацию (см.Приложение А).

Выполненное исследование открывает новые возможности для дальнейшей разработки вопросов методики применения ИПС как в самостоятельном, так и в дистанционном обучении. Более того проделанная работа поможет пользователям грамотно использовать поисковые системы, и быстро находить нужную и актуальную информацию для дальнейшего её использования. Данное исследование может служить теоретической основой применения информационных средств в обучении.

Список литературы

1. Ашманов, И. С. Продвижение сайта в поисковых системах / И. С. Ашманов. - М.: «Вильямс», 2010. - 304 с.

2. Байков, В. Д. Интернет. Поиск информации. Продвижение сайтов / В. Д. Байков. - СПб: БХВ - Петербург, 2010. — 288 с.

3. Блог WebMilk.ru. - [Электронный ресурс]. URL:http://webmilk.ru/2008/01/24/yandeks-voshel-v-top-10-poiskovyih-mashin - Режим доступа: (дата обращения: 8.04.2014);

4.Гаврилов, А. В. Локальные сети ЭВМ / А. В. Гаврилов.- М. : «Мир», 1990.- 154 с.

5. Гайдамакин, Н. А. Автоматизированные информационные системы, базы и банки данных / Н. А. Гайдамакин.- М. : «Гелиос», 2012.- 280 с.

6. ГОСТ 7.74-96 «СИБИД. Информационно-поисковые языки. Термины и определения - [Электронный ресурс]. URL: http://www.standartov.ru/norma_doc/33/33984/index.htm- Режим доступа: (дата обращения: 8. 04. 2014);

7. Информатика. Базовый курс: учебник / под ред. С. В. Симоновича. - СПб: «Питер», 2007.- 110 с.

8. Информационные поисковые системы - [Электронный ресурс]. URL: http://oka2o1o.narod.ru/ips.htm - Режим доступа: (дата обращения: 1.06.2014).

9. Итоги года - Sostav.ru. - [Электронный ресурс]. URL: http://www.sostav.ru/itogi/s/2009/6 - Режим доступа: (дата обращения: 8.04.2014);

10. Кадеев, Д. Н. Информационные технологии и электронные коммуникации / Д. Н. Кадеев.- М.: «Электро», 2011.- 250 с.

27. Как все начиналось - Google, Yahoo, Яндекс, Mail.ru, Rambler. TvoiExpert.

11. Колисниченко, Д. Н. Поисковые системы и продвижение сайтов в Интернете / Д. Н. Колисниченко. - М.: «Диалектика», 2007. – 272 с.

12. Ландэ, Д. В. Поиск знаний в Internet / Д. В. Ландэ. - М. : «Диалектика», 2005. — 272 с.

13. Маннинг, К. Введение в информационный поиск / К. Маннинг. – М.: «Вильямс», 2011.- 200 с.

14.Описание поисковой системы Bing. - [Электронный ресурс]. URL: http://anokalintik.ru/opisanie-poiskovoj-sistemy-bing.html - Режим доступа: (дата обращения: 10.03.2014)

15.Поисковаясистема Google- история компании Bbcont.ru. - [Электронный ресурс].URL:http://bbcont.ru/business/poiskovaya_sistema_google_istoriya_kompanii.html- Режим доступа: (дата обращения: 12.04.2014);

16. Путеводители в лабиринте Интернета. - [Электронный ресурс]. URL:http://rutracker.org/forum/viewtopic.php?t=1117865 - Режим доступа: (дата

обращения: 8.05.2014);

17. Поисковая система Yahoo! - [Электронный ресурс]. URL: http://www.egonika.ru/forum/poiskovye_sistemy/poiskovaya_sistema_yahoo - Режим доступа: (дата обращения: 8.04.2014);

18. Поисковая система Байду. ЦИТ-Форум - журнал о поисковых системах. - [Электронный ресурс]. URL:http://www.cit-forum.com/baidu/poiskovaja-sistema-bajdu.html - Режим доступа: (дата обращения: 14.05.2014);

19. Поисковая машина Yandex.Ru. - [Электронный ресурс]. URL:http://spravki.se-ua.net/yandex - Режим доступа: (дата обращения: 8.04.2014);

20. Поисковая оптимизация веб страниц SEO. - [Электронный ресурс]. URL: http://creng.ru/seo/seo-poiskovaya-optimizaciya-veb-stranic - Режим доступа: (дата обращения: 8.04.2014);

21. Просвещение W3. Google. - [Электронный ресурс]. URL: http://w3pro.ru/tematika/google - Режим доступа: (дата обращения: 8.04.2014);

23. Сахарова, Е. В. Информатика. Методические указания / Е. В. Сахарова.- Ставрополь: СТИС, 2011.- 200 с.

24. Схемы и рисунки ИПС - [Электронный ресурс]. URL: http://ssofta.narod.ru/bd/ets2.htm - Режим доступа: (дата обращения: 10.05.2014).

25. Структура и классификация автоматизированных информационных систем - [Электронный ресурс]. URL: http://do.rksi.ru/library/courses/opais/tema1_3.dbk - Режим доступа: (дата обращения: 8.12.2011).

26. Терехов, И. В. Автоматизированные информационные системы в образовании и науке [Электронный ресурс]: семинар / И. В. Терехов: М.-2009. http://ou.tsu.ru/seminars/sem13/tezis/section6.htm - Режим доступа: (дата обращения: 8.12.2011).

27. Чурсин, Н. А. Популярная информатика / Н. А. Чурсин.- М.: «Вильямс», 2011.- 300 с.

28. Якубайтис, Э. А. Информатика – электроника- сети / Э. А. Якубайтис.- М.: «Финансы и статистика», 2010.- 300 с