

## **Содержание:**

### **Введение**

Поисковая система (поисковик) – это сайт, обратившись к которому пользователь может найти интересующую его информацию по заданному ключевому запросу. На сегодняшний день поисковые системы самый лучший инструмент для поиска информации в Интернете.

Рассмотрим принцип работы поисковика, который довольно прост. Пользователю, пришедшему на сайт системы необходимо ввести в форму, располагающуюся на сайте ключевую фразу, по которой он ищет информацию, и послать запрос, нажав кнопку поиск. После чего он получит результат в виде списка текстовых ссылок на сайты соответствующие данному запросу. Это принцип работы поисковика со стороны пользователя. Ниже рассмотрим процесс работы (который не заметен пользователю) и внутреннее устройство.

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

Одним из первых способов организации доступа к информационным ресурсам сети стало создание открытых каталогов сайтов, ссылки на ресурсы в которых группировались согласно тематике. Первым таким проектом стал сайт Yahoo.com, открывшийся весной 1994 года. После того, как количество сайтов в каталоге Yahoo значительно увеличилось, была добавлена возможность поиска нужной информации по каталогу. В полном смысле это еще не было поисковой системой, так как поисковая область была ограничена только ресурсами, присутствующими в каталоге, а не всеми Интернет ресурсами.

Каталоги ссылок широко использовались ранее, однако практически полностью утратили свою популярность в настоящее время. Так как даже современные, огромные по своему объему каталоги, содержат информацию лишь о ничтожно малой части сети Интернет. Самый большой каталог сети DMOZ (его еще называют Open Directory Project) содержит информацию о 5 миллионах ресурсов, тогда как база поисковой системы Google состоит из более чем 8 миллиардов документов.

Первой полноценной поисковой системой стал проект «WebCrawler», вышедший в свет в 1994 году. Основным отличием поисковой системы от своих предшественников является предоставление возможности пользователям осуществлять поиск по любым ключевым словам на любой веб-странице. Сегодня эта технология является стандартом поиска любой поисковой системы. Поисковая система «WebCrawler» стала первой системой, о которой было известно широкому кругу пользователей.

В 1995 году появились поисковые системы Lycos и AltaVista. В 1996 году AltaVista запустила морфологическое расширение для русского языка и стала первой поисковой системой, которая была доступна русскоязычным пользователям Интернета. В этом же году были запущены первые отечественные поисковые системы – «Rambler.ru» и «Aport.ru». Появление первых отечественных поисковых систем ознаменовало новый этап развития Рунета, позволяя русскоязычным пользователям осуществлять запрос на родном языке, а также оперативно реагировать на изменения, происходящие внутр

## **ГЛАВА 1.ИНТЕРНЕТ**

### **1.1 Сети**

С запуском в 1997 году поисковой системы «Яндекс» отечественные поисковые машины начали конкурировать между собой, улучшая систему поиска и индексации сайтов, выдачи результатов, а также предлагая новые сервисы и услуги.

В 1997 году Сергей Брин и Ларри Пейдж создали поисковую машину Google в рамках исследовательского проекта в Стэнфордском университете. В настоящий момент Google - самая популярная поисковая система в мире, которая дала пользователям возможность осуществлять качественный поиск с учетом морфологии, ошибок при написании слов, а также повысить релевантность в результатах выдачи запросов. Сегодня компания Google обрабатывает более 40 миллиардов запросов в месяц, что соответствует 62,4 % всех поисковых запросов в мире.

## **1.2 Задачи поисковых систем.**

Все поисковые системы объединяют несколько основных задач. Это поиск новых сайтов, оценка сайта и максимально точный ответ пользователю на запрос.

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

Это означает, поисковая система должна «думать» так же, как думает пользователь при поиске информации. Когда пользователь обращается с запросом к поисковой машине, он хочет найти то, что ему нужно, максимально быстро и просто. Получая результат, он оценивает работу системы, руководствуясь несколькими основными параметрами. Для того, чтобы удовлетворить пользователя, разработчики поисковых машин постоянно совершенствуют алгоритмы и принципы поиска, добавляют новые функции и возможности, всячески пытаются ускорить работу системы.

## **1.3 Состав и принцип работы поисковой системы.**

Поисковая машина – это аппаратно-программный комплекс, осуществляющий быстрый поиск необходимой информации внутри сервера или Интернет-ресурса. Основа поисковой машины у всех поисковых систем примерно одинаковая. Как правило, это поисковый бот, необходимый для индексации и поиска сайта, программное обеспечение, отвечающее за составление каталога запроса и ранжирование результатов по релевантности поискового запроса. Но многие крупные поисковые системы держат в секрете содержание своей поисковой машины. Ключевым отличием является база проиндексированных сайтов, релевантность и учет морфологии языка запроса. Все это в совокупности и определяет критерий качества работы поисковых машин.

Классифицируется поисковая машина по области поиска информации:

1. Локальный поиск. Предназначен для осуществления поиска информации по какой-либо части всемирной сети, например, по одному или нескольким сайтам, либо по локальной сети. Примером служит поисковый скрипт на сайте или внутренние серверы крупных компаний.

2. Глобальный поиск. Предназначен для поиска информации по сети Интернет, либо по региональной части, группе сайтов и т.д. Глобальный поиск используют крупные поисковые системы Яндекс, Google, Yahoo и т.д.

Поисковые машины осуществляют различный поиск информации по сети Интернет. Например, картинки, музыка, географическое положение, личная информация и т.д. Файлы, с которыми работает поисковая машина, могут быть разных форматов (например .html, .htm, .txt, .doc, .rtf, ...), графического (.gif, .png, .svg,) или мультимедийного (видео, звука и другой информации). Но наиболее распространенным является поиск по текстовым документам (web-страницы, документы в формате doc, rtf, txt и др.). Поиск по изображениям, видео, звукам более сложен с технологической точки зрения, поэтому массово не реализован. Такие системы, как, например, Яндекс.Картинки искали не по самим изображениям, а по альтернативным текстам, соответствующим этим изображениям. А каталог поиска картинок в компании Google составляется вручную, что увеличивает релевантность запроса, но тормозит обновление баз изображений.

Модуль индексирования:

Модуль индексирования состоит из трех вспомогательных программ (роботов):

Spider (паук) – программа, предназначенная для скачивания веб-страниц. «Паук» обеспечивает скачивание страницы и извлекает все внутренние ссылки с этой страницы. Скачивается html-код каждой страницы. Для скачивания страниц роботы используют протоколы HTTP. Работает «паук» следующим образом. Робот на сервер передает запрос “get/path/document” и некоторые другие команды HTTP-запроса. В ответ робот получает текстовый поток, содержащий служебную информацию и непосредственно сам документ.

Ссылки извлекаются из тэгов a, area, base, frame, frameset, и др. Наряду со ссылками, многими роботами обрабатываются редиректы (перенаправления).

Каждая скачанная страница сохраняется в следующем формате:

- URL страницы
- дата, когда страница была скачана
- http-заголовок ответа сервера
- тело страницы (html-код)

Crawler («путешествующий» паук) – программа, которая автоматически проходит по всем ссылкам, найденным на странице. Выделяет все ссылки, присутствующие на странице. Его задача - определить, куда дальше должен идти паук, основываясь на ссылках или исходя из заранее заданного списка адресов. Crawler, следуя по найденным ссылкам, осуществляет поиск новых документов, еще неизвестных поисковой системе.

Indexer (робот - индексатор) - программа, которая анализирует веб-страницы, скаченные пауками. Индексатор разбирает страницу на составные части и анализирует их, применяя собственные лексические и морфологические алгоритмы. Анализу подвергаются различные элементы страницы, такие как текст, заголовки, ссылки структурные и стилевые особенности, специальные служебные html-теги и т.д.

Таким образом, модуль индексирования позволяет обходить по ссылкам заданное множество ресурсов, скачивать встречающиеся страницы, извлекать ссылки на новые страницы из получаемых документов и производить полный анализ этих документов.

База данных:

База данных или индекс поисковой системы - это система хранения данных, информационный массив, в котором хранятся специальным образом преобразованные параметры всех скачанных и обработанных модулем индексирования документов.

Поисковый сервер:

Поисковый сервер является важнейшим элементом всей системы, так как от алгоритмов, которые лежат в основе его функционирования, напрямую зависит качество и скорость поиска.

Поисковый сервер работает следующим образом:

- Полученный от пользователя запрос подвергается морфологическому анализу. Генерируется информационное окружение каждого документа, содержащегося в базе (которое и будет впоследствии отображено в виде сниппета, то есть соответствующей запросу текстовой информации на странице выдачи результатов поиска).
- Полученные данные передаются в качестве входных параметров специальному модулю ранжирования. Происходит обработка данных по всем документам, в результате чего, для каждого документа рассчитывается собственный рейтинг, характеризующий релевантность запроса, введенного пользователем, и различных составляющих этого документа, хранящихся в индексе поисковой системы.
- В зависимости от выбора пользователя этот рейтинг может быть скорректирован дополнительными условиями (например, так называемый «расширенный поиск»).
- Далее генерируется сниппет, то есть, для каждого найденного документа из таблицы документов извлекаются заголовки, краткая аннотация, наиболее соответствующая запросу и ссылка на сам документ, причем найденные слова подсвечиваются.
- Полученные результаты поиска передаются пользователю в виде SERP (Search Engine Result Page) – страницы выдачи поисковых результатов.

Как видно, все эти компоненты тесно связаны друг с другом и работают во взаимодействии, образуя четкий, достаточно сложный механизм работы поисковой системы, требующий огромных затрат ресурсов.

# ГЛАВА 2.ПОИСКОВЫЕ СИСТЕМЫ

## 2.1 Классификация поисковых систем.

Наиболее известные международные поисковые системы: Google, Yahoo, Baidu(лидер среди китайских поисковых систем), Bing, AOL, Lycos, Ask.

Русскоязычные — большинство «русскоязычных» поисковых систем индексируют и ищут тексты на многих языках — украинском, белорусском, английском, татарском и др. Отличаются же они от «всеязычных» систем, индексирующих все документы подряд, тем, что в основном индексируют ресурсы, расположенные в доменных зонах, где доминирует русский язык или другими способами ограничивают своих роботов русскоязычными сайтами.

В России основной поисковой системой является «Яндекс», далее - Rambler, Aport, Mail.ru, Нигма.

А теперь более подробно об этих поисковых системах.

## 2.2 Поисковые системы.

### Google.

Google — поисковая система, принадлежащая корпорации Google Inc.

Первая по популярности (84,65 %), обрабатывает 41 млрд 345 млн запросов в месяц (доля рынка 62,4 %), индексирует более 8 миллиардов веб-страниц, может находить информацию на 191 языке (с 15 октября 2009).

Поддерживает поиск в документах форматов PDF, RTF, PostScript, Microsoft Word, Microsoft Excel, Microsoft PowerPoint и других.

В качестве учебного проекта двое студентов Стендфорского университета, Лари

Пейдж и Сергей Брин, предложили новую поисковую систему, которая на сегодняшний день стала одной из самых влиятельных компания во всемирной сети Интернет – Google.

Своё название система и соответствующая компания получили от математической величины "гугол" (googol), которая равна 10<sup>100</sup>. Создатели посчитали, что такое название лучше всего передаст основную идею их системы – организовать миллиарды байтов информации, содержащейся в Интернете.

Поисковая система Google представляет собой мощный механизм. Без таких поисковых систем найти информацию в глобальной сети Интернет было бы практически невозможно. Подобно всем поисковым серверам Google использует специальный поисковый алгоритм для получения результатов поиска. Некоторые основные характеристики алгоритма компания не скрывает, особенность же алгоритма является строгой тайной компании. Именно благодаря этому Google сохраняет свое лидерство в сети Интернет и не позволяет никому взломать ее систему.

Как и большинство поисковых серверов, Google использует программный поисковый механизм, называемый «паук» или «ползунок», для автоматического выбора всех документов, на которые есть ссылки в первом выбранном документе. Ключевые слова вписываются в специальную строку ввода, после чего начинается поиск. Особенность Google заключается в том, как и по какому критерию эта поисковая система классифицирует результаты поиска на своей странице. Используемый алгоритм PageRank сортирует все веб-страницы по критерию смыслового соответствия.

Алгоритм PageRank зависит от нескольких факторов:

- От частоты повторения и местоположения ключевых слов на веб-странице – если искомое слово или фраза только один раз встречается на сайте, то страница получает низкий бал.
- От времени существования страницы – каждый день в Интернете появляются новые сайты, но не многие задерживаются надолго. Поэтому преимущество отдается уже зарекомендовавшим себя сайтам, существующим долгое время.



- От количества веб-страниц, связанных с «главной страницей» - Google «смотрит» на количество веб-страниц, относящихся к определенному сайту и определяет ее рейтинг среди всех остальных.

Систему Google практически невозможно обхитрить, потому что все ссылки на веб-страницы она воспринимает как «голоса». Самый лучший способ обеспечить своему сайту первые места, это напичкать его самой разнообразной информацией, которая будет привлекать больше людей. Чем больше ссылок на веб-странице, тем выше будет оценка поискового агента PageRank.

Для некоторых результатов поиска Google ранее предоставлял повторное поле для поиска, которое позволяет пользователю найти то, что он ищет внутри конкретного веб-сайта. С 2011 года эта опция недоступна, что весьма отрицательно сказалось на удобстве, популярности, и востребованности Google.

«Википоиск» – поисковая технология, позволяющая пользователю настраивать результаты выдачи по поисковым запросам. Пользователь может удалять результаты из списка и поднимать вверх списка. Технология была запущена компанией «Google» весной 2009 года и проработала до осени.

Google предоставляет не только возможность эффективного поиска, но и ряд бесплатных сервисов, большинство из которых не требует установки дополнительного ПО на компьютер пользователя. Наиболее популярными веб-приложениями от Google являются «Gmail» и «Gtalk». Оба проекта прекрасно дополняют друг друга, но могут работать и по-отдельности. «Gmail» представляет собой почтовый сервис с автоматической фильтрацией спама, мобильным доступом и большим объемом почтового ящика (на апрель 2009 г. он составляет более 7 Гб, и это значение неуклонно растет). «Gtalk» – сервис, позволяющий обмениваться голосовыми и текстовыми сообщениями, как в окне браузера, так и с помощью специального ПО.

Для владельцев популярных сайтов Google предлагает несколько программ контекстной рекламы. Сервисы «AdSense» и «AdWords» позволяют заработать на посещаемости своих Интернет-страниц и привлечь новых посетителей.

Обычным пользователям «Google» предлагает доступ к справочной информации,

новостям, обмену фотографиями, и многим другим ресурсам.

Yahoo!



Yahoo! (произносится как яху!) — американская компания, владеющая второй по популярности (6.35 % ) в мире поисковой системой и предоставляющая ряд сервисов, объединённых интернет-порталом «Yahoo! Directory»; портал включает в себя популярный сервис электронной почты «Yahoo! Mail», один из старейших и наиболее популярных в Интернете. Не так давно была запущена новая версия почтового интерфейса, основанная на AJAX (русскоязычный обзор нового интерфейса).

В январе 1994 года аспиранты Стэнфордского университета Дэвид Файло и Джерри Янг создали веб-сайт, который назывался «Путеводитель Джерри по Всемирной Паутине». «Путеводитель» представлял собой каталог других сайтов.

В апреле 1994 года сайт был переименован в Yahoo!. Существует две версии происхождения названия. Согласно первой, слово было взято из книги Джонатана Свифта «Путешествия Гулливера» (в русском переводе Йеху, еху), где обозначает расу грубых и тупых человекообразных существ. Именно на этой версии настаивают основатели компании. Согласно второй, Yahoo! — акроним, образованный от фразы «Еще один иерархический неотесанный (неофициальный) прорицатель» (англ. Yet Another Hierarchical Officious Oracle). Но есть и третья версия происхождения названия. Вполне возможно, Yahoo произошло от японского Yahooo, что в переводе означает «Привет» (неформальный вариант). Однако к тому времени Yahoo уже был зарегистрированной торговой маркой соуса для барбекю, поэтому к названию был добавлен восклицательный знак. Янг и Файло быстро оценили коммерческий потенциал проекта и 2 марта 1995 года основали корпорацию Yahoo!

Согласно статистике Alexa Internet, на сегодняшний день Yahoo! — четвертый по посещаемости веб-сайт в сети Интернет, и примерно 28% посещений состоят из просмотра только одной страницы.

Bing.



Bing — поисковая система, разработанная международной корпорацией Microsoft. Доступна по адресу <http://www.bing.com/>. Ранее имела следующие наименования и адреса:

- MSN Search (<http://search.msn.com/>) — с момента появления и до 11 сентября 2006;
- Windows Live Search (<http://search.live.com/>) — до 21 марта 2007;
- Live Search (<http://www.live.com/>) — до 1 июня 2009.

В настоящее время Bing занимает третье место по счёту самых используемых поисковых систем, в отличие от которых, обладает рядом эксклюзивных возможностей, таких как просмотр результатов поиска на одной странице (вместо пролистывания многочисленных страниц результатов поиска), а также динамическое корректирование объёма информации отображаемой для каждого результата поиска (например, только название, краткая или большая сводка).

В американской версии Bing присутствуют некоторые интересные нововведения, касающиеся интернет-поиска, например:

- 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.

При том, что Bing неплохо ищет по англоязычным ресурсам, для российских пользователей он фактически бесполезен. «Яндекс» и Google дают гораздо более релевантные результаты.

Яндекс.



Поисковая система "Яндекс" является восьмой среди крупнейших поисковых сайтов мира по количеству обработанных поисковых запросов (1,892 млрд., статистика за декабрь 2009 г.) и первым крупнейшим неанглоязычным поисковым сервером. Ежедневная посещаемость этой поисковой системы – более 9 миллионов человек в сутки (январь 2010). В поисковой системе представлено более 5,4 миллиона сайтов с количеством страниц более 2,5 миллиарда, что составляет более 62 терабайта информации.

Поисковая система Yandex.ru была официально анонсирована 23 сентября 1997 года, и первое время развивалась в рамках компании CompTek International. Как отдельная компания "Яндекс" образовался в 2000 году.

"Яндекс" постоянно совершенствует свои поисковые алгоритмы, что позволяет ему соответствовать современным критериям поиска и становиться на уровень Google хотя бы в России. А пока так и есть, если судить по уровню освоения обеих поисковых систем аудиторией СНГ. "Яндекс", как и Google, работает на кластерной системе организации компьютерных вычислительных сетей. Каждый кластер отвечает за определённый сегмент сохранённой информации.

Сканирующие роботы поисковой системы бывают 2 видов:

1) основной сканирующий робот

2) быстрый робот, который регулярно сканирует сайты, на которых информация обновляется очень часто, и вносит информацию с этих сайтов в поисковую систему, что обеспечивает быстрое обновление индекса поисковой системы.

## **Обновления (апдейты) поисковой системы бывают 2 видов:**

1) Обновления поисковой базы. Происходят несколько раз в месяц. В поиске начинают выдаваться обновлённые страницы сайтов, собранные основным сканирующим роботом.

2) Обновления программной части (движка) поисковой системы. Сутью данных обновлений является изменение алгоритмов ранжирования документов в поисковой системе. Обычно подобные обновления анонсируются и им даются определённые названия.

"Яндекс" учитывает морфологию русского языка (мощная система определения словоформ), имеет возможность сузить запрос до максимально точного при помощи специальных поисковых формул и геотаргетинга, имеет качественно разработанный алгоритм оценки релевантности документов (точность отношения текста документа к поисковому запросу), обладает высокой скоростью реакции на запросы и практически полным отсутствием перегрузок серверов.

Совсем недавно (с 2009 года) с вводом нового алгоритма "Снежинск" поисковая система "Яндекс" научилась определять региональную принадлежность сайтов и формировать результаты поиска по географическому расположению пользователя. Сегодня это самый точный географический поисковик Рунета.

Вместе с тем, интернет-портал «Яндекс» представляет собой не только мощную поисковую систему. Под этим брендом собрано множество самых разнообразных сервисов.

Так, с помощью «Яндекса» можно узнать свежие новости («Новости»), пообщаться с друзьями и коллегами («Блоги», «Почта»), заработать («Мой Круг», «Директ», «Рекламная сеть»), продать или приобрести различные товары («Маркет», «Авто»), а также получить море полезной информации (пробки, погода, котировки валют, телепрограмма, афиша). Не менее популярна у русскоязычных пользователей и платежная система «Яндекс Деньги», которая позволяет осуществлять электронные платежи с помощью веб-интерфейса или Интернет-кошелька.

Рамблер.



Rambler – поисковая система, разработанная в 1996 году. До недавнего времени Rambler был первым по популярности поисковиком Рунета. И хотя сейчас виртуальная пальма первенства «перешла» к Яндекс и Google, Rambler прочно обосновался на ведущих позициях. По оценкам SpyLog на Rambler приходится 20-25% от всех поисковых запросов Рунета.

Название поисковика как нельзя лучше отражает суть деятельности компании («rambler» - по-англ. «бродяга»). Rambler позволяет искать документы на английском, русском, украинском и многих других языках, а при формировании списка к выдаче учитываются не только ключевые запросы, но и их разнообразные словоформы. Полученные в результате поиска данные структурируются по степени соответствия запросу (релевантности) и распределяются по сайтам.

Сравнительно недавно (в начале 2009 г.) разработчики Rambler'a начали внедрение вертикального поиска, в основе которого лежит технология XAG (eXtended AGgregator). Благодаря этой системе, запрашиваемая информация сортируется по темам, а работать с поисковой машиной становится несколько проще.

Использование вертикалей поисковиком имеет и еще одно преимущество — при недостатке информации в одном документе, она может дополняться данными из другого. Что очень удобно для «очистки» данных от повторов, спама и так далее. Например, при поиске вакансий по телефонному номеру той или иной компании система определяет ее название. Таким образом, поисковая система Rambler может отслеживать на соответствующих сайтах дублирующиеся объявления, сомнительные вакансии и т.д.

Особенностью поисковой системы является то, что Rambler обслуживает только сайты, находящиеся в следующих доменах первого уровня: Российская Федерация: .ru, .su; Украина: .ua; Белоруссия: .by; Казахстан: .kz; Киргизия: .kg; Узбекистан: .uz; Грузия: .ge.

Вместе с тем, Rambler — это не только поисковая машина. Под эгидой Rambler организовано множество Интернет-проектов, с помощью которых можно узнать новости, посетить наиболее популярные сайты, посмотреть видео, послушать музыку, получить полезную информацию, а также развлечься и познакомиться с интересными людьми. Помимо таких традиционных и знакомых всем функций, как Топ 100, почта, новости и игры, Rambler предлагает и другие сервисы.



Нигма.РФ – российская интеллектуальная

метапоисковая система, первая кластеризующая поисковая система в Рунете. Проект создан при поддержке факультетов ВМиК и психологии МГУ, а также Стэнфордского университета. Название — Nigma (один из родов пауков семейства Dictynidae, en:Nigma) было выбрано по ассоциации со Всемирной паутиной.

Nigma осуществляет поиск как по своему индексу, так и по индексам Google, Yahoo, Bing, Яндекс, Rambler, AltaVista, Aport. По состоянию на 28 февраля 2009 в суммарном индексе всех этих поисковых систем находится более 7 160 000 000 русскоязычных документов.

На основе введённого пользовательского запроса Нигма формирует список документов, разделённых на несколько множеств (кластеров). Пользователь может уточнить в каком множестве продолжить поиск, тем самым улучшив релевантность результатов поиска. Пользователь также может исключить ненужные ему множества сайтов, например, документы, пришедшие с интернет-магазинов (для них формируется специальный кластер). Список кластеров выводится слева от списка результатов поиска. Для каждого кластера указывается образующая его фраза и количество документов в кластере. Пользователь может управлять кластерами при помощи специальных ссылок под списком кластеров.

- Нигма поддерживает русскую морфологию.
- Используется морфологический модуль для русского языка собственной разработки.
- Нигма также может выводить ход решения многих математических задач.

- Система позволяет производить поиск по более чем 12000 неорганических реакций.
- В ответ на запрос пользователи увидят три самые свежие новости прямо на странице результатов поиска. Разработчики проиндексировали и обработали данные более 3500 rss-лент СМИ и популярных блогов. На сайте также есть форма, через которую пользователи сами могут добавлять новостные ресурсы в индекс поисковой системы. База данных новостей обновляется каждые 5 минут.
- Поддерживается орфография на базе словаря, составленного из реальной русской орфографии, которую авторы веб-сайтов используют на своих страницах.
- Система предлагает варианты на основе предыдущих запросов пользователей Нигма.РФ.
- В улучшенной версии поисковик отвечает на вопросы пользователя до того, как он полностью введёт свой вопрос в строку поиска. Система предлагает варианты заполнения, учитывая популярности предыдущих запросов пользователей. Справа от варианта выдаётся ответ на вопрос.
- При выборе нужного варианта в поисковой подсказке пользователь может увидеть определения понятий и слов, которые хочет найти. База данных содержит 340928 определения. Информация берется из Википедии. Предполагается подключение и других источников информации.

Mail.ru и GoGo.ru





Mail.ru — крупный

коммуникационный портал российского Интернета. Его ежемесячная аудитория превышает 50 миллионов уникальных посетителей. Ресурс принадлежит р.



В 1998 году программисты, работающие в

петербургском офисе американской софтверной компании DataArt, создали новое программное обеспечение для почтового веб-сервера, которое в дальнейшем предполагалось продавать западным компаниям. Чтобы протестировать сервис, его временно выложили в открытый доступ для российских пользователей, а сервис вдруг стал стремительно набирать популярность.

GoGo.ru - относительно новый и вполне автономный проект от создателей Mail.ru. Несмотря на то, что соответствующий домен был зарегистрирован компанией еще

в 2000-м году, работы над созданием поисковика начались только в 2006. Руководство над новым «детищем» Mail.ru доверила организатору поисковой системы Апорт – Михаилу Костину.

Поисковик был запущен в июне 2007 года, и первое время имел небольшой охват и выдавал довольно странные результаты. Однако работа над поисковой системой продолжалась, и к февралю 2008 г. GoGo.ru осуществляла поиск по более чем 2,5 млрд. документов, 140 млн. изображений, 2 млн. видео-файлов, и т.д.

Успехи новой поисковой системы были отмечены высшей наградой на выставке SofTool-2007. Именно там, в рамках конкурса «Продукт года», в номинации «Интернет-технологии», GoGo.ru получила первое место.

От остальных поисковых систем GoGo.ru отличается возможностью поиска по русскоязычным видеороликам, WAP-сайтам, анализа базы данных Ответы@Mail.ru, а также набор функций для веб-мастеров и владельцев собственных Интернет-проектов.

В настоящее время, основными видео-хостерами, по материалам которых осуществляет поиск GoGo.ru, являются такие ресурсы, как Video.Mail.ru, RuTube.ru, LiveInternet.ru, Teledu.ru, Smotri.com, Myvi.ru, Video.i.ua и некоторые другие. Кроме того, GoGo.ru может искать видео-ролики по размещенным на YouTube или Metacafe файлам с описанием на русском языке.

Следует отметить, что при поиске роликов в выдаче оказываются не только видео, отвечающее запросу, но и изображения. При этом поисковый бот GoGo.ru имеет встроенный фильтр, позволяющий отделять фотографии от всех остальных картинок.

Высокая эффективность графического поиска в GoGo.ru обеспечивается наличием XML-синдикации с ведущими фотохостерами Рунета.

Заслуживает внимания и обработка поисковой системой разнообразных синонимов. Данной проблеме разработчики GoGo.ru уделяли особое внимание, и в настоящее время проект имеет обширный словарь словоформ и синонимии, пополняющийся в полуавтоматическом режиме.

# ЗАКЛЮЧЕНИЕ

Поисковые системы и существующие к ним приложения, способны облегчить работу представителей многих профессий: Web-мастера, аналитика, руководителя, лингвиста. Информационный бум продолжается, происходит дальнейшее развитие электронно-компьютерных технологий, а следовательно и в будущем без поисковых систем обойтись будет крайне сложно.

Итак, первоочередная задача любой поисковой системы – доставлять людям именно ту информацию, которую они ищут.

Основные характеристики, которыми должны обладать поисковые системы:

**Полнота** – Представляет собой отношение количества найденных по запросу документов к общему числу документов в сети Интернет, удовлетворяющих данному запросу.

**Точность** – Определяется степенью соответствия найденных документов запросу пользователя.

**Актуальность** – Характеризуется временем, проходящим с момента публикации документов в сети Интернет, до занесения их в индексную базу поисковой системы.

**Скорость поиска** – Скорость поиска тесно связана с его устойчивостью к нагрузкам. Посетитель желает получить результаты как можно быстрее, а поисковая машина должна обрабатывать запрос максимально оперативно, чтобы не тормозить вычисление следующих запросов.

**Наглядность** – Наглядность представления результатов является важным компонентом удобного поиска. По большинству запросов поисковая машина находит сотни, а то и тысячи документов. Вследствие нечеткости составления

запросов или неточности поиска, даже первые страницы выдачи не всегда содержат только нужную информацию. Это означает, что пользователю зачастую приходится производить свой собственный поиск внутри найденного списка.

Как показывает статистика, пользователи русскоязычной части Интернета предпочитают несколько поисковых машин. Прежде всего, это мультязычная платформа Google, являющаяся своеобразным эталоном универсального поискового механизма.

Чуть менее популярный поисковик — Yahoo! - объединяет не одну, а целых 3 поисковых машины (Inktomi, AltaVista, Alltheweb).

Лидер среди русскоязычных поисковых систем — Яндекс — индексирует документы форматов pdf, rtf, doc, txt, swf, rss и так далее. С помощью Яндекса можно искать информацию на русском, английском, украинском, белорусском, румынском, немецком и французском языках.

Rambler — оценивает преимущественно посещаемость Интернет-страниц. Соответственно, чтобы оказаться в верхних строчках рейтинга Rambler, необходимо обеспечить своему сайту постоянный приток посетителей.

В настоящее время, практически каждая поисковая система имеет свои механизмы расчета рейтинга Интернет-страниц, и алгоритмы эти постоянно изменяются, совершенствуются. Однако в целом можно сказать, что наибольшее внимание современные поисковики уделяют внешним критериям.

## **Список использованных источников**

- <http://www.getinfo.ru/article108.html>
- <http://www.thepromo.ru/articleview.php?id=60>
- <http://www.wmz-portal.ru/list-c-searchsystems.html>

- <http://re-port.ru/articles/74607/>
- <http://www.homearchive.ru/internet/i0028.html>
- [http://ru.wikipedia.org/wiki/Поисковая\\_система](http://ru.wikipedia.org/wiki/Поисковая_система)
- <http://www.seonews.ru/masterclasses/detail/29814.php>
- <http://ru.wikipedia.org/wiki/Google>
- <http://ru.wikipedia.org/wiki/Bing>
- <http://www.bastau.com/websitepromotion-searchengines.htm>
- <http://www.seonews.ru/masterclasses/detail/29812.php>
- <http://mp.ustu.ru/Users/?????/?/>

## **ПРИЛОЖЕНИЯ**