Учреждение образования

«Брестский государственный Университет имени А.С.Пушкина»

математический факультет

кафедра информатики и прикладной математики

Курсовая работа

**Сравнительный обзор  
современных поисковых систем**

Брест, 2008

# Содержание

[Введение - 2 -](#_Toc217097874)

[Поисковые системы - 2 -](#_Toc217097875)

[Самые популярные поисковые системы - 3 -](#_Toc217097876)

[Сравнительный обзор и классификация поисковых системы - 5 -](#_Toc217097877)

[Наиболее яркие представители - 9 -](#_Toc217097878)

[Преимущества и недостатки поисковых систем - 16 -](#_Toc217097879)

[Сравнительный анализ работы различных поисковых систем: - 18 -](#_Toc217097880)

[Модель "идеальной" поисковой системы - 19 -](#_Toc217097881)

[Рекомендации по применению поисковых систем: - 21 -](#_Toc217097882)

[Заключение - 23 -](#_Toc217097883)

[Используемые источники: - 25 -](#_Toc217097884)

# Введение

Первые поисковые системы появились в сети Интернет более десяти лет назад. Тогда они выполняли лишь одну функцию – поиска ссылок к недавно созданным страницам.

На начальном этапе развития интернета, число пользователей сети было невелико и количество информации относительно небольшим. В подавляющем большинстве случаев пользователями Интернет были сотрудники различных университетов или научных организаций. В то время поиск нужной информации в сети был не столь актуален, как теперь. Сегодня же поисковые системы превратились в многофункциональный сервис. Они позволяют пользователям находить в сети Интернет самую разнообразную информацию, благодаря чему пользуются огромным успехом.

# Поисковые системы

Одной из первых попыток организации доступа к информационным ресурсам сети стало создание тематических каталогов сайтов. Первым, открывшимся в апреле 1994 г, стал Yahoo. Это еще не было поисковой системой, в современном понимании, т.к. возможность поиска информации ограничивалась ресурсами, зарегистрированными в каталоге Yahoo. Каталоги ссылок ранее использовались довольно широко, но в настоящее время практически утратили свою популярность. Объяснение этому очень простое – даже современные, содержащие огромное количество ресурсов каталоги, представляют информацию лишь о довольно незначительной части сети. Для сравнения - самый полный каталог сети интернет - DMOZ содержит информацию примерно о 12.000.000 ресурсов, в то время как база данных самой полной поисковой системы Google состоит более чем из 28.000.000.000 документов.

Первой полноценной поисковой системой в 1994г. стал проект WebCrawler. Далее в 1995 году появились поисковые системы AltaVista и Lycos. В 1997 году в Стэнфордском университете, в рамках исследовательского проекта, была создана Google - самая популярная поисковая система на данный момент в мире. В 1997 году появилась поисковая система - Yandex, лидер в русскоязычной части Интернета. На данный момент основными поисковыми системами являются три международных – Google, Yahoo и MSN Search. Остальные, коих не мало, используют целиком или частично базы и (или) алгоритмы выше приведенных систем. В Рунете основной поисковой системой является Яндекс, далее по популярности идут Rambler, Google.ru, Mail.ru и Aport.

Поисковая система - это сумма следующих компонентов:

Web server (веб-сервер) – сервер поисковой машины, который осуществляет взаимодействие между пользователем и остальными компонентами системы.

Spider (паук)- программа написанная по принципу браузера, предназначена для скачивания веб-страниц. Браузер предназначен для визуального использования страниц, а паук работает с HTML кодом напрямую. Чтобы посмотреть "сырой" исходник нажмите в меню браузера: Вид- Просмотр HTML кода.

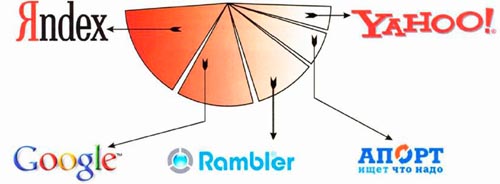
Crawler («путешествующий» паук) – программа, которая автоматически уходит по всем внешним ссылкам страницы. Ее задача - поиск не известных (или измененных) документов и в расстановке приоритетов, куда дальше должен идти Spider.

Indexer (индексатор) - программа-анализатор скаченных пауками веб-страниц. Она "разбирает" на части скачанную страницу и анализирует ее элементы, такие как текст, служебные html-теги, заголовки, особенности стилистики и структурные формы.

Database (база данных) – хранилище для скачанных и обработанных страниц - общая база данных поисковой машины.

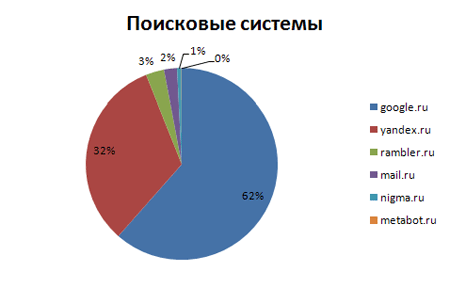
Search engine results engine (система выдачи результатов) – извлекает результаты поиска из базы данных поисковой системы. Именно она решает, какие страницы более соответствуют запросу пользователя и отсортировывает их в нужном порядке. Модуль работает согласно заданным поисковой системой алгоритмам ранжирования.

# Самые популярные поисковые системы

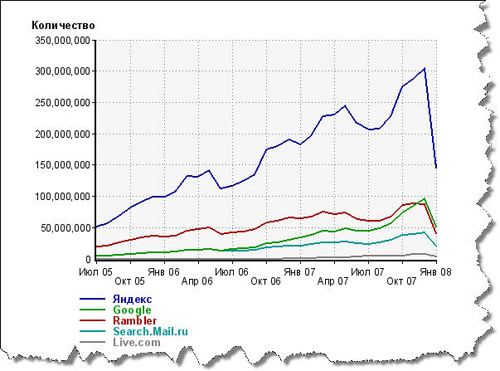


Доля каждой поисковой системе в Рунете

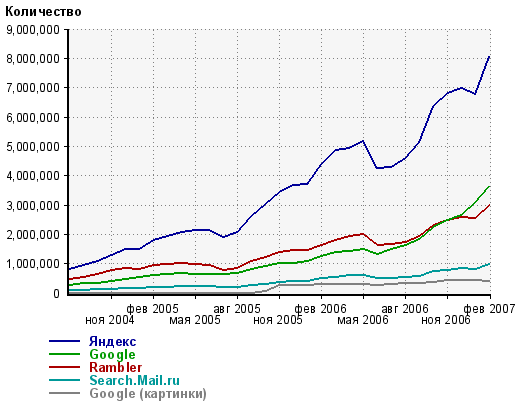
|  |  |
| --- | --- |
| **Русскоязычные** | **Англоязычные** |
| **Яndex** | **Google** |
| **Google** | **MSN search** |
| **Rambler** | **Yahoo** |
| **Аport** | **Aol** |
| **MSN Search** | **Lucos** |
| **Yahoo** | **Cuil** |
| **AltaVista** | **Ask Jeeves** |
| **Mail.ru** | **LookSmart** |



**Анализ трафика – июнь 2008г.**

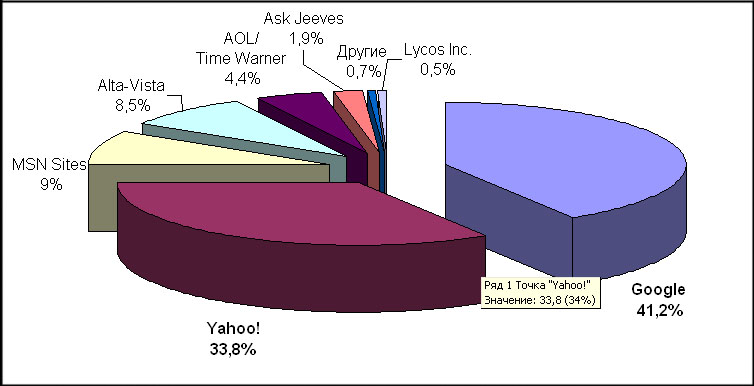


**Рейтинг поисковых систем в России**



**Количество переходов с поисковых систем**

# Сравнительный обзор и классификация поисковых системы

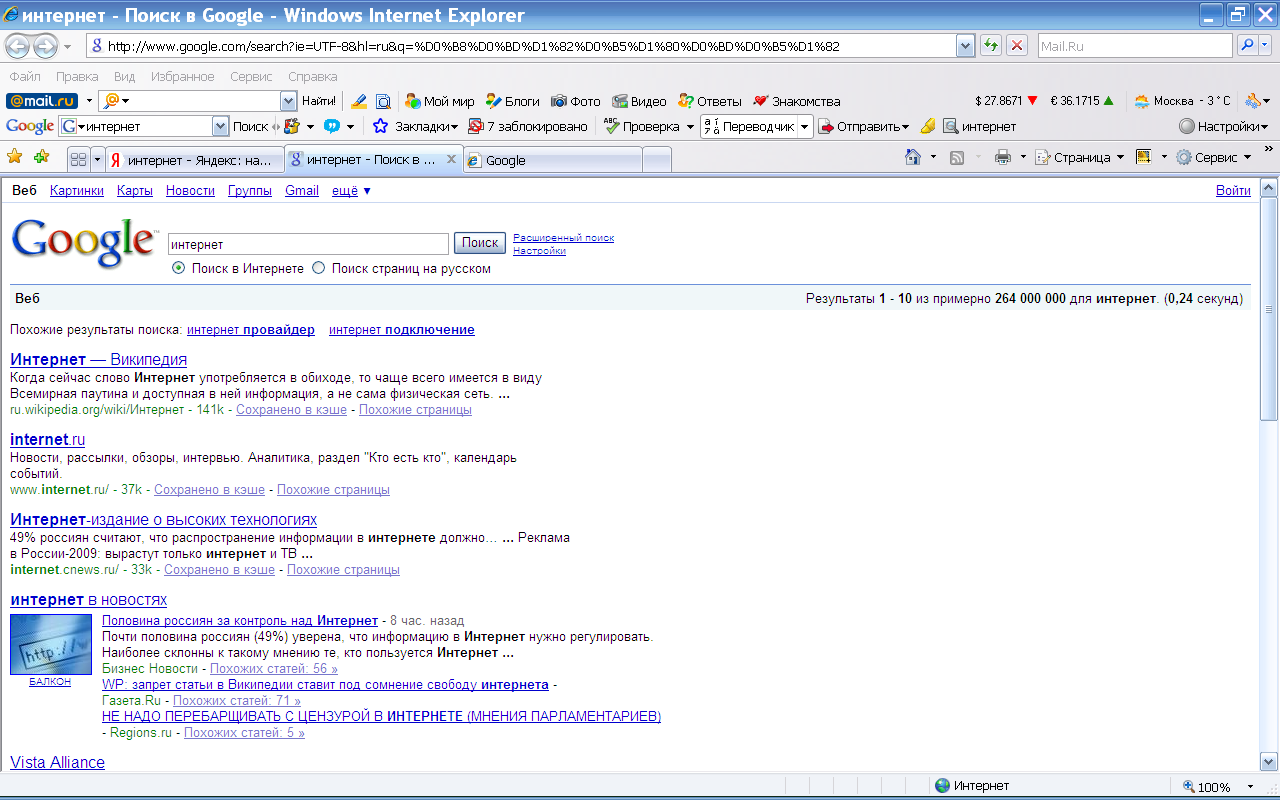


Общее соотношение поисковых систем в англоязычных странах

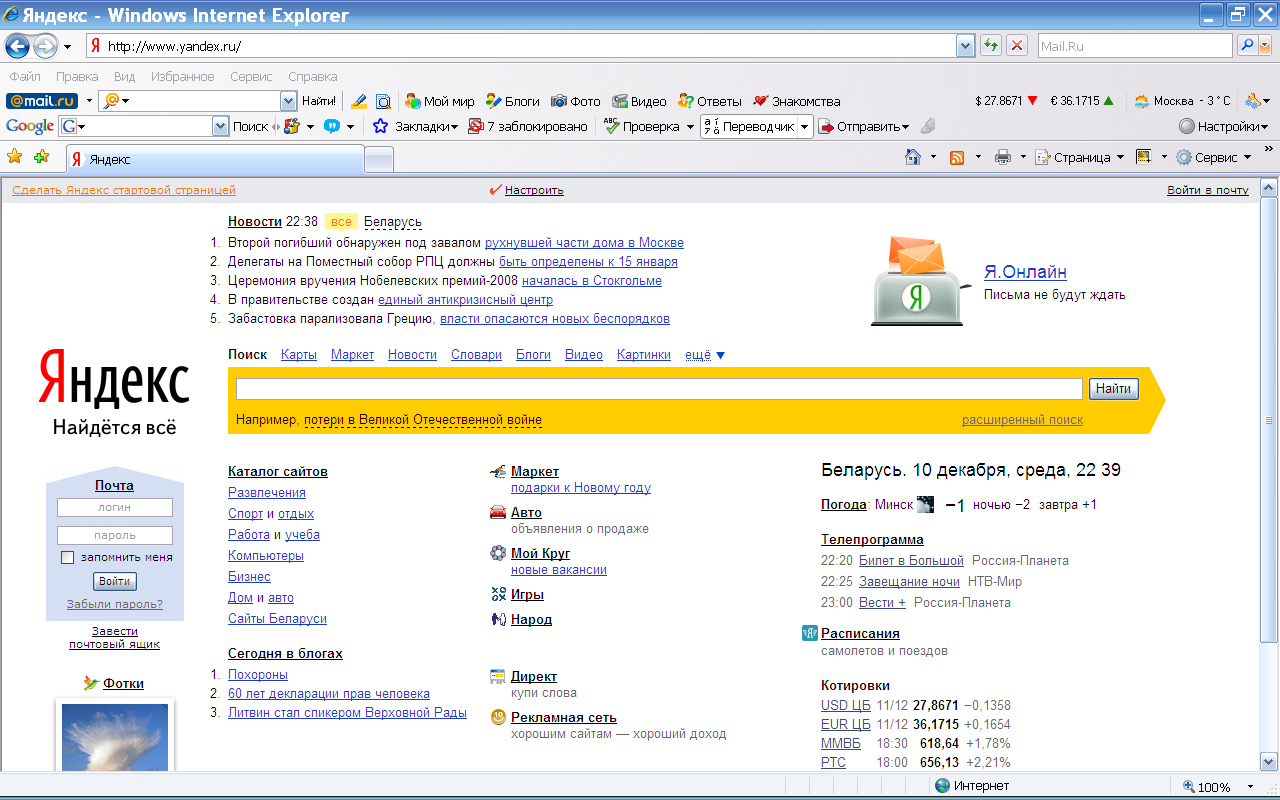
**Таблица 1**

|  |  |  |
| --- | --- | --- |
| **Google** | **Яndex** | **Rambler** |
| 1997г.  Служба основана на работе crawler’ов, охватывающей всю Сеть. Приоритетное место занимает релевантность результатов и их соответствие запросу. В отличие от многих поисковиков, роботы Google индексируют все страницы, а не только самые главные. Есть меню, где с легкостью сможете находить в Сети изображения, обсуждения, проходящие в группах новостей, читать новости или проводить поиск товаров. Можно получить доступ к информации из Open Directory (Открытого Каталога, DMOZ)**.** Система долго хранит проиндексированные роботами страницы в своем кэше, что позволяет «оживлять» мертвые, удаленные с серверов страницы, или видеть старые версии только что обновленных страниц. Система предоставляет отличную проверку правописания, легкий доступ к словарным определениям, просмотр биржевых ставок, дорожные атласы, базы номеров телефонов и многое другое. Программа AdWords размещает рекламные объявления на сайтах Google и его партнеров. Также Google является бесплатным поставщиком результатов поисковых запросов некоторых других поисковых систем**.** | 1997г.  В системе производится поиск с учетом морфологии русского языка, поиск с учетом расстояния, и тщательно разработанный алгоритм оценки релевантности.  Реализован естественно-языковой запрос: теперь поисковик можно спрашивать «по-русски», задавая длинные вопросы. Например: «где купить холодильник».  Поисковый робот позволяет предоставить возможности поиска по разным зонам текста, ограничение поиска на группу сайтов, поиск по ссылкам и изображениям. Существует «индекс цитирования» - количество сайтов, ссылающихся на данный ресурс. Работает фильтрация результатов поиска от мата и порнографии.  Службы: почта, новости, открытки и закладки, автоматическое объединение новостей в сюжеты и выделение главных тем дня. Плюс, стилизованный под Google Toolbar, спартанский поисковик ya.ru. Поиск теперь ведется не только по веб-страницам, но и по специализированным массивам данных – новостям и товарам. Находит документы не только в формате HTML. | 1996г.  Система работает с большой скоростью поиска, обновление поискового индекса происходит несколько раз в день. Поисковик всегда находит самые свежие документы и последние новости. Rambler знаком с русским языком, понимает термины типа «б/у», «у.е.» и «а/я». Услуги: строка поиска, почта и новости. Особенность - механизм ассоциаций. Помимо стандартной ответной страницы, в которой найденные документы расположены в порядке убывания релевантности, появляется строка - «У нас также ищут». В ней приведены слова и словосочетания тематически связанные с запросом. Новый механизм также исправляет ошибки запросов. На введенное gjujlf пользователь получит теперь положительный результат: сведения о погоде (gjujlf – это слово «погода» набранное латиницей). Rambler Mass Media - первый и по-прежнему лидирующий на рынке информационный дайджест. Помимо текстовых материалов и фоторепортажей, ставших визитной карточкой проекта, Rambler Mass Media представляет пользователям аудио и видеосюжеты. |
| **Aport** | **MSN Search** | **Yahoo!** |
| 1996г.  Продолжает работать под управлением OC Windows. К важным свойствам первой версии поисковой системы Апорт можно отнести перевод запроса и ответа на английский язык и обратно. Второе свойство – реконструкция проиндексированных страниц из собственной базы. Это дает возможность просмотра уже несуществующих страниц.  Система построена на основе выдачи результатов по отдельно взятым сайтам. Для разделения ресурсов на сайты Апорт использует информацию, которую предоставляет каталог AtRus, или владельцы ресурсов.  В поисковой системе Aport впервые была реализована возможность поиска по новостным лентам. | 1994г.  Ранее у MSN никогда не было собственного паука или каталога. С 1997 года для выдачи результатов поиска использовались разные базы данных. Только с начала 2005 года MSN запустил бета-версию собственного поискового алгоритма.  Система МSN особенно важна, т.к. именно этот поисковик по умолчанию используется, когда пользователи Internet Explorer'а вводят в адресную строку поисковый запрос.  Локализована примерно для 30 стран, таких как: Великобритания, Дания, Бельгия, Новая Зеландия, Япония и др. Также эта система предоставляет пользователям возможность сортировать результаты поиска: по дате, по алфавиту, по релевантности. | 1995г.  На сегодняшний день это самый старый и наиболее полный каталог Интернет-ресурсов. Система зародилась как простая коллекция закладок, которую пополняли всего 2 человека - Дэвид Фило и Джерри Янг. Вдобавок к релевантной выдаче результатов поисковых запросов Yahoo предлагает возможность использовать ярлыки над поисковой формой на своей главной странице для поиска изображений, товаров или перехода на Желтые Страницы Yahoo. Можно производить поиск по каталогу Yahoo. Есть панель для браузера Firefox. Для  хранения почты система предлагает 1 гигабайт бесплатного места или 2 по годовому абонементу. Yahoo продает место на сайте для рекламных ссылок; продажа ведется через программу Overture. Система учитывает орфографические ошибки, позволяет искать слово лишь по первым буквам. |
| **AltaVista** | **Teoma** | **LookSmart** |
| 1995г.  Особенность этого поисковика заключается в возможность вести поиск по усложненным критериям отбора.  Обеспечение поддержки множества языков (Китайский, Японский, Корейский). Содержит услуги по индексации больших объемов информации и возможность мгновенного поиска в огромных базах данных.  Есть схема «спонсорских» сайтов. Осуществляет разработку корпоративных поисковых систем внутреннего пользования. Лицензирует технологии поиска предприятиям, в том числе для использования во внутренних сетях. Она пользуется каталогами Yahoo, и ярлыки над поисковой строкой позволяют находить в сети изображения, музыку, видео, а также тематические разделы, наполненные вручную. Если вы хотите получать информацию не менее надежную чем от Yahoo, но поставляемую через более простой интерфейс – AltaVista ваш выбор. | 2001г.  Поисковая система, основанная на работе crawlerов (с англ. пауков" - специальных программных роботов, сканирующих Интернет в поисках информации, которая затем заносится в базу данных), принадлежащих Ask Jeeves. Большая релевантность поиска, встроенная система проверки правописания запросов, динамические описания страниц, появляющиеся рядом с результатами поиска. Запросы можно вводить на десяти европейских языках. В баз данных поисковика проиндексированы свыше миллиарда уникальных страниц. Существует возможность уточнения запроса в тематических топиках, ссылки на которые выдаются на странице с результатами поиска. Раздел «Ресурсы», также находящийся на странице результатов, в свою очередь уникален тем, что перенаправляет пользователя на сайты, которые послужили исходным материалом для приведенных в качестве результатов статей. | 1995г.  LookSmart – это интернет-компания, имеющая свой интернет-каталог и механизм поиска Wisenut. LookSmart - это, в первую очередь, вручную пополняемый перечень сайтов. Он получает списки двумя путями. Во-первых, коммерческие сайты платят за место в торговых топах, уподобляя LookSmart электронным Желтым Страницам. Во-вторых, существуют добровольцы, которые заносят сайты с некоммерческим содержанием в каталог www.zeal.com, принадлежащий LookSmart. Хотя это и отдельный сайт, списки, находящиеся на нем, используются при обработке поисковых запросов LookSmart. Изюминка LookSmart – это возможность доступа к сотням периодических изданий через один из ярлыков над строкой поиска. Содержит службу Furl, предназначенную для архивирования веб-страниц и ссылок на них. Для архива страниц пользователя предоставляется 5 Гб свободного пространства. Запустила систему вертикального поиска из 161 сайта. |
| **Scirus** | **GigaBlast** | **Aol Search** |
| 2001г.  Специализированная поисковая система научной и околонаучной информации. Scirus использует около 200 миллионов страниц научной тематики и отфильтровывает результаты ненаучного содержания, которые по релевантности в какой-то степени соответствуют запросу.  Стартовая страница Scirus радует своей непритязательностью – там только Вы и строка запроса.  Если Вы не очень сведущи в научных терминах, Scirus помогает ввести корректный запрос для лучшего поиска.  Это лучшая из научных специализированных поисковых систем. Она предоставляет информацию конкретного направления, помогает в составлении запросов и использовании терминов, а кроме того Scirus находит информацию, которая может быть недоступна в случае использования других поисковых машин. | 2000г.  Поисковая система с более чем двух миллиардным количеством индексированных страниц. GigaBlast имеет высокую релевантность запросу и множество различных инструментов поиска. Стартовая страница GigaBlast.com проста и аккуратна. В дополнение к обычной строке запроса в наличии имеются Каталог, Поиск по блогам, Поиск Путешественника и Правительственный Поиск.  Система использует обычный прямой поиск -  Соответствие результатов довольно высоко, кроме того, GigaBlast проводит проверку правописания. Поддерживает запросы в буквальной форме («Почему трава зеленая?»). Проводит проверку правописания. В расширенном поиске GigaBlast.com можно искать по точной фразе, ставить исключение на слово, искать на определенном адресе и т.д. Все стандартно.  В целом GigaBlast является удобной машиной поиска с высоко релевантными результатами. | 1985г.  Огромный портал, на котором можно найти все, начиная с игр и заканчивая форумами самой различной тематики + простота навигации и хорошо запоминающееся название обозначающие America Online. AOLbyPHONE – Сервис позволяет получать заказанную ранее информацию по телефону. AOLSchool – сервис для школьников. DigitalCity – проект виртуального города. AOL Anywhere – сервис позволяет получить доступ с любого мобильного устройства к контенту America Online.  AOL Search предоставляет пользователям страницы с результатами поиска на основании индекса Google, «Встроенная» версия AOL Search в результате поисковых запросов дает ссылки только на содержимое, доступное внутри онлайн системы AOL. А пользуясь AOL Search, вы можете одновременно проводить поиск и по Интернет, и по Сети AOL. Во «внешней» поисковой системе AOL Search такой опции нет. Если вам по душе Google, вам будет не хватать некоторых его особенностей в AOL Search, например, кэширования страниц. |
| **Blink TV** | **Infoseek** | **Mail.ru** |
| 2004г.  Машина поиска, позволяющая находить в интернете аудио и видео файлы, используя не только ключевые слова и фразы, но и фактическое содержание аудио и видеоклипов. Можно вносить поправки в поиск, например, по дате (если ищете материал за определенный период времени) и по релевантности (если хотите получить результаты, наиболее соответствующие тому, что Вы ищете). Есть Smart папки – это папки, содержимое которых автоматически пополняется из интернет источников без участия пользователя. Они формируются из результатов поиска по вашему запросу. Использование Blinkx TV достаточно просто и удобно (единственное требование – это высокоскоростная связь с Интернетом). | 1994г.  В этой системе индекс создает робот, но он индексирует не весь сайт, а только указанную страницу. Имеется возможность указания ключевых фраз, представляющих собой единое целое, вплоть до порядка слов. Ранжирование при выдаче осуществляется по числу терминов запроса в документе, по числу фраз запроса за вычетом общих слов. Все эти факторы используются как вложенные процедуры. Infoseek относится к традиционным системам с элементом взвешивания терминов при поиске, а так же позволяет проводить то, что называется контекстным поиском. Это значит, что используя специальную форму запроса, можно потребовать последовательной совместной встречаемости слов, обладает довольно развитым информационно-поисковым языком (не просто указывает, какие термины должны встречаться в документах, но и своеобразно взвешивать их). | 1997г.  Не является поисковой системой, это огромный портал со множеством сервисов. Но в сервисы mail.ru входит и поисковая система. Ведет активную борьбу со спамом и вирусами, сотрудничая по этим направлениям с создателями «Dr.Web» и «Антивирус Касперского». В 2004 году присоединилась к инициативе почтовых служб мира и начала использовать стандарт SPF в работе фильтров своих почтовых серверов. В поисковом механизме существует опция морфологии русского языка и возможность ведения истории поисковых запросов. Обзавелась двумя инновационными сервисами: доступом по протоколу IMAP и голосовой почтой - возможностью прослушивания содержимого ящика по сотовому телефону. К порталу Mail.ru добавились такие новые сервисы как «Хостинг» и «Энциклопедии». |
| **Lucos** | **NetScape Search** | **Ask Jeeves** |
| 1945г.  Система предоставляет доступ к результатам, собранным вручную из LookSmart по результатам наиболее частых запросов, а также из Yahoo на основе индексов сайтов, составленных роботами.Система Fast Forward позволяет видеть на одной стороне экрана результаты поиска, а на другой – первые страницы найденных сайтов. Релевантные запросу категории из каталога, собранного вручную, показываются вверху страницы с результатами поиска. | 1994г.  Система пользуется индексной базой Google, как и основной поисковый сайт AOL – AOL Search. Основное отличие Netscape Search от Google в том, что Netscape Search в первую очередь будет производить поиск по собственной базе и выводить эти результаты первыми. Есть разница и в дизайне. | 1998г.  Прославилась как поисковая машина, понимающая «естественную речь» - в поисковую строку вводится предложение в вопросительной форме, и система адекватно отвечает на него. В настоящее время Ask Jeeves работает на основе технологии crawlerов, обрабатывая запросы, поступающие в поисковую систему Teoma. |
| **Cuil** | **Alexa Internet** | **HotBot** |
| 2007г.  Создана выходцами Google. Отличительная черта – большой объем обрабатываемых страниц. Понимает запросы только на английском языке. | 1996г.  Крупнейшая поисковая система, предоставляющая также интернет статистику о трафике web сайтов.  Сегодня Alexa – это машина поиска, использующая базы данных двух гигантов мирового интернет поиска: Google и каталог DMOZ. | 1996г.  HotBot предоставляет простой доступ к трем основным поисковым системам, в основе которых лежит работа crawler’ов: Google, MSN и ASK. Не смешивает результаты, полученные из трех этих источников. Тем не менее, это быстрый и простой путь к тому, чтобы получить несколько независимых «мнений» по интересующей вас теме. |

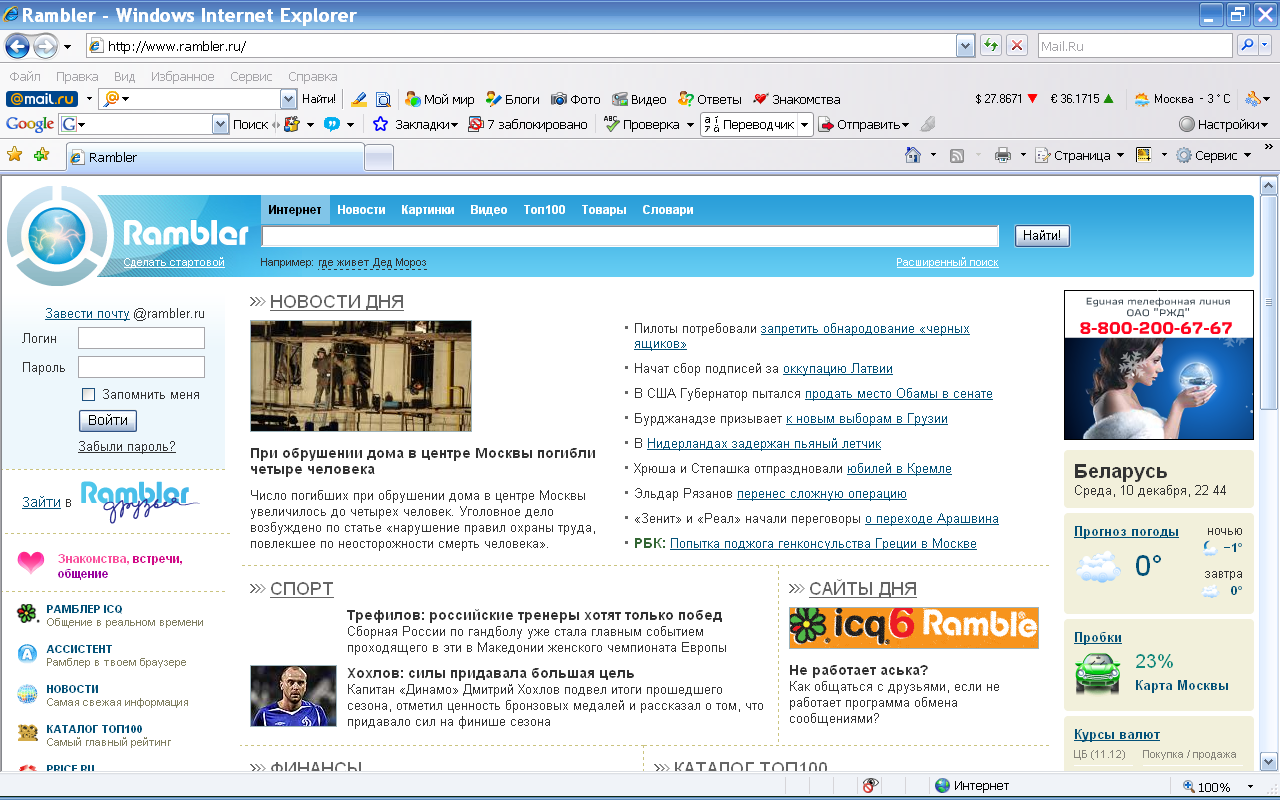
# Наиболее яркие представители



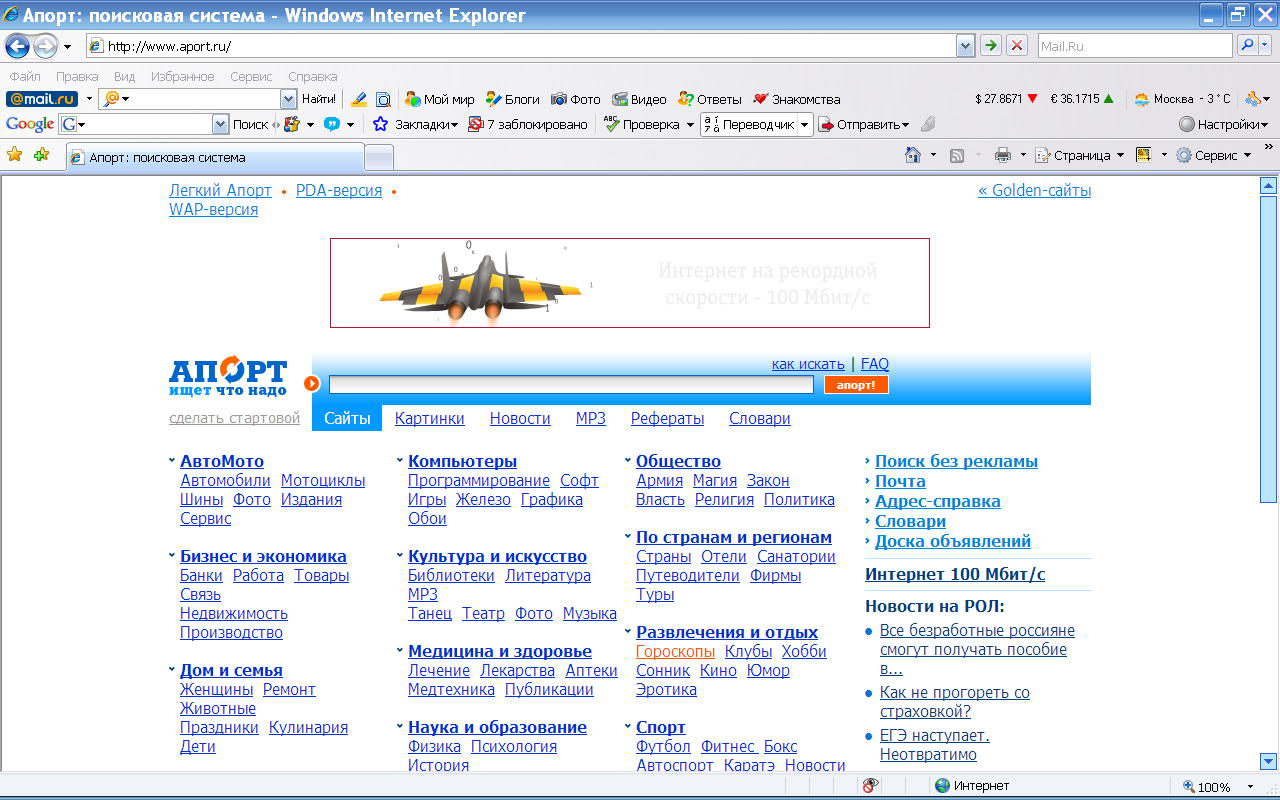
|  |
| --- |
| **Google (www. google.ru)** |
| По статистике является самой популярной поисковой системой в мире.  Основатели Google - аспиранты Стэнфордского университета Сергей Брин и Ларри Пейдж. Из скромной затеи двух энтузиастов она выросла в интернет-гиганта, услугами которого ежедневно пользуются миллионы человек в 80 странах мира. В одном из интервью Брин, родившийся в Москве, объяснил появление Google так: "Когда мы начинали, было несколько поисковиков, но качество их работы оставляло желать лучшего.  Каждый запрос возвращал вам около тысячи результатов, отсортированных в случайном порядке". Именно релевантность результатов, их соответствие запросу, была поставлена во главу угла новой системы.  Помимо просто поиска, здесь можно сравнивать цены на товары в интернет-магазинах, читать новости и многое другое. Есть и служба блокировки назойливой интернет-рекламы.  Google первым применил новую технологию поиска и выдачи результатов - релевантность интернет-страниц - то, что не лежит на данной конкретной странице, но тем ни менее влияет на ранг ее выдачи. Поисковая система Google учитывает гипертекстовую структуру Интернета, и какие страницы на какие ссылаются. Google сама определяет релевантность интернет-страниц Страница, на которую ссылаются чаще, более релевантна и значит более популярна.  Лидерство Google никто не может оспорить. Акции компании за 2005 год выросли почти в два с половиной раза. Поисковый трафик за тот же период вырос примерно на 30%.  Коммерческое использование Google запущен в 2000 году и благодаря своему программному обеспечению, обеспечивающему прекрасную релевантность результатов поиска, а также дополнительным сервисам (такому, как поиск новостей, встраиваемый в браузер "Google bar" с "гасителем всплывающих окон") быстро набрал популярность среди "продвинутых" пользователей, которые использовали Сеть в первую очередь как рабочий инструмент. Эффективную систему "тематической рекламы" также оценили рекламодатели. Не случайно Washington Post назвала сервис Google одним из 10 важнейших интернет-достижений 2003 года.Сейчас Google ежедневно обрабатывает более 200 миллионов запросов. В своей работе в настоящее время Google применяет более 15 тысяч компьютеров.  Объем проиндексированных страниц более 10 миллиардов!  Справка: ... на сегодняшний день, благодаря дистрибьюторским соглашениям с Yahoo, AOL и Ask Jeeves, Google ежедневно обрабатывает до 80% всех поисковых запросов, сделанных в интернете.  Для сравнения. Яндекс в настощее время имеет в день 600 тысяч посетителей в день, количество компьютеров 40 единиц, проиндексировано 15 миллионов страниц.  Google - лидер в списке самых популярных поисковиков. В результате исследования, проведенного компанией comScore, выяснилось, что Google - самая популярная из поисковых систем в США.  В августе этот поисковик обработал 63% поисковых запросов, что на 1,1% превышает показатели июля. На втором месте оказалась поисковая система Yahoo, обработавшая 19,6% запросов (20,5% в июле), а на третьем - Microsoft (доля обработанных запросов - 8,3%). Замыкают пятерку самых популярных поисковиков Ask.com - 4,8% (доля обработанных запросов возросла на 0,3%) и AOL - 4,3% (рост - 0,1%). Первое место занял Google (41,3 млрд. поисковых запросов; 62,4% рынка), второе — Yahoo! (8,5 млрд. запросов; 12,8% рынка) и китайский поисковик Baidu.com (3,4 млрд. запросов; 5,2% рынка). |



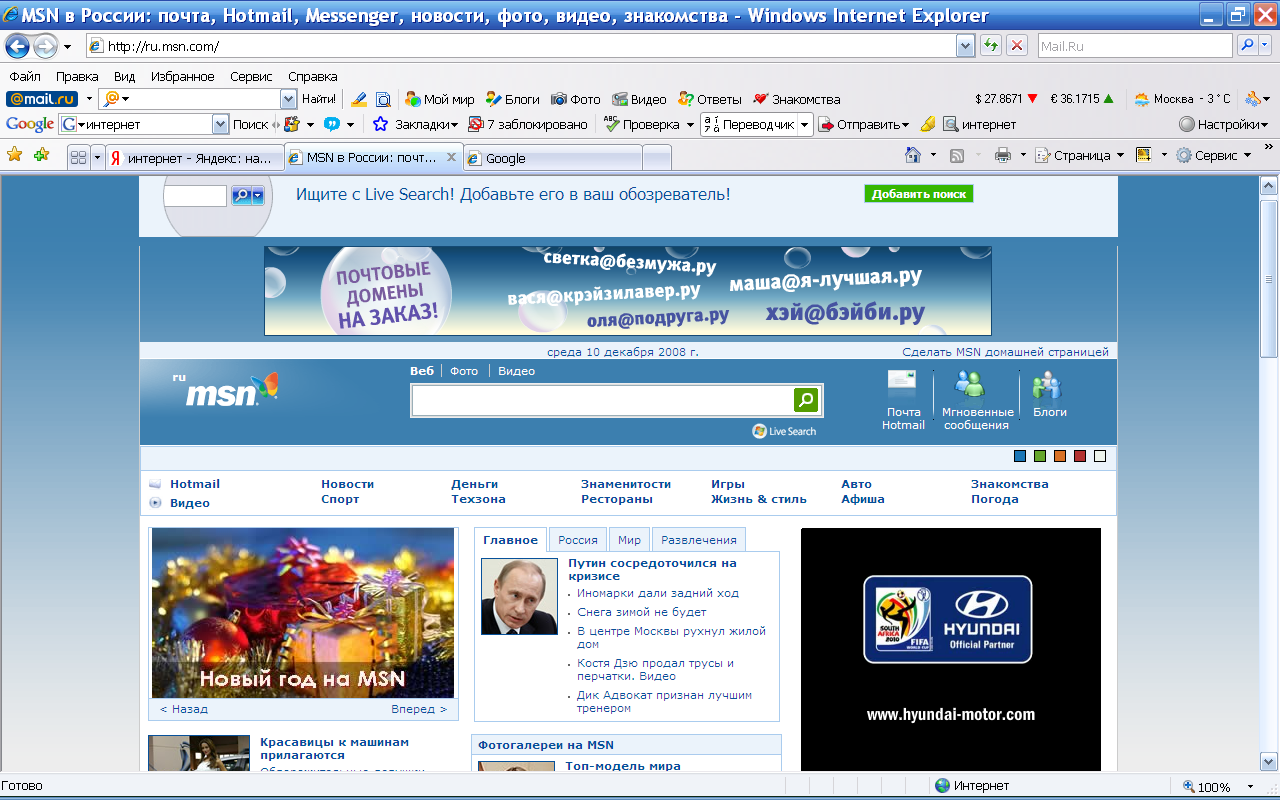
|  |
| --- |
| **Яndex (www.yandex.ru)** |
| Поисковик Yandex появился 23 сентября 1997 г. Разработчик, компания CompTek, представила новую поисковую систему на выставке Softool. Ранее эта компания занималась производством информационно-поисковых систем. Одним из ее проектов стала система, обеспечивающая поиск с учетом морфологии русского языка. Она занимала 300кб, грузилась в оперативную память полностью и работала очень быстро. На основе этой технологии был создан «Библейский компьютерный справочник» (поисковая система, обеспечивающая работу с Ветхим и Новым Заветом).  Основными нововведениями поисковой системы www.yandex.ru были: проверка уникальности документа – этим достигалось исключение копий в различных кодировках. И отличительные свойства поискового алгоритма Yandex. Поиск с учетом морфологии русского языка, поиск с учетом расстояния, и тщательно разработанный алгоритм оценки релевантности. Несколько месяцев спустя, в поисковой системе www.yandex.ruбыл реализован естественно-языковой запрос. Теперь поисковик можно спрашивать «по-русски», задавая длинные вопросы. Например: «где купить холодильник» или «генетически модифицированные продукты».  В 1998-99гг. вместе с ростом Рунета выросли и услуги поисковой системы Yandex. Увеличение количества запросов привело к необходимости оптимизации поискового механизма. Новый поисковый робот позволил предоставить возможности поиска по разным зонам текста, ограничение поиска на группу сайтов, поиск по ссылкам и изображениям. Также, впервые в Рунете, было введено понятие «индекс цитирования» - количество сайтов, ссылающихся на данный ресурс. Был открыт «Семейный www.yandex.ru», фильтрация результатов поиска от мата и порнографии.  В 2000 году поисковая система Yandex была преобразована в юридическое лицо. Тогда же начинается активное продвижение поисковика в традиционных, оффлайновых СМИ. Агрессивная, но эффективная рекламная кампания принесла свои плоды – слоган «Найдется все», некоторое время спустя, становится нарицательным. Для привлеченных рекламой пользователей поисковая система www.yandex.ru открывает новые службы – почта, новости, открытки и закладки. Плюс, стилизованный под Google Toolbar, спартанский поисковик ya.ru. Поиск теперь ведется не только по веб-страницам, но и по специализированным массивам данных – новостям и товарам.  2003 год - поисковая система Yandex продолжает развиваться. Яндекс научился находить документы не только в формате HTML. Служба Yandex Новости радикально преобразилась: пользователям предложен оригинальный сервис – автоматическое объединение новостей в сюжеты и выделение главных тем дня.  На сегодняшний день www.yandex.ru- крупнейший российский портал: его доля среди поисковых машин Рунета составляет около 45%, в то время как идущий следом Rambler может похвастаться лишь 22%. Каждый день более 500 серверов компании отдают 2.7 терабайт трафика двум с половиной миллионам посетителей.  “Яндекс” вошел в топ-10 поисковых машин. Согласно отчету, опубликованному компанией comScore, поисковая система Яндекс попала в десятку самых популярных поисковиков в мире по результатам декабря прошлого года. Согласно представленному отчету, в декабре было сделано 66,2 млрд. поисковых запросов. 566 миллионов из них были заданы Яндексу, что позволило сервису занять 9-е место в списке поисковиков. Доля Яндекса на мировом рынке поисковых сервисов составила 0,9%. |



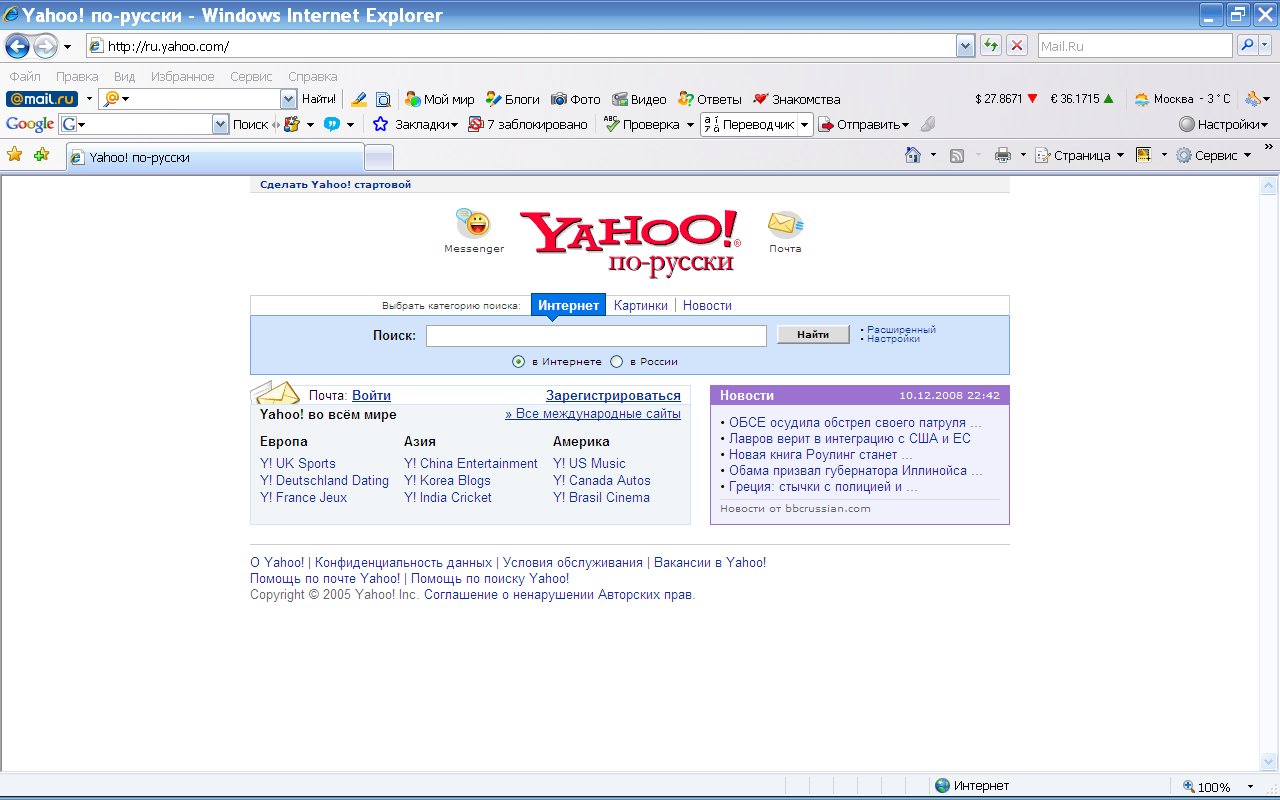
|  |
| --- |
| **Rambler (www.rambler.ru)** |
| Rambler – первая поисковая система русского Интернета, появилась на свет в 1996 году. Ее создатели, группа «Стек» из подмосковного Пущино, начинали с создания локальной городской сети, и подключения к Интернету. В 1992 году «Стек» запускает собственные ftp- и mail-серверы. Следующим шагом к светлому будущему, стало создание поисковой системы. Дмитрий Крюков и Сергей Лысаков начали с того что дали проекту название «Rambler». Им понравилось значение этого слова «скиталец, бродяга», выражающее саму суть поисковика. К осени 1996 года, окончательный вариант поисковой системы был готов, а 8 октября «Rambler» был запущен.  Еще год система отлаживалась, к чести создателей поисковик не работал всего одни сутки – сгорели жесткие диски. Существовавшие на то время в Рунете две – три поисковых системы не выдержали конкуренции и «канули в Лету».  Весной 1997 года появляется уникальный рейтинг-классификатор Rambler’s Top –100, который не только оценивал популярность российских ресурсов, но и позволял одним кликом попасть на них. Rambler’s Top –100 быстро стал универсальным барометром сети, общим стандартом медиаизмерений.  Следующий шаг в развитии поисковой системы Rambler произошел в июне 2003 года. Компания запустила новую версию поисковой машины, которая отличается от предыдущей по двум основным параметрам. Первое и главное – значительно увеличилась скорость поиска; второе, благодаря новой архитектуре системы, обновление поискового индекса происходит несколько раз в день.  Теперь поисковик всегда находит самые свежие документы и последние новости. Рамблер знаком с русским языком, понимает термины типа «б/у», «у.е.» и «а/я». Как дань моде, открылась лаконичная версия поисковой системы по адресу «r0.ru». Содержание ее стандартно – строка поиска, почта и новости.  Еще одна уникальная особенность поисковика Rambler, это механизм ассоциаций. Помимо стандартной ответной страницы, в которой найденные документы расположены в порядке убывания релевантности, появляется строка - «У нас также ищут». В ней приведены слова и словосочетания тематически связанные с запросом. Новый механизм также исправляет ошибки запросов. На введенное gjujlf пользователь получит теперь положительный результат: сведения о погоде (gjujlf – это слово «погода» набранное латиницей).  Необходимо отметить проекты Rambler-Группы и Rambler-Фото. Оба этих проекта предоставляют пользователям сети возможность использования заложенных в проекты программных алгоритмах в собственных целях. Проект Rambler-Группы дает людям возможность начинать обсуждения самых разных тем, которые существуют не только в виде форумов, но также и в виде открытых списков рассылки. Проект Rambler-Фото представляет пользователям сети возможность создания фотоальбомов в объеме до 20 мегабайт, причем это дисковое пространство зарегистрированные пользователи поисковой системы Rambler получают в дополнение к тем мегабайтам почты, на которые они могут претендовать по новым условиям e-mail обслуживания (15 мегабайт).  Rambler Mass Media - первый и по-прежнему лидирующий на рынке информационный дайджест. Помимо текстовых материалов и фоторепортажей, ставших визитной карточкой проекта, Rambler Mass Media представляет пользователям аудио и видеосюжеты. В числе поставщиков - ВГТРК и BBC, Deutsche Welle, радио "Маяк" и др. Заголовки важнейших новостей размещаются на первой странице портала Rambler.ru и в блоке "Главные новости дня" на главной странице Rambler Mass Media. Наиболее актуальные новости могут читать и посетители других сервисов Рамблера - Рамблер Почты, Погоды на Рамблере. |



|  |
| --- |
| **Aport (www.aport.ru)** |
| Поисковая система Апорт впервые была презентована на пресс-конференции компании «Агама» в феврале 1996 года. Тогда система работала только по сайту russia.agama.com. К моменту официальной презентации, 11 ноября 1997 года, в поисковой системе Апорт был проиндексирован первый миллион документов, расположенных на 10 тысячах серверов. Отцом www.aport.ruстала компания «Агама» - разработчик программного обеспечения для Windows. Это отразилось и на создании поисковой системы, в отличие от большинства конкурентов Aport продолжает работать под управлением ОС Windows.  К важным свойствам первой версии поисковой системы Апорт можно отнести перевод запроса и ответа на английский язык и обратно. Второе свойство – реконструкция проиндексированных страниц из собственной базы. Это дает возможность просмотра уже несуществующих страниц.  Ноябрь 1998 года – Aport продан гражданину Израиля Джозефу Авчуку. Сумма сделки составила 55 тысяч долларов. Торговые марки «Апорт» и «Агама» сохранились, каталогу Ау! повезло гораздо меньше. В 1999 году Авчук окончательно покупает каталог и переименовывает его в AtRus, а после и вовсе уничтожает при экспорте на сайты «Омен», «Россия он-лайн» и «Апорт».  Конец 1999 года – в поисковую систему Апорт вложен первый миллион долларов, это позволило, некоторое время спустя, представить на компьютерных выставках «Апорт - 2000». Полностью интегрированный с AtRus; теперь Каталог – Апорт (Catalog - Aport).  Поисковая система Aport 2000 была построена на основе выдачи результатов по отдельно взятым сайтам. Для разделения ресурсов на сайты Апорт использует информацию, которую предоставляет каталог AtRus, или владельцы ресурсов.  www.aport.ru – первым из поисковых систем Рунета реализовал базовые технологии Google. «Page rank» - характеристика популярности ресурса по формуле «обратных ссылок»: ссылки с других сайтов на данный ресурс. Причем учитывается не только количество, но и важность ссылок. Вес ссылки с популярного сайта больше, а ссылки включающие слова запроса ценятся выше, чем ссылки со словом «здесь». К тому же, при обработке запроса поисковая система Aport 2000 ориентируется на HTML – код страницы, и наличие слов запроса в URL.  Еще одна особенность - приоритет сайтам, вошедшим в высшую и элитную лиги Апорт – Каталог. В поисковой системе Aport впервые была реализована возможность поиска по новостным лентам.  Первым www.aport.ru стал использовать платную нулевую строку в выдаче. Но на Апорт купить не нулевое, просто более высокое место в ответной странице.  Масштабируемость в архитектуре Aport 2000 построена таким образом, что его поисковую базу можно дробить на несколько отдельных фрагментов, каждый из которых будет работать на своем компьютере. После поиска по фрагментам для пользователя рассчитывается общий ответ.  31 июля 2000 года Golden Telecom купил семейство интернет-проектов "Агама", включающее "Апорт" и AtRus, для включения в "Россию-он-лайн".  Май 2001 - года окончательно завершилась сделка по смене владельца "Апорт" самого "Golden Telecom", новым владельцем стал "Альфа-Банк". NASDAQ к тому времени переживал бурный спад, и шансов перепродать Интернет проекты за приемлемую сумму не было. Это обусловило решение новых хозяев "Golden Telecom" минимизировать расходы на поддержку дорогостоящих Интернет проектов. |



|  |
| --- |
| **MSN Search (www.msn.ru)** |
| Когда-то MSN Search был одним из лучших поисковиков, но и сегодня он по-прежнему стоит внимания. MSN Search пользовался каталогами сайтов LookSmart и получал лучшие отзывы пользователей, поскольку за проектом стояла команда редакторов, которая отслеживала наиболее частые запросы, вводимые в поисковую форму, и подбирала сайты, наиболее релевантные тематике запроса. Такая методика давала хорошие результаты.  Сегодня MSN Search ищет новые пути развития. Система производит поиск по материалам Yahoo, но не отличается той функциональностью и многообразием параметров, которые предлагает сам Yahoo. В то же время, MSN Search разрабатывает собственные роботы и планирует другие перемены, которые должны полностью обновить систему.  (www.msn.com)  У поисковой системы МSN (msn.com) никогда не было собственного паука или каталога. С 1997 года система МSN для выдачи результатов поиска использовала разные базы данных, такие как: Yahoo!, LookSmart, Altavista, DirectHit, Inktomi и RealNames. Сейчас результаты поиска берутся из Inktomi, LookSmart, Direct Hit.  Но, все-таки, MSN.com один из самых популярных порталов Интернета. Там же есть и поисковая система search.mns.com. Система МSN особенно важна, т.к. именно этот поисковик по умолчанию используется, когда пользователи Internet Explorer'а вводят в адресную строку поисковый запрос. Система МСН также является информационным узлом, и по данным MediaMetrix, является одним из наиболее посещаемых сайтов.  msn.com использует базу данных Inktomi, но эта база данных сейчас содержит более более 550 миллионов URL и не известно, какую часть использует поисковая система MSN. Известно, что из базы данных LookSmart используется около 2 миллионов URL.  Поисковая система МСН (msn.com) локализована примерно для 30 стран, таких как: Великобритания, Дания, Бельгия, Новая Зеландия, Япония и др. Также эта система предоставляет пользователям возможность сортировать результаты поиска: по дате, по алфавиту, по релевантности.  Недавно, в конце 2004 года, портал MSN.com объявил о запуске официальной версии своего поиска, спустя два года после объявления о начале разработки собственного поисковика. Ранее продемонстрированная бета-версия поиска выкачена на всех сайтах портала и потеряла отметку "beta". По сравнению с уже знакомой бета-версией, нынешний поиск приобрел несколько новых возможностей, а в частности:  - Интеграция с MSN Direct Answers - службы, базирующейся на электронной Microsoft Encarta и расширенной теперь до 1,5 миллионов статей.  - Выдача результатов поиска в RSS-формате - при этом существует ограничение на использование этих результатов и запрет на их использование в коммерческих целях.  - Поиск по RSS.  Хотя при создании поисковика MSN.com использовалось большинство принципов оптимизации, которые и в других поисковых системах не будут сюрпризом для большинства профессиональных участников он-лайн рынка.  В связи с запуском поиска проведен также редизайн всего сайта MSN.com, а на портале опубликовано обращение Билла Гейтса к пользователям. В ближайшее время MSN планирует запустить также поиск по блокам, несколько специализированных поисков (можно предположить, что это будет поиск по видеофайлам - он уже есть у Yahoo и Google) и собственную программу контекстной и поисковой рекламы. Сейчас основная масса рекламы, демонстрируемой на результатах поиска, является рекламой Overture, которая принадлежит одному из конкурентов новой системы MSN - Yahoo! Inc.  Но, поисковик уже приобрел значительное влияние и, если верить слухам, MSN Search все-таки бросил вызов Google и Yahoo в борьбе за первенство на рынке. |



|  |
| --- |
| **Yahoo! (www.yahoo.ru)** |
| В 1994 году, студенты Стэндфордского университета, Джерри Янг и Дэвид Фило, готовились к защите диссертации в области компьютерного проектирования интегральных схем. Для этого им приходилось много времени проводить в сети Интернет, в поисках нужной информации и копить ссылки. Списки со ссылками росли, потом Янг и Фило забросили диссертацию и принялись исключительно коллекционировать ссылки. К середине 1994 года их стало много, они отсортировали ссылки по категориям, потом в категориях ссылок стало тоже много, появились подкатегории. Но список Джерри и Дэвида не был предназначен для всеобщего обозрения - он составлялся исключительно для друзей. Время шло, а посещаемость все росла и росла. Адрес сайта пошел по рукам.  Первым шагом к успеху стало новое, запоминающееся название – Yahoo! Следуя пожеланиям пользователей, создатели www.Yahoo.com, стали преобразовывать сайт. Появились новые категории, и разделы «What's New» и «What's Cool». К концу 1994 Янг и Фило забросили свои диссертации и полностью отдались работе над поисковиком Яху.  В это время на дороге появилась компания Netscape, предложившая ресурсы для содержания поисковой системы Yahoo!. В результате у Yahoo! появился свой домен - yahoo.com, и каталог переехал на 10 станций Silicon Graphics Indy. Примерно в это же время Yahoo! получил и первого инвестора - инвестиционный фонд «Seqouia Capital». Джерри и Янг обзавелись офисами и наняли энергичную команду web-серферов. Темп роста составил, в среднем, 1000 страниц в день.  12 апреля 1996 года www.Yahoo.com выпустил свои акции. При номинальной цене в 12 долларов за штуку к концу дня они продавались уже по 33 доллара. Создатели Яху первыми стали применять рекламу своего ресурса в печатных СМИ и на телевидении.  Новый управляющий поисковой системы Yahoo! - Тим Кугл решил развить Яху из простого перечня упорядоченных линков до настоящего портала. А из финансовых новостей сделать аукционы, добавляя к ядру Yahoo.com элементы онлайнового доступа.  Следующие вложения в поисковик Yahoo! составили 560 миллионов долларов. Yahoo.com взрослеет и начинает действовать на рынке как самостоятельная акула капиталистического мира. Конкуренты, в лице America Online и Microsoft, сбиваются в различные корпоративные альянсы, чтобы укреплять свои позиции в бизнесе. Подсмотрев у "врагов" методы, Yahoo в 1999 году покупают поставщика бесплатной электронной почты rocketmail.com. В то время на просторах Интернета появляются различные сайты, претендующие на звание Портал. С конкурентами расправляются по-разному. Webring.com покупают, с остальными ведут войну на бирже.  В 2000 году поисковая система Yahoo запустили проект под названием Corporate Yahoo! и купили тайваньский бизнес-сайт Kimo, чем заявили о желании двигаться на обширный китайский Интернет-рынок.  На сегодняшний день Yahoo.com одна из основных поисковых систем в Интернет. Предоставляющая услуги 200 миллионам человек в месяц.  Нынешнее развитие Яху можно определить как движение в он-лайн, интерактивность. Яху 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 "он-лайновая" 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 Яху, 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, Яху 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.  Для начала www.Yahoo.com 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 comScore Media Metrix 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).  11 октября 2005 года, стремясь ни в чем не отстать от крупнейшего конкурента, компания Yahoo объявила о намерении создать цифровую библиотеку. Причем не слишком успешный опыт Google, погрязшего в судебных разбирательствах с правообладателями, компанию не пугает, а скорее напротив — подстегивает.  Напомним, что Google некоторое время тому назад заключила договор на оцифровку книг в крупнейших европейских и американских университетах и библиотеках, но попытка создать крупнейшую в мире библиотеку электронных текстов вызвала яростное сопротивление опасающихся за свои доходы издателей и авторов.  Целями же Яху стали книги из библиотеки Калифорнийского университета, от издательства О`Рейли, прославившегося своими учебниками, и Национального Архива Великобритании. Войдут в «стартовый пакет» и книги из фонда Европейского Архива. Избежать затяжных судов компания намерена путем предложения потребителям получить разрешение на просмотр данных текстов от правообладателей перед переходом на «полновесный» текст. При этом поисковая система Yahoo рассчитывает поделиться с другими поисковыми сервисами возможностью проводить поиск в этой «библиотеке», в то время как Google возможности своей несостоявшейся библиотеки делить ни с кем не планировала.  В середине 2005 года, компания Yahoo, официально объявила о достижении важного преимущества в борьбе со своим главным конкурентом на рынке сетевого поиска — компанией Google. Менеджеры считают, что поисковик Yahoo позволяет найти в два раза больше документов, чем Google. По их мнению, база данных поисковой системы Yahoo включает в себя 20,5 млрд. объектов — 19 миллиардов текстовых документов и 1,5 миллиардов изображений. Таким образом, ее поисковый индекс (число объектов, которые пользователь может найти в интернете при помощи поисковика), почти в два раза превышает аналогичный показатель поисковой системы интернет-корпорации Google — 11,3 миллиардов объектов, из которых 8,2 миллиардов текстовых документов и 3,1 миллиардов изображений.  Ранее Yahoo никогда не раскрывала значение индекса своей поисковой системы. Отраслевые аналитики не были единодушны в оценках ее базы данных. По мнению различных экспертов, поисковик Yahoo охватывал от 6 до 8 миллиардов объектов в сети. Согласно официальному сообщению, впервые в истории Yahoo, качественный прорыв в конкурентной борьбе с Google обеспечила последняя, весьма успешная модернизация поисковой системы. «У пользователей интернета появилась очень веская причина обратить на нас внимание, — заявляет вице-президент корпорации Yahoo Экхарт Уолтер. — Наша поисковая система охватывает намного больше информации, чем любые другие».  «Yahoo заставила пользователей усомниться в том, что Google — самая мощная поисковая система. Это дает компании Яху дополнительное преимущество в борьбе за лидерство на рынке. Бренд Google для многих стал синонимом поиска информации в интернете. Задача Yahoo заключается в том, чтобы пошатнуть доверие к этому бренду», — считает аналитик компании Forrester Research Карлин Ли.  Вот чем живет www.Yahoo.com сегодня, когда стоимость одной акции уже подкралась к $200. Сейчас поисковая система Yahoo! достигла критического размера - команде серферов с трудом удается поддерживать ее актуальность и добавлять тысячи новых сайтов ежедневно. |

# Преимущества и недостатки поисковых систем

**Таблица 2**

|  |  |  |
| --- | --- | --- |
| **Поисковая система** | **Преимущества** | **Недостатки** |
| **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. |
| **Яндекс** | 1)Постоянное 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. |
| **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. |
| **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. |
| **MSN** | 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. |
| **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 |

# Сравнительный анализ работы различных поисковых систем:

1. Сформулировала тему поиска.
2. Создала поисковый запрос (одинаковый для всех поисковых систем).
3. Результаты поиска (количество ссылок) поместила в таблицу.
4. По результатам поиска создала уточняющие поисковые запросы (с учетом правил работы в конкретной поисковой системе), чтобы уменьшить количество нерелевантных ссылок. Уточняющие запросы создавала до тех пор, пока результаты поиска станет возможно проанализировать – 10 ссылок.
5. Результаты уточняющего поиска (количество ссылок) поместила в таблицу.
6. Проанализировала 10 первых найденных ссылок на предмет релевантности.
7. Сделала вывод о качестве поиска каждой поисковой системой и поставила оценки по пятибалльной шкале.

**Таблица 3**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Что такое тезаурус и каковы правила его создания?** | | | | | | | | | |
| **Поисковые системы** | **Яndex (стр.)** | **Rambler (док-ов)** | **Google (сайтов)** | **Aport (сайтов)** | **MSN (сайтов)** | | | **Yahoo! (сайтов)** | |
| **Поисковый Запрос** | 75.000 | 2.676 | 24.600 | 1.946 | | 9.770 | | | 16.800 |
| **Уточняющий Запрос** | 7.453 | 189 | 20.400 | 443 | | | 543 | | 417 |
| **Анализ первых 10 найденных ссылок** | Нет ничего  Нужная ссылка оказалась на 2 странице под номером 13 | Ничего нет  Нужная ссылка оказалась на 5 странице | Ничего нет  После просмотра 10 страниц = 100 ссылкам – ничего не найдено | Нужная ссылка на первой странице под номером 10 | | Ничего нет  Нужная ссылка оказалась на 3 странице | | | Ничего нет  Нужная ссылка оказалась на 2 странице под номером 16 |
| **Кол-во релевантных ссылок** | 0 | 0 | 0 | 1 | | 0 | | | 0 |
| **Вывод-Оценка по пятибалльной шкале** | 5 | 4 | 2 | 5 | | 4 | | | 4 |

# Модель "идеальной" поисковой системы

Поисковые системы являются инструментом, которым сегодня пользуется большинство посетителей Интернета. Современные поисковые системы начинают тяготеть к порталам, расширяя предлагаемые возможности и не ограничивая пользователей только поиском.

Поисковики – это своего рода монополисты. Они ведут скрытую политику, диктуют свои правила, не разглашая их при этом. А подчас, даже просто беспредельничают и показывают поразительную необъективность по отношению к сайтам. Возникает ситуация, что оптимизаторы не могут ясно понять, каким он должен быть, «хороший» сайт в понимании поисковика. Как сделать его таким, чтобы поисковик считал его наиболее релевантным по запросам. И именно из-за этой неопределённости и возникают у вэб-мастеров желание обхитрить систем.

Выход из этой проблемы – создание полностью ОТКРЫТОЙ поисковой системы, в которой бы ранжированием занимался не только поисковик но и сами веб-мастера**.** Чтобы в ней возможно было открыть все механизмы и алгоритмы ранжирования и при этом объективно оценивать сайты между собой.

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 конфидициальности. 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 минипоисковой системы. Эта минипоисковая система будет индексировать содержимое сайта по правилам самого обладателя сайта. Только вэбмастер будет решать, какие страницы его сайта по каким запросам более релевантны. 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:

* **Primary nouns (**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.

Для получения всех данных, необходимых для составления идеальной модели, было проведено интервью небольшого числа пользователей, а затем – Интернет-опрос. Интервью позволило понять, какие вопросы и каким образом нужно задать, чтобы получить полную картину: кто является пользователем поисковой системы, каковы его потребности и задачи, каким образом он привык решать задачи и каким видит путь решения новых, нестандартных задач.

Проанализированные ответы позволили построить модель идеальной поисковой системы. В результате исследования были получены выводы о том, чего не хватает пользователям в имеющихся поисковых системах, а также того, что не нравится пользователям. Эта информация позволила выделить прототипы, опираясь именно на портрет целевой аудитории, которая будет работать с поисковой системой, учитывая предпочтения пользователей и особенности их работы. Итак, что у нас получилось?

**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 MSN Search. 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.

Согласно моему исследованию 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;
* Апорт;
* Google;

**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:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 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.

**Общий вывод:** по рассмотренным критериям 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. Отметим также, что поисковая система MSN лидирует в системе ранжирования.

# Рекомендации по применению поисковых систем:

Первый вопрос, который возникает у начинающего пользователя Интернета: "Как найти необходимую мне информацию?" Прежде всего, мы должны отметить, что нет "правильных" или "неправильных" путей поиска в Интернете. Несомненно, одни из них могут потребовать больше времени, чем другие, но в конечном счете предпочтение должно отдаваться тому методу, который кажется наиболее удобным и эффективным конкретному пользователю Интернета.

Представьте себе, что Вам нужно найти потерявшуюся вещь в Вашем доме. Вы можете использовать несколько способов поиска:

**1**. Планомерно обыскать все шкафы и полки;

**2**. Искать только там, где, по Вашему мнению, может находиться эта вещь;

**3**. Спросить родных: маму, детей, жену или мужа о месте, где может находиться искомый предмет или нечто похожее на него.

Конечно, могут быть и другие способы... Но в каждом отдельном случае только один из трех перечисленных методов будет наиболее успешным. Аналогичный принцип применяется при поиске в любой поисковой системе: Вы можете искать в новостях, файлах или во всем Интернете. Вы даже можете искать только на определенном сайте (только в одном шкафу), если Вы знаете, что искомый объект спрятался именно там. Но следует учитывать, что поисковые роботы могут обойти только те места, куда его пускают (заглянуть только в те шкафы, дверцы которых не заперты).

Точно так же, как и дома, в каждом отдельном случае один из способов может оказаться эффективнее другого: в Интернете нужный объект или сайт может найтись вовсе не там, где ожидалось

**Начните искать — это легко**

Поисковые машины не требуют от вас знания специальных команд для поиска. Достаточно набрать вопрос — так же, как бы вы его задали библиотекарю или всезнайке-эрудиту. Например, где раки зимуют? Ярчайшая звезда северного полушария?

**Морфология учитывается**

Независимо от того, в какой форме вы употребили слово в запросе, поиск учитывает все его формы по правилам русского языка. Это называется учет морфологии слов. При необходимости вы всегда можете его отключить.

**Что такое правильный запрос?**

Правильный запрос состоит из нескольких слов, потому что по одному слову обычно трудно понять, что вы хотите найти. Например, запрос дизайн может означать желание найти информацию о дизайне интерьера, или о веб-дизайне, а может — о ландшафтном дизайне. По такому запросу поисковик найдет страницы обо всех известных ему видах дизайна.Обычно поисковые системы хорошо понимает запросы из нескольких слов и способны самостоятельно их интерпретировать. Они знают, в каких случаях надо искать страницы, в которых слова запроса расположены рядом, а когда достаточно словам встретиться в одном документе, чтобы страница хорошо соответствовала запросу.

**Ищете цитату?**

Если вы хотите найти точную цитату, просто задайте искомую фразу в кавычках: "о закрой свои бледные ноги". По такому запросу будут найдены все страницы, в которых есть нужная вам цитата.

**Ищете имена?**

Слова с большой и маленькой буквы считаются разными формами одного слова, поэтому все равно, какой регистр использовать в запросе. Исключением является оператор точной формы. По запросу '!лужков' будут найдены все документы, содержащие эту словоформу в любом регистре, а по запросу '!Лужков' — только документы, в которых имеется форма 'Лужков' с большой буквы. Это полезно, если искомое имя собственное совпадает с распространенным именем нарицательным, например, село !Кошки или !Прост.

**Обязательно найдите слово!**

Чтобы отобрать документы, где определенное слово присутствует обязательно, поставьте перед ним плюс (без пробела). Поэтому, если вас интересует цитата из «Гамлета», вы можете задать запрос +быть +или +не быть. Маленькая хитрость: тот же результат можно достичь, задав «обязательные» слова в кавычках: "быть или не быть".

**Как исключить слово?**

Чтобы исключить документы, в которых встречается определенное слово, поставьте перед этим словом два символа тильды или один минус (без пробела). Например, если вам нужно описание Парижа, а не предложения многочисленных турагентств, задайте запрос путеводитель по Парижу ~~агентство ~~тур. Или запрос путеводитель по Парижу -агентство –тур.

# Заключение

Поисковые системы обычно состоят из трех компонент:

* агент (паук или кроулер), который перемещается по Сети и собирает информацию;
* база данных, которая содержит всю информацию, собираемую пауками;
* поисковый механизм, который люди используют как интерфейс для взаимодействия с базой данных.

Во время путешествия по Интернету, вам обязательно понадобиться помощь поисковой машины. Очень часто приходится искать информацию в сети не зная 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. База данных содержит URL-адреса и проиндексированную информацию, связанную с этими адресами.

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 (насколько велика база данных по документам), 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).

Кроме релевантности, существуют важные пользовательские характеристики: скорость поиска (медленная поисковая машина неэффективна в работе), поисковые возможности (как именно происходит индексация: только по ключевым словам web-страницы или по всему тексту, с учетом морфологии или без него, с поиском по тэгам HTML - заголовкам, ссылкам, подписям к изображениям и др.), а также дополнительные удобства (удобный интерфейс, наличие специальных функций, например, поиск по датам и серверам). Здесь все зависит от того, что вы предпочитаете.

Среди ведущих поисковых машин на данный момент - Яndex, Google, Rambler, Апорт! и др.

**Анализ поисковых систем Интернет**

Исходя из темы курсовой работы **"Сравнительный обзор современных поисковых систем"**.

Результаты поиска представлены в следующей таблице.

**Таблица 4**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ***Yandex***  страниц | ***Google***  сайтов | ***Aport***  cайтов/  документов | ***Rambler***  документов | ***Yahoo***  сайтов | ***MSN***  сайтов |
| Современные поисковые системы | 5 000 000 | 4 730 000 | 13 054/  65 133 | 107 045 | 483 000 | 591 000 |
| Популярные поисковые системы | 10 000 000 | 438 000 | 4 238 /  42 618 | 54 016 | 944 000 | 148 000 |
| Обзор поисковых систем | 1 000 000 | 149 000 | 7 038/  84 553 | 13 364 | 139 000 | 28 200 |
| Преимущества и недостатки поисковых систем | 53 000 | 8 620 | 218/ 298 | 2 258 | 11 200 | 3 230 |
| Сравнительный анализ работы поисковых систем | 809 000 | 71 400 | 369 / 917 | 10 069 | 28 100 | 13 900 |
| «Идеальная» поисковая система | 83 000 | 9 210 | 4 458 /  23 838 | 5 172 | 9 060 | 43 600 |
| Рекомендации по применению поисковых систем | 635 000 | 124 000 | 926/  1 811 | 29 147 | 35 200 | 72 800 |

**Общее заключение о наличии материалов в Интернет по теме курсовой**

Степень актуальности того или иного предмета исследования определяется, главным образом, исходя из объема существующей по данному вопросу литературы. В ходе осуществленного поиска в Internet мною было найдено большое количество информации, касающейся классификации, обзора и анализа современных поисковых систем. Исходя из объема представленной литературы как на английском, так и на русском языках, можно сделать вывод, что к настоящему времени  поисковые системы пользуются огромным спросом среди пользователей сети Internet.

**Анализ статистики по таблице 4**

По данным, представленным в таблице поиска, можно анализировать эффективность каждой из представленных семи поисковых систем с точки зрения количества найденной информации, а именно сайтов, имеющихся в системе по отдельно заданному запросу. Анализ я осуществляла по следующему принципу: Учитывалась особенность поисковых систем по языковой принадлежности, анализ производился по 6-ти системам. В ходе анализа статистики мною были выявлены следующие закономерности: наибольшее количество сайтов по запросу на русском языке среди русскоязычных систем было выдано поисковыми системами Yandex и Google, при этом показатели Yandex значительно превышают Google, далее в порядке снижения эффективности следуют системы Aport, Rambler и MSN соответственно Yahoo, как англоязычная система, показала наихудший результат при поиске ключевых слов, заданных на русском языке. Таким образом, лидирующие позиции по показателям количества представленной информации занимают поисковые системы Yandex и Google и Aport.

**Анализ различных поисковых систем по степени эффективности**

Лидерами по показателям качества представленной информации оказались Yandex, Google и Aport.

Yandex оказался одной из наиболее эффективных систем с точки зрения ее релевантности и соответствия выданных результатов заданному запросу. Хоть страниц было много, но нужная информация находилась на самых первых из них. Мало затраченного времени – необходимые результаты. При этом немаловажную роль сыграла также относительная новизна представленной информации.

Google выдавал результаты страниц, на которых не всегда первое место занимали релевантные документы. Зато жалоб на разнообразие просто не было, т.к. в представленном количестве материала можно было найти что угодно ( при этом было потрачено времени в два раза больше, по сравнению с поисками в других поисковых системах).

Система Aport оказалась менее эффективной, чем вышеназванные из-за ее чрезмерной ориентации на частные случаи, но результаты, которые она выдавала, значительно отличались от результатов других поисковых систем. Они были единственные в своем роде, не всегда релевантны, но неповторимы.

Rambler, несмотря на прочно занимаемое четвертое место в количественном рейтинге, оказался намного менее эффективной по релевантности системой. Здесь преобладает ориентация на российские источники информации, что снижает ее адекватность в оценке ситуации в других странах. К этой же категории по степени релевантности можно отнести и поисковую систему Google.

Поисковую систему Yahoo можно рассматривать как наиболее эффективную наравне с Yandex, но только в англоязычном поиске. На русском языке в данной системе имеется незначительное количество сайтов и их релевантность минимальна.

**Общий анализ ситуации с поиском и предоставлением информации в современном Internet**

Исходя из личного опыта постоянного пользователя я должна сказать, что на данный момент в условиях постоянных качественных изменений во всех сферах жизни и постоянном движении человеческой мысли вперед, Internet является наиболее важным и наиболее часто используемым источником информации. Наибольшая полезность глобальной сети состоит в доступности информации любому пользователю и регулярной обновляемости ресурсов, что позволяет постоянно быть в курсе новых событий.

# Используемые источники:

1. www. google.ru
2. www.yandex.ru
3. www.rambler.ru
4. www.aport.ru
5. www.msn.ru
6. www.yahoo.ru
7. http://company.yandex.ru/articles/article10.xml
8. http://www.citforum.ru/internet/search/searchsystems.shtml
9. http://www.intelsib.ru/
10. http://www.seonews.ru/news/.info\_news/2385/
11. http://www.seo-gu.ru/im\_stat.html
12. http://www.relevantno.ru/news/html/1138782965.html
13. http://www.vadimstepanov.ru/f\_texts/column6.htm
14. http://book.itep.ru/4/45/retr4514.htm
15. http://www.seoexp.com/ru/history/search\_engines/
16. http://smo.me/uncategorized/история-создания-компании-google/
17. www.google.com/codesearch