МИНОБРНАУКИ РОССИИ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет теории и методики преподавания

Кафедра романской филологии

Курсовая работа

Информационно-поисковые системы

Студент \_\_\_\_\_\_\_\_\_\_\_ Быстрова З.М 25.09.18

Руководитель \_\_\_\_\_\_\_\_\_ Донина. О.В к.ф.н. 22.10.18

Воронеж 2018

Содержание

**Введение**…………………………………………………………………

……3 - 4

**Глава 1. Информационно-поисковые системы4-14**

1. Понятие информационно-поисковых систем………………………………………………………………………....4-6
2. История развития ИПС………………………………………………………………………...6-7
3. Структура ИПС…………………………………………………………………….....8-11
4. Виды ИПС4-11

**Глава 2. Современные информационно-поисковые системы………...14-23**

1. Сферы использования современных ИПС……………………………………………………………………...14-16
2. Архитектура современных ИПС……………………………………………………………………...16-18
3. Популярные ИПС18-23

**Заключение**………..…………………………………………………………23-25

**Список использованной литературы**25-26

Введение

Актуальность. Современный этап развития цивилизации характеризуется переходом наиболее развитой части человечества от индустриального общества к информационному. Одним из наиболее ярких явлений этого процесса является возникновение и развития глобальной информационной компьютерной сети.

Проблема поиска и сбора информации - одна из важнейших проблем информационно поисковых систем. Конечно, нельзя сравнивать в этом отношении, скажем, средние века, когда поиск информации был проблемой потому, что этой информации было мало, и требовались усилия только для того, чтобы найти хоть что-то по более или менее значительному интересующему вопросу. Так, сначала появилась возможность пойти в библиотеку и, потратив там время на выбор нужной книги по каталогу, найти необходимую информацию. Но каталоги не решают полностью проблем поиска информации даже в рамках одной библиотеки, так как в каталожную запись входит относительно мало информации: заголовок, автор, место издания. Проблема поиска информации приобрела новый характер в 20-м столетии, с началом развития века информационных технологий. Теперь она заключается не в том, что информации мало и поэтому ее трудно найти, а в том, что ее теперь наоборот становится все больше и больше, и от этого найти ответ на интересующий вопрос может оказаться тоже довольно сложной задачей. **Проблема поиска информации** значительно усложняется при использовании виртуальных источников. Здесь используется технология онлайновых каталогов, в результате применения которой пользователь имеет возможность выполнять поиск в каталогах сразу нескольких библиотек, чем, на самом деле, еще больше усложняет себе задачу, но, с другой стороны, увеличивает шансы решить ее.

На современном этапе все информационное пространство, в котором мы живем, все больше погружается в Internet. Internet становится основной формой существования информации, не отменив традиционных, такие как журналы, радио, телевидение, телефон, всевозможные справочные службы.

**Целью исследования** является изучение автоматизированных информационно - поисковых систем.

Задачей в данной курсовой работе рассматриваются теоретические основы автоматизированного информационного поиска, классификация и разновидности информационно поисковых систем. Также анализируется материал по применяемым в настоящее время информационно - поисковым каталогам полнотекстовых и гипертекстовых поисковых систем.

При появлении сети Internet проблема поиска становилась более актуальной. Internet - всемирная компьютерная сеть, представляющая собой единую информационную среду и позволяющая получить информацию в любое время. Но с другой стороны в Интернете хранится очень много полезной информации, но для поиска её требуется затрачивать много времени. Эта проблема послужила поводом к появлению поисковых систем. В данной курсовой работе будут рассмотрены поисковые системы в сети Internet.

**Глава 1. Информационно-поисковые системы**

1.Понятие информационных поисковых систем   
  
Автоматизированная поисковая система – система, состоящая из персонала и комплекса средств автоматизации его деятельности, реализующая информационную технологию выполнения установленных функций. Опыт и практика создания систем в различных сферах деятельности позволяет дать более широкое и универсальное определение, которое полнее отражает все аспекты их сущности. Под информационной системой в дальнейшем понимается – организованная совокупность программно–технических и других вспомогательных средств, технологических процессов и функционально–определённых групп работников, обеспечивающих сбор, представление и накопление информационных ресурсов в определённой предметной области, поиск и выдачу сведений, необходимых для удовлетворения информационных потребностей установленного контингента пользователей – абонентов системы.  
  
Следовательно, информационно-поисковая система выполняет следующие функции:   
  
1. хранения больших объемов информации;   
  
2. быстрого поиска требуемой информации;   
  
3. добавления, удаления и изменения хранимой информации;   
  
4. вывода информации в удобном для человека виде.  
  
Создание и использование информационно-поисковой системы началось в конце 1950-х - начале 1960-х годов. Информационно-поисковая система опирается на базу данных, в которой осуществляет поиск нужных документов по заявкам пользователей. Частным случаем информационно-поисковой системы является информационно-справочная система, которая в ответ на запросы выдает сведения, в явной форме отсутствующие в базе данных. Широко информационно-поисковые системы используются и в базах знаний.

Далее мы будем, в основном, рассматривать ИПС для всемирной паутины (WorldWideWeb). Основными показателями ИПС для WWW являются *пространственный масштаб* и *специализация*.

По пространственному масштабу ИПС можно разделить на локальные, глобальные, региональные и специализированные. Локальные поисковые системы могут быть разработаны для быстрого поиска страниц в масштабе отдельного сервера. Региональные ИПС описывают информационные ресурсы определенного региона, например, русскоязычные страницы в Интернете. Глобальные поисковые системы в отличие от локальных стремятся объять необъятное - по возможности наиболее полно описать ресурсы всего информационного пространства сети Интернет.

2. История развития ИПС

Обратимся к истории возникновения сети Internet, которая была создана в связи с возникшей необходимостью совместного использования информационных ресурсов, распределенных между различными компьютерными системами. Большинство первых приложений, включая FTP и электронную почту, были разработаны исключительно для обмена данными между хост- компьютерами Internet. Другие приложения, такие как Telnet, создавались для того, чтобы пользователь получил возможность доступа не только к информации, но и к рабочим ресурсам удаленной системы. По мере развития Internet (увеличения пользователей и хост- компьютеров) прежние методы обмена данными перестали отвечать возросшим потребностям пользователей. Возникла необходимость разработки новых способов поиска сетевых ресурсов и доступа к ним, которые позволяли бы использовать информацию независимо от ее формата и расположения . Для удовлетворения таких потребностей сначала были созданы поисковая система Archie, решающая задачу локализации ресурсов на FTP-сервере, и система Gopher, упрощающая доступ к различным сетевым ресурсам. Затем были разработаны сетевые информационные системы World Wide Web и WAIS, предлагающие абсолютно новые методы получения информации. Принципы работы этих систем позволяют легко ориентироваться в огромном количестве информационных ресурсов без необходимости предоставления механизмов работы самой сети Internet. Такой подход позволяет говорить уже не просто о ресурсах взаимосвязанных компьютерных систем, а об особых информационных пространствах сети .

Система Archie представляет собой комплекс программных средств, работающих со специальными базами данных. В этих базах данных содержится постоянно пополняющаяся информация о файлах, к которым можно получить доступ через сервис FTP. Пользуясь услугами системы Archie, можно осуществить поиск файла по шаблону его имени. При этом пользователь получит список файлов с точным указанием места их хранения в сети, а также с информацией о типе, времени создания и размере файлов. Доступ к информационно-поисковой системе Archie может осуществляться различными путями, начиная от запросов по электронной почте и с помощью сервиса Telnet и заканчивая использованием графических Archie-клиентов.

Система Gopher была разработана для упрощения процесса локализации FTP-ресурсов Internet и для более удобного представления сведений о содержании хранящихся на FTP-серверах файлов. Система Gopher дает возможность в удобной форме (в виде меню) представлять пользователям об имеющихся файлах и их содержании. Меню Gopher-серверов могут содержать ссылки на другие Gopher- и FTP-серверы. Таким образом, пользователь получает возможность путешествовать по Internet, не обращая внимания на местонахождение интересующих его ресурсов, и получать доступ к этим ресурсам.

Система Veronica используется для поиска информации в Gopher-пространстве по заголовкам пунктов меню. После ввода ключевого слова, система Veronica выясняет, встречается ли оно в меню на каком-либо Gopher-сервере, и в качестве результатов поиска выдает список заголовков пунктов меню, содержащих ключевое слово. Поскольку система Veronica не является автономной поисковой программой, а тесно связана с системой Gopher, она обладает тем же, что и система Gopher, недостатком: далеко не всегда по заголовку можно сказать, что собой представляет тот или иной информационный ресурс. Достоинства системы заключается в том, что нет необходимости узнавать, где расположена найденная информация, достаточно выбрать требуемую запись из списка.

3.Структура ИПС

Работа поискового указателя происходит в три этапа, из кото­рых два первых являются подготовительными и незаметны для пользователя. Сначала поисковый указатель собирает инфор­мацию из *World Wide Web.*Для этого используют специальные программы, аналогичные браузеры. Они способны скопи­ровать заданную Web-страницу на сервер поискового указателя, просмотреть ее, найти все гиперссылки, которые на ней имеют те ресурсы, найденные там, снова разыскать имеющиеся в них гиперссылки. Подобные программы называют *червяками, пауками, гусеницами, краулерами, спайдерами*и другими подобными именами. Каждый поисковый указатель эксплуатирует для этой цели свою уникальную программу, которую нередко сам и разрабатывает. Многие современные поисковые системы родились из экспериментальных проектов, связанных с разработкой и внедрением автоматических про­грамм, занимающихся мониторингом Сети. Теоретически, при удачном входе *спайдер*способен прочесать все Web-простран­ство за одно погружение, но на это надо очень много времени, а ему еще необходимо периодически возвращаться к ранее посе­щенным ресурсам, чтобы контролировать происходящие там изменения и выявлять «мертвые» ссылки, т. е. потерявшие актуальность.  
  
После копирования разысканных Web-ресурсов на сервер поис­ковой системы начинается второй этап работы — индексация. В ходе индексации создаются специальные базы данных, с помощью которых можно установить, где и когда в Интернете встречалось, то или иное слово. Считайте, что индексированная база данных — это своего рода словарь. Она необходима для того, чтобы поисковая система могла очень быстро отвечать на запросы пользователей. Современные системы способны выда­вать ответы за доли секунды, но если не подготовить индексы заранее, то обработка одного запроса будет продолжаться часами.  
  
На третьем этапе происходит обработка запроса клиента и выдача ему результатов поиска в виде списка гиперссылок. Допустим, клиент хочет узнать, где в Интернете имеются Web-страницы, на которых упоминается известный голландский механик, оптик и математик Христиан Гюйгенс. Он вводит слово Гюйгенс в поле набора ключевых слов и нажимает кнопку «Найти» (Search). По своим базам указателей поисковая система в доли секунды разыскивает подходящие Web-ресурсы и фор­мирует страницу результатов поиска, на которой рекомендации представлены в виде гиперссылок. Далее клиент может пользоваться этими ссылками для перехода к интересующим его ресурсам.  
  
Все это выглядит достаточно просто, но на самом деле здесь есть проблемы. Основная проблема современного Интернета связана с изобилием Web-страниц. Достаточно ввести в поле поиска такое простое слово, как, например, футбол, и российская поис­ковая система выдаст несколько тысяч ссылок, сгруппировав их по 10-20 штук на отображаемой странице.  
  
Впрочем, для рядового потребителя совершенно все равно, выдадут ему тысячу результатов поиска или миллион. Как правило, кли­енты просматривают не более 50 ссылок, стоящих первыми, и что там делается дальше, мало кого беспокоит. Однако клиен­тов очень и очень беспокоит качество *самых первых*ссылок. Клиенты не любят, когда в первом десятке встречаются ссылки, утратившие актуальность, их раздражает, когда подряд идут ссылки на соседние файлы одного и того же сервера. Самый же плохой вариант — когда подряд идут несколько ссылок, веду­щих к одному и тому же ресурсу, но находящемуся на разных серверах.  
  
Клиент вправе ожидать, что самыми первыми будут стоять наи­более полезные ссылки. Вот здесь и возникает проблема. Чело­век легко отличает полезный ресурс от бесполезного, но как объяснить это программе? Поэтому лучшие поисковые сис­темы проявляют чудеса искусственного интеллекта в попытке отсортировать найденные ссылки по качественности их ресур­сов. И делать это они должны быстро — клиент не любит ждать.  
  
Все поисковые системы черпают исходную информацию из одного и того же Web-пространства, поэтому исходные базы данных у них могут быть относительно похожи. И лишь на третьем этапе, при выдаче результатов поиска, каж­дая поисковая система начинает проявлять свои лучшие (или худшие) индивидуальные черты. Операция сортировки полу­ченных результатов называется ранжированием*.*Каждой най­денной Web-странице система присваивает какой-то рейтинг, который должен отражать качество материала. Но качество — понятие субъективное, а программе нужны объективные критерии, которые можно выразить числами, пригодными для сравнения.  
  
Высокие рейтинги получают Web-страницы, у которых клю­чевое слово, использованное в, запросе, входит в заголовок. Уровень рейтинга повышается, если это слово встречается на Web-странице несколько раз, но не слишком часто. Благопри­ятно влияет на рейтинг вхождение нужного слова в первые 5-6 абзацев текста — они считаются самыми важными при индек­сации. По этой причине опытные Web-мастера избегают давать в начале своих страниц таблицы. Для поисковой системы каж­дая ячейка таблицы выглядит, как абзац, и потому содержательный основной текст как бы далеко отодвигается назад (хотя на экране это и не заметно) и перестает играть решающую роль для поисковой системы.  
  
Очень хорошо, если ключевые слова, использованные в запросе, входят в альтернативный текст, сопровождающий иллюстра­ции. Для поисковой системы это верный признак того, что дан­ная страница точно соответствует запросу. Еще одним призна­ком качества Web-страницы является тот факт, что на нее есть ссылки с каких-то других Web-страниц. Чем их больше, тем лучше. Значит, эта Web-страница популярна и обладает высо­ким показателем цитирования*.*Самые совершенные поиско­вые системы следят за уровнем цитирования зарегистрирован­ных ими Web-страниц и учитывают его при ранжировании.

4.Виды ИПС

Информационно-поисковые системы (ИПС) Интернет, при всем их внешнем разнообразии, также попадают в один из этих классов. Поэтому, прежде чем знакомиться с этими ИПС, рассмотрим абстрактные алфавитные (словарные), систематические и предметные ИПС. Для этого дадим определение некоторым терминами из теории информационного поиска.

**Классификационные информационно-поисковые системы**

В классификационных ИПС используется иерархическая (древовидная) организация информации, которая называется КЛАССИФИКАТОРОМ. Разделы классификатора называются РУБРИКАМИ. Библиотечный аналог классификационной ИПС — систематический каталог. Классификатор разрабатывается и совершенствуется коллективом авторов. Затем его использует другой коллектив специалистов, называемых СИСТЕМАТИЗАТОРАМИ. Систематизаторы, зная классификатор, читают документы и приписывают им классификационные индексы, указывающие, каким разделам классификатора эти документы соответствую.

**Предметная ИПС Web-кольца**

Предметная ИПС с точки зрения пользователя устроена наиболее просто. Ищи название нужного предмета своего интереса (предметом может быть и нечто невещественное, например, индийская музыка), а с названием связаны списки соответствующих ресурсов Интернет. Это было бы особенно удобно, если полный перечень предметов невелик.

**Словарные ИПС**

Культурные проблемы, связанные с использованием классификационных ИПС, привели к созданию ИПС словарного типа, с обобщенным англоязычным названием search engines. Основная идея словарной ИПС — создать словарь из слов, встречающихся в документах Интернет, в котором при каждом слове будет храниться список документов, из которых взято данное слово.

Теория информационного поиска предполагает два основных алгоритма работы словарных ИПС: с использованием ключевых слов и с использованием дескрипторов. В первом случае, для оценки содержимого документа используются только те слова, которые в нем встречаются, и по запросу ИПС сопоставляет слова из запроса со словами документа, определяя по количеству, расположению, весу слов из запроса в документе его релевантность. Все работающие ИПС по историческим причинам используют этот алгоритм, в различных модификациях.

При работе с дескрипторами индексируемые документы переводятся на некоторый дексрипторный информационный язык. [2] Дескрипторный информационный язык, как и любой другой язык, состоит из алфавита (символов), слов, средств выражения парадигматических и синтагматических отношений между словами. Парадигматика предусматривает выявление скрытых в естественном языке лексико- семантических отношений между понятиями. В рамках парадигматических отношений можно рассматривать, например, синонимию, омонимию. Синтагматика исследует такие отношения между словами, которые позволяют объединять их в словосочетания и предложения. Синтагматика включает правила построения слов из элементов алфавита (кодирование лексических единиц), правила построения предложений (текстов) из лексических единиц (грамматика).

То есть, запрос пользователя переводится в дескрипторы и обрабатывается ИПС уже в этой форме. Такой подход более затратен по вычислительным ресурсам но и потенциально более продуктивен, так как позволяет отказаться от критерия релевантности и работать непосредственно с пертинентностью документов.

**Ранжирование результатов поиска**

Словарные ИПС способны выдавать списки документов, содержащие миллионы ссылок. Даже просто просмотреть такие списки невозможно, да и не нужно. Было бы удобно иметь возможность задать формальные критерии (хотя бы относительной) важности (с точки зрения пертинентности) документов с тем, чтобы наиболее важные документы попадали бы в начало списка. Все ИПС в настоящее время уделяют основное внимание именно алгоритму ранжирования полученных ссылок.

Наиболее часто используемыми критериями при ранжировании в ИПС являются:

* наличие слов из запроса в документе, их количество, близость к началу документа, близость к друг другу;
* наличие слов из запроса в заголовках и подзаголовках документов (заголовки должны быть специально отформатированы);
* Количество ссылок на данный документ с других документов;   
    «рекспектабельность» ссылающихся документов.

**Глава 2. Современные ИПС**

1. Сферы использования современных ИПС

Современные ИПС характерны для так называемой информационной индустрии - новейшей области экономики и социальной сферы, занятой обработкой, систематизацией, накоплением и распространением информации. Бурное развитие ИПС связано с успехами информатики (Информатика). Предметами запроса в ИПС могут быть библиографические данные, управленческая и фактографическая информация, экспертные оценки, ретроспективный опыт, результаты исследования моделей и т.д.

Такой широкий круг задач обусловливает большое разнообразие типов ИПС. Они различаются своими целями, объемом содержащихся сведений, видами информации, способами доведения ее до потребителя. Наряду с локальными ИПС, действующими в рамках одного учреждения (например, поликлиники или больницы), существуют национальные и интернациональные центры информационного обслуживания (например, в области охраны окружающей среды). Широкое распространение получили библиографические ИПС (например, содержащие библиографию по всем областям медицины и медико-биологических наук). Массовое производство персональных ЭВМ, развитие средств коммуникаций, возможность объединения ЭВМ в информационные сети и обращения со своего рабочего места к сведениям, находящимся в памяти других ЭВМ, существенно расширили диапазон применения информации, широту и глубину ее поиска. Качественно новый этап развития ИПС связан с формированием баз данных на машиночитаемых носителях. Такие базы данных позволяют обращаться к ним дистанционно, одновременно по многим запросам, получая результаты поиска оперативно и в удобном виде. Медицина и здравоохранение являются чрезвычайно специфической областью внедрения ИПС. Это связано со сложной структурой и многообразием форм медико-санитарной информации, которая включает трудно формализуемые понятия и категории, а также значительные массивы подлежащих учету данных. Особенностью медицинской информации является и то, что результаты единичных клинических или экспериментальных наблюдений по мере накопления и обобщения становятся основой для осуществления крупных здравоохранительных и социальных мероприятий. Медико-санитарная информация является базой принятия управленческих решений - от выбора наиболее важных направлений научно-исследовательской работы до проведения экстренных санитарно-профилактических мероприятий.

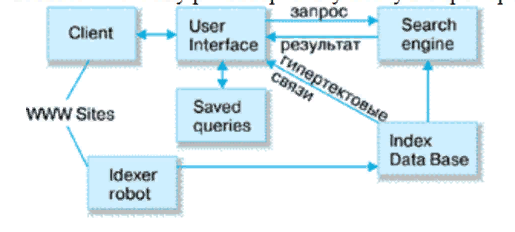
В массивы информации, на основании анализа которой осуществляется управление здравоохранением, входят статистика (демографическая и популяционная, статистика кадров, данные о заболеваемости и смертности и пр.), обобщенные данные о состоянии и достижениях медицинской и ряда смежных научных дисциплин, опыт предшествующих лет.

Именно комплексный характер сведений послужил причиной разработки единой концепции ИПС. Она включает поэтапное создание отдельных подсистем, объединение которых достигается как на уровне обмена базами данных, так и (или) с помощью средств коммуникаций. Процесс разработки и интеграции подсистем в ИПС может осуществляться по вертикали и по горизонтали по мере их создания. Подсистемы, являющиеся вспомогательными (например, учет и движение кадров, планирование и финансирование), могут создаваться независимо от других. На нижнем уровне учреждения здравоохранения (больницы, клиники, НИИ) пользуются ИПС для ведения историй болезни, контроля эффективности лечебных мероприятий, сбора и обработки первичных статистических данных, а также для решения управленческих задач своего уровня компетенции (использование коечного фонда и лабораторно-диагностического оборудования, лекарственное обеспечение и др.). Осуществляя оперативные функции, эти ИПС одновременно накапливают, а затем передают необходимую информацию на более высокий уровень (городской, областной).

Отдельно создаются подсистемы справочно-информационного обслуживания (в области библиографии и научных исследований, нормативных материалов, стандартов). В рамках общей ИПС могут разрабатываться подсистемы для поддержки и развития отдельных служб (например, психиатрической, онкологической) или целевых программ (например, побочное действие лекарственных препаратов).

2.Архитектура современных ИПС для WWW

Прежде чем описать проблемы построения информационно-поисковых систем Web и пути их решения рассмотрим типовую схему такой системы. В различных публикациях, посвященных конкретным системам, приводятся схемы, которые отличаются друг от друга только способом применения конкретных программных решений, а не принципом организации различных компонентов системы. Поэтому рассмотрим эту схему на примере, взятом из работы.



**Client** (клиент) на этой схеме - это программа просмотра конкретного информационного ресурса. Наиболее популярны сегодня мультипротокольные программы типа Netscape Navigator. Такая программа обеспечивает просмотр документов WWW, Gopher, Wais, FTP-архивов, почтовых списков рассылки и групп новостей Usenet. 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 происходил 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 хорошие ответы.

**Index robot** (робот-индексировщик) - служит для сканирования Internet и поддержания базы данных индекса в актуальном состоянии. Эта программа является основным источником информации о состоянии информационных ресурсов сети.

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

3.Популярные ИПС (2018 год)

**Краткая характеристика поисковика Яндекс**

Яндекс – самая крупная отечественная поисковая система. Дата основания – 23 сентября 1997г. В последнее время Яндекс активно выходит на международный уровень и уже имеет локализованные версии сервиса в Украине, Казахстане, Беларуси и Турции. Помимо поиска Яндекс предоставляет много дополнительных возможностей, в том числе, бесплатный хостинг Народ.ру, сервис для ведения блога Я.ру, почтовый сервис, рекламная сеть Яндекс Директ, и, кроме того, с недавнего времени Яндекс активно продвигает свой собственный браузер. Помимо основного поиска, Яндекс предоставляет возможность поиска по блогам, картинкам и видео.

В 2011 году был разработан и внедрен новый, метод машинного обучения «Матрикснет», который значительно улучшил качество поиска.

Кроме того, в 2011 году компания Яндекс разместила акции на американской бирже высоких технологий Nasdaq, что является знаковым событием для отечественных интернет-компаний.

В декабре 2012г. Был внедрен новый алгоритм «Калининград», который позволил сделать поиск персонализированным. Это значит, что теперь основным фактором ранжирования является сам пользователь, который формирует запрос для поисковой системы. Другими словами, результаты поиска по одному и тому же вопросу для двух разных пользователей теперь будет различаться в зависимости от нужд и предпочтений самих пользователей. Это новый шаг на пути эволюции поисковых систем.

Среди нововведений Яндекса можно отметить:

* введение геозависимости запросов в зависимости от региональной принадлежности пользователя и сайта;
* учет поведенческих факторов;
* разработка механизма подсказок, исправления ошибок и распознавания аббревиатур;
* активная борьба с продажными ссылками и переоптимизированными текстами;
* введение персонализированного поиска;
* учет добавочной смысловой стоимости сайта.

**Краткая характеристика поисковика Google**

Общепризнанный лидер среди поисковых систем. Поисковая система появилась в 1996 году, и изначала предназначалась для поиска в картотеке библиотеки Стэнфорда. Корпорация Google была основана в 1998 году. В настоящий момент имеет более 100 региональных версий в различных странах. Кроме того, Google это не только поиск, но и еще более 50 различных сервисов, в том числе самый популярный браузер Google Chrome.

* Проиндексировано страниц - более 3,000,000,000 страниц.
* Поддержка фреймов - полностью.
* Типы индексируемых файлов - PDF, CGI, ASP, JSP, CFM, PHP.
* Поддержка метатэгов - самый важный тэг - title.
* Обновление базы - раз в 60 дней.
* Среднее время индексации - от 48 часов до 2 недель.
* Быстрая индексация - нет.
* Предоставляет данные для - Topclick.com, Yahoo, Netscape, EarthLink.net
* Платные ссылки - Google AdWords - собственная программа контекстной рекламы по выбранным ключевым словам.
* Результаты поиска по каталогу - используется каталог Open Directory Project.
* Частота ключевых слов - ключевые слова должны упоминаться в первой трети документа. Необходимо использование синонимов и слов по выбранной теме по всему сайту. Рекомендуемая частота ключевых слов - от 0 до 12% - возможны случаи высокого ранжирования только за счет высокого PageRank.
* Link Popularity - очень важна. При отсутствии внешних ссылок документ не будет проиндексирован.

**Краткая характеристика поисковика Bing**

В настоящее время сайт Bing занимает 3-е место в списке самых популярных поисковых сайтов по объёму трафика, в отличие от которых обладает рядом эксклюзивных возможностей, таких как динамическое корректирование объёма информации, отображаемой для каждого результата поиска (например, только название, краткая или большая сводка).

* ежедневно изменяющиеся темы оформления стартовой страницы с информационными блоками;
* вывод уточняющих вариантов поисковых запросов по отдельным категориям;
* видео с автоматически запускающимся предварительным просмотром;
* дополнительные данные по каждому результату поиска;
* встроенный сервис для поиска маршрутов (другие специальные поисковые сервисы появятся в скором времени);
* функции, повышающие удобство в использовании при поиске информации, изображений и видео.

Несмотря на все возможности популярных поисковых систем, они не могут решить все проблемы. В частности, ими не всегда удобно пользоваться для поиска научной информации. Для этих целей существуют специальные научные поисковые системы.

**Краткая характеристика поисковика Mail.ru**

Каждый месяц свыше 15 млн человек из более чем 100 миллионной аудитории этого проекта используют его поиск для работы с интернетом. На сегодняшний день поисковая система маил расположилась позади двух гигантов поиска Google и Яндекс. Вот только эмэил ру не является просто очередным поисковиком, поскольку обладает исключительными возможностями представления информации.

Проект не сразу обзавёлся своим поисковиком, а определенное время функционировал на базе своих более старших братьев Яндекс и Google. Но быстрое развитие собственной технологической платформы дало возможность внедрить собственный уникальный поисковый механизм, что обладает рядом инновационных функций:

* Автоматическая генерация результатов на популярные ключи;
* Работа с соцсетью «Мой Мир» и поиск людей по ней;
* Создание проекта Товары@Mail.Ru и поиск необходимых товаров;
* Социальный сервис Ответы@Mail.Ru и поиск по проекту.’

**Краткая характеристика поисковика Rambler**

История развития и становления поисковой системы «Рамблер» берет начало в подмосковном городе Пущино. В 1991 году там появилась группа единомышленников вдохновленных идей только зарождающегося интернета. Дмитрий Крюков, Сергей Лысаков, Виктор Воронков, Владимир Самойлов, Юрий Ершов - это имена создателей Поисковой системы, которая в дальнейшем получит название Rambler. Один из основателей Сергей Лысаков при помощи своей компании "Стек" занялся локальными сетями и подключением к Интернету. Они сами создали внутри города сеть и подключили ее к Москве, а уже через Москву к мировому Интернету, присоединившись так же к сети Курчатовского института атомной энергии. И уже на следующий 1992 год компания запустила свои собственные ftp и mail серверы, а двумя годами позже в 1994, свой первый www сервер.

Самый продуктивный и "ключевой" год для российского Рунета 1996 год ознаменовался для основателей тем, что было принято решение по разработке первой русской поисковой системы для Интернета. Придумав название Rambler, к осени 1996 года был создан окончательный вариант поисковика, а 26 сентября зарегистрирован домен с названием rambler.ru. В начале 1997 года стартовал Ramblers Top100-рейтинг классификатор, который в некоторой степени явился переломным моментом в истории Рунета. С возникновением такого универсального рейтинга и счетчика, каждый вебмастер старался занять как можно более высокую позицию в рейтинге, что в свою очередь позволило Rambler-у окрепнуть и укорениться в сознании пользователей и вебмастеров. В середине 2003 года компания запустила обновленную версию Поисковой системы, которая отличалась от предыдущей по нескольким параметрам в частности поисковый индекс стал обновляться по нескольку раз в день, появилась возможность находить самые свежие документы в том числе и последние новости. Появился механизм ассоциаций, позволивший точнее формулировать свой запрос пользователям. Изменения так же коснулись и почтового сервиса и защиты данных.

2004 год ознаменовался тем, что был запущен проект Rambler Vision -это интерактивный мультимедийный портал в Рунете. В этом же году Rambler стал обладателем ряда наград, премий и дипломов. Следующий 2005 год стал для Rambler-а стал удачным в плане сотрудничества с новыми компаниями. Появилась Rambler ICQ. В 2006 году компания Rambler получает титул "супербренда" и одной из самых узнаваемых торговых марок России. В этом же году празднуется юбилей десятилетия компании Rambler.2007 год для компании стал годом совершенствования и ростом серьезных партнерских отношений с другими компаниями, благодаря чему были существенно развиты сервисы и материалы портала. В этом же году компания Rambler становится основным акционером сервиса контекстной рекламы "Бегун".К 2008 году компания Rambler продолжает четко следовать выбранному направлению, развивая при этом выбранные приоритетные направления, которые обозначены как: "Поиск, коммуникации, медиа и развлечения".

**Заключение**

Считается, что идеальная поисковая машина должна отвечать следующим требованиям:  
  
1. простота в использовании  
2. чётко организованный и обновляемый индекс.  
3. быстрый поиск в базе данных и быстрое реагирование.

4..надёжность и точность результатов поиска.

Масштабы информационных ресурсов и их количество постоянно расширяется. Становится ясно, что база данных не является совершенной. Интеллектуальные агенты – новое направление лежащее в основе нового поколения поисковых машин, которые могут фильтровать информацию и получать более точный результат. Internet продолжает развиваться с неослабевающей интенсивностью, по сути дела стирая ограничение на распространение и получение информации в мире. Однако в этом информационном океане бывает не очень легко найти необходимый документ следует также иметь в виду, что в сети наряду с давно действующими серверами возникают новые. Поисковые системы Интернета, пожалуй один из способов двигаться вперед по направлению развития и совершенствования. Развиваясь сами Поисковые системы развивают все остальное вокруг себя выгодно принуждая совершенствовать качество самого Интернета а так же технологий связанных с этим. Как и во всех других отраслях, в Интернете существуют свои лидеры, которые прошли испытание временем и добились каких либо значимых результатов, что в свою очередь позволило занять определенные ниши и позиции в мировой паутине. Поисковые системы Интернета словно огромные столпы на которых держится весь процесс движения. Конечно, много может быть несовершенно как в поиске так и в технологиях, и в наши дни несмотря на то что уже достаточная доля Интернет трафика имеет свои предпочтения все же в Интернете появляются новые поисковые системы, которые к сожалению так и остаются неизвестными.  
  
Имеется ли альтернатива для Поисковых систем Интернета? На этот вопрос однозначно ответить пожалуй не получится, так же как не получится ответить на вопрос имеется ли альтернатива автотранспорту. Возможно и есть и в случае приложения каких либо умственных усилий в поисках альтернативного направления Поисковым системам можно создать нечто новое, однако стоит ли изобретать колесо повторно... Конечно, в том виде, в котором существуют Поисковые системы сейчас в Интернете долго продержаться не смогут, так как подгоняемый технологическим процессом Интернет заставляет считаться с нововведениями и отставание от этого может пагубно отразиться на Поисковых системах, поэтому происходит постоянное совершенствование и развитие как сами Поисковых систем так и всех остальных участников Интернета.

**Список литературы**

1. Ашманов, И. С. Продвижение сайта в поисковых системах / И. С. Ашманов. - М. : «Вильямс», 2007. - 304 с.

2. Байков, В. Д. Интернет. Поиск информации. Продвижение сайтов / В. Д. Байков. - СПб.: БХВ- Петербург, 2000. - 288 с.

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

4. Чурсин, Н. А. Популярная информатика / Н. А. Чурсин.- М.: «Вильямс», 2007.- 300 с.

5. Схемы и рисунки ИПС [Электронный ресурс]. - Режим доступа : http://ssofta.narod.ru/bd/ets2.htm

6. Структура и классификация автоматизированных информационных систем Режим доступа: http://do.rksi.ru/library/courses/opais/tema1\_3.dbk

7. Схемы и рисунки ИПС [Электронный ресурс]. - Режим доступа: http://ssofta.narod.ru/bd/ets2.htm  
**8.http://poisk.sandoor.ru/**  
9.http://ru.wikipedia.org/wiki/Поисковая\_система  
10Петров, В.Н. Информационные системы [Текст]//В.Н. Петров - СПб. : Питер - 2003.   
11.Косарев, В.П., Еремин, Л.В. Экономическая информатика [Текст]//В.П. Косарев, Л.В. Еремин - Москва : Финансы и статистика – 2002.  
12. Семакин, И.Г., Хеннер, Е.К. Информационные системы и модели [Текст]//И.Г. Семакин, Е.К. Хеннер – Москва : БИНОМ – 2005.