

Содержание:

ВВЕДЕНИЕ

Актуальность.

Поисковая система – это аппаратно-программный комплекс, который предназначен для осуществления функции поиска в интернете, и реагирующий на пользовательский запрос который обычно задают в виде какой-либо текстовой фразы (или точнее поискового запроса), выдачей ссылочного списка на информационные источники, осуществляющейся по релевантности.[\[1\]](#)

Каждый день большинство пользователей сети Интернет начинают с выхода во Всемирную сеть. Одним из основных веб-серверов является поисковая система, где люди стремятся найти столь нужные им сведения и решить собственные проблемы. К сожалению, поисковые системы нередко не могут точно и справедливо интерпретировать ресурсы. Как результат, на верхних позициях поиска нередко оказываются сайты «далекие» от решаемого вопроса.

В то же время ресурсы, представляющие реальную пользу, оказываются «за бортом» поиска. Причина этого положения проста и кроется в технологии получения и представления полученных результатов поисковыми системами. Здесь следует понимать, что основная проблема выражается в отсутствии каких-либо чётких правил, открытых и доступных для любых желающих.

Чем больше неопределенности в алгоритмах формирования поисковых индексов (особый черный ящик), тем меньше поисковые системы отражают процесс формирования реальной информации. И соответственно, тем меньше окажется уровень доверия к результатам поиска поисковых систем. Как это ни удивительно, но это вина не поисковых систем, так как они должны прятать правила построения поисковых индексов. Это вина самой технологии во время осуществления поиска.

По собственной сути технология поисковых систем ориентируется на пассивного пользователя. Следует зарегистрировать только сайт, затем всё сделает поисковый робот. Он просканирует этот ресурс - страницу за страницей, стараясь провести анализ содержания каждой из них. Трудоёмкость пользователя окажется минимальной, что позволит использовать различные методики по «обману»

поисковых роботов при довольно низких затратах средств и сил.

В этой схеме работы поисковым системам следует поменять правила и алгоритмы индексирования ресурсов и построения поискового индекса. Конечно, основная часть пользователей пользовались, пользуются, и будут пользоваться классическими поисковиками. Это удобно, просто и распространено. Это - как сама привычка пользоваться поисковиками.[\[2\]](#)

Учитывая все вышеизложенное, цель настоящей работы – анализ поисковых систем в сети Интернет.

Цель исследования обуславливает решение следующих задач:

1. изучить историю создания поисковых систем;
2. рассмотреть классификацию поисковых систем в сети Интернет;
3. провести сравнительный анализ поисковых систем.

ГЛАВА 1. ПОИСКОВЫЕ ТЕХНОЛОГИИ В СЕТИ ИНТЕРНЕТ

Поисковая система – это программное оснащение, предоставляющее доступ к набору полуструктурированной информации. Ориентация на слабоструктурированные сведения, то есть данные, которые нельзя представить в форме реляционной таблицы, отличает поисковую систему от системы управления базами данных. В этом определении поисковой системы понимается информация разного рода, то есть аудио, текст, изображения, видео и т.д. Однако следует обратить внимание на то, что как раз текстовые данные идеально подходят для характеристики полной функциональности поисковой системы, так как алгоритмы отбора мультимедийной информации, в первую очередь, строятся на алгоритмах поиска текста.[\[3\]](#)

Поисковые системы при пользовании Интернетом играют очень важную роль. В Интернете сконцентрировано такое количество информации, что ее поиск уже трансформируется в отдельную задачу и занимает очень много времени. Поисковые сервера выдают на запрос несколько тысяч ссылок вместо нескольких страниц, где действительно содержатся нужные сведения. Пользователи всемирной сети Интернет, поняв преимущества, предоставляемые возможностью анализа пространственных данных, нуждаются в особом инструменте,

позволяющем осуществлять удобный и быстрый поиск, а также доступ к разным цифровым снимкам местности и иным пространственным данным, сконцентрированным во многих коммерческих, правительственных и академических организациях.[\[4\]](#)

1.1 История развития поисковых систем

Во времена, когда только начиналось развитие интернета, объём общедоступной информации был относительно мал, и пользователей сети было немного. На первоначальных стадиях развития сети, ею пользовались сотрудники университетов и исследовательских лабораторий для обмена информацией между учреждениями. В то время поиск данных в сети интернет был не актуальным, нежели в нынешнее время.

Первым методом систематизации и организации доступа к информационным ресурсам стало создание разных каталогов сайтов. В них начали группировать ссылки по четко определенной тематике.[\[5\]](#)

Первопроходцем оказался сайт Yahoo, запущенный в апреле 1994 года. После того, как количество сайтов в каталоге Yahoo начало существенно расти, стала доступна возможность выполнения поиска информации по каталогу. Это, естественно, не было поисковой системой в полном смысле, так как сфера поиска ограничивалась лишь средствами, имеющимися в каталоге, а не всеми ресурсами сети Интернет. Реестры ссылок часто применялись и ранее, но почти потеряли свою популярность в настоящее время. Причина этого довольно проста – даже современные каталоги, включающие в себя огромное количество ресурсов, представляют информацию только об очень незначительной части сети Интернет. Наиболее крупный каталог сети Open Directory Project (раньше Directory Mozilla) содержит сведения только о 5 миллионах ресурсов. При этом база поисковой системы Google включает свыше 8 миллиардов документов.

Первой полноценной поисковой системой оказался проект WebCrawler, запущенный в 1994 году. В 1995 году появились поисковые системы Lycos и AltaVista. При этом последняя на протяжении многих лет была лидером в области поиска данных и информации в Интернет. В 1997 году Сергей Брин и Ларри Пейдж создали Google в ходе исследовательского проекта в Стэнфордском университете. Сейчас момент Google – наиболее популярная поисковая система во всём мире. 23 сентября 1997 года была официально анонсирована поисковая система Yandex, наиболее

популярная в русскоязычной части Интернет. Сейчас существует 3 главные международные поисковые системы – Google, Yahoo и MSN Search, имеющие свои базы и алгоритмы поиска. Основная часть остальных поисковых систем (которых можно насчитать большое количество) использует в разном виде результаты 3 перечисленных. К примеру, поиск AOL (search.aol.com) и Mail.ru пользуются базой Google, а AltaVista, Lycos и AllTheWeb – базой Yahoo. В России главной поисковой системой является Яндекс, за ним следуют Rambler, Google.ru, Aport, Mail.ru и KM.ru.

Поисковая система включает в себя из следующие главные компоненты:

1. Spider (паук) – браузероподобная программа, скачивающая веб-страницы. Crawler (краулер, «путешествующий» паук) – программа, автоматически проходящая по всем ссылкам, обнаруженным на странице.
2. Indexer (индексатор) – программа, проводящая анализ веб-страниц, скаченных пауками. Database (база данных) – хранилище скаченных и обработанных страниц. Search engine results engine (система выдачи результатов) – программа, получающая результаты поиска из базы данных.
3. Web server (веб-сервер) – веб-сервер, осуществляющий взаимодействие между пользователем и другими составными частями поисковой системы. Детальная реализация поисковых механизмов может отличаться друг от друга (к примеру, связка Spider + Crawler + Indexer может быть выполнена в форме единой программы, скачивающая известные веб-страницы, после чего проводит анализ их и ищет по ссылкам новые ресурсы), однако всем поисковым системам характерны рассмотренные общие черты.
4. Spider (паук) – это программа, скачивающая веб-страницы тем же методом, что и браузер пользователя. Отличие заключается в том, что браузер отображает сведения, содержащиеся на странице (графические, текстовые и т.п.), а паук не обладает визуальными компонентами и работает напрямую с html-текстом страницы (вы можете сделать «просмотр html-кода» в вашем браузере, чтобы увидеть «сырой» html-текст).
5. Crawler – программа, выделяющая все ссылки, имеющиеся на странице. Её задача – выяснить, куда дальше должен направиться паук, основываясь на ссылках или исходя из заранее заданного перечня адресов. Краулер, следуя по обнаруженным ссылкам, выполняет поиск новых документов, еще неизвестных поисковой системе.
6. Indexer (индексатор) – программа, которая разбирает страницу на составные части и проводит их анализ. Выделяются и анализируются разные элементы страницы, такие как текст, заголовки, стилевые и структурные особенности,

специальные служебные html-теги и т.п.

7. Database (база данных) – это хранилище всех сведений, скачиваемых и анализируемых поисковой системой. Изредка базу данных считают индексом поисковой системы.
8. Search Engine Results Engine является системой выдачи результатов и занимается ранжированием страниц. Она решает, какие именно страницы удовлетворяют пользовательскому запросу, и в какой последовательности они должны быть отсортированы. Это происходит по алгоритмам ранжирования поисковой системы. Такая информация оказывается самой интересной и ценной для нас – как раз с этим элементом поисковой системы взаимодействует оптимизатор, стараясь улучшить позиции сайта в выдаче, поэтому в последующем мы подробно изучим все факторы, оказывающие воздействие на ранжирование результатов.
9. Web server. Обычно на сервере есть html-страница с полем ввода, в котором пользователем задаётся интересующий его поисковый запрос. Ещё веб-сервер отвечает за выдачу полученных результатов пользователю в форме html-страницы.[\[6\]](#)

1.2 Классификация поисковых систем

Все поисковые системы условно можно поделить на три больших класса:

- 1) Поисковые машины (search engines). Иначе их ещё называют программы-пауки и программы-червяки. Эти программы ползают от одного сайта к другому, методично индексируя его контент. Все, что находит подобный червяк, оказывается в базе данных, куда каждый пользователь может обратиться с запросом. Преимущество этих программ – обширная база данных, то есть почти вся сеть. Минус – это то, что в ответ на каждый запрос вы получите тысячи адресов веб-страниц.
- 2) Веб-каталоги или поисковые порталы (directories). Информация в них организована в виде древовидной структуры, чаще всего по тематическому признаку и на основании рейтинга. Адреса и описания веб-сайтов переносятся в каталог по заявке. Записи редактируются вручную модератором, web-мастером. Во многие каталоги попасть сложно, некоторые такую услугу делают платной. Желание хозяев сайтов попасть в каждый каталог (чем больше, тем лучше) обусловлено стремлением разместить свою ссылку на чужом сайте и, соответственно, увеличить рейтинг в метапоисковых системах, о чем будет сказано

ниже. Если Вы стараетесь повысить количество посетителей на собственном сайте, то следует выбрать такой каталог, где Вас будет видно. При этом эффект от размещения ссылки в малом каталоге в случае наличия интересного контента порой оказывается довольно высоким, так как с этих сайтов посетители переходят чаще всего по какой-либо ссылке. Однако регистрация в таких каталогах как Yahoo и Open Directory желательна, так как их информационные базы используются метапоисковыми системами во время определения рейтинга.

3) Метапоисковые системы – это поисковые системы, у которых нет своей базы данных с адресами и характеристикой ресурсов. Они пользуются базами данных каталогов. В своей базе данных имеются лишь адреса ресурсов. Поиск с помощью таких систем сейчас считается самым популярным.

1. Поисковые машины:

AltaVista (вид сверху) – торговая марка популярной поисковой машины. В настоящее время база данных AltaVista считается наиболее крупной в Интернете. Помимо разветвленных средств поиска текстовой информации, она содержит такие инструменты, как Photo Finder – поиск изображений, технологию перевода документов в режиме «онлайн» и возможность индексирования на разных языках. В сотрудничестве с фирмой AskJeeves, AltaVista создала базу данных, управляемую с помощью команд на поддерживаемых ею языках.

Excite – это поисковая машина. Технология Excite лицензирована фирмой Netscape Communications для применения на портале NetCenter, а также корпорацией America Online (с правом собственного дополнения базы данных). Она имеет интеллектуальные алгоритмы поиска с использованием технологии Intelligent Concept Extraction, помогающие работать не только с отдельными ключевыми словами, но и с объединяющими их терминами. Если, например, вы введете фразу «система образования», то поисковая машина просмотрит и страницы, где есть слова «учебник», «школа» и т.п. Итак, ПИ оказывается довольно эффективной при поиске разных материалов по смежным понятиям.

Goto – поисковая машина. Фирма начала собственную деятельность с покупки старой и довольно известной поисковой машины WWW Worm. Затем решили подобрать ссылки на наиболее популярные темы поиска и разместить соответствующие ключевые слова на заглавной странице. Результаты поиска нередко получаются такие же, что и в HotBot, Snap и Yahoo, а порой даже лучше.

HotBot (Wired Digital) - с 1998 года поисковая машина принадлежит фирме Lycos. Здесь применяется оригинальная технология Inktomi, позволяющая осуществлять полноценный текстовый поиск по произвольному ключевому слову. Главными посетителями её поискового сервера являются компьютерщики-профессионалы, применяющие HotBot для поиска нужного программного обеспечения и данных, связанных с разными информационными технологиями. В то же время компьютерной тематикой эта поисковая система, безусловно, не ограничена. Непрерывно ведется работа, направленная на последующее её развитие, то есть происходит пополнение базы данных Inktomi, подготавливаются новые версии алгоритмов и т.п.

InfoSeek – поисковая машина, когда-то входящая в десятку лучших. Сейчас она больше ориентирована на электронную коммерцию. После создания в 1999 году совместного с фирмой Walt Disney нового суперпортала Go Network, сайт входит в первую десятку по посещаемости.

Northern Light – поисковая машина создавалась для повышения информативности поиска в Интернете. Ее главная идея – поиск по контексту. Спайдер фирмы каждый день индексирует тысячи сайтов, в число которых входят электронные издания, службы новостей, периодика, академические библиотеки и электронные архивы текстов. Еще одной чертой поисковой машины Northern Light, выгодно отличающей её от других, является возможность сортирования полученных сведений по адресам сайтов и тематикам. Желающие могут подписаться на все доступные тематические подборки, к примеру, на материалы конкретных рубрик из любимых газет и журналов, а потом на протяжении года получать специально подготовленную и отсортированную информацию. Проект Northern Light считается одним из наиболее масштабных в Интернете.

SearchKing – поисковая система с упором на увеличение достоверности информации (релевантности документов). Во время построения рейтинга поисковая машина учитывает число «кликов» (щелчков) на ссылках, посещаемых в результате поиска. В связи с этим каждый пользователь невольно «голосует» за наиболее популярный сайт.

WebCrawler – с 1996 года проект принадлежит фирме Excite, поэтому на его заглавной странице находится логотип этой компании, а поисковая машина WebCrawler является составной частью Excite Network. Ее авторы обращают внимание на то, что многие более поздние технологии, включая Lycos и InfoSeek, были разработаны уже после появления WebCrawler. База данных проекта

продолжает пополняться на регулярной основе, но использовать эту систему рекомендуется в ситуациях, когда необходим поиск по одному или двум ключевым словам.

2. Веб-каталоги и поисковые порталы:

LookSmart - Каталог ссылок, редактируемый вручную, усилен одной из наиболее мощных поисковых машин AltaVista - это помогло создать один из наиболее информативных порталов. Гигантская база данных AltaVista поможет отыскать необходимый сайт по заданным ключевым словам, а рубрики LookSmart помогут точнее определиться с предметом поиска.

Lycos - специализацией поисковой системы является сфера Интернет-торговли. База данных на регулярной основе пополняется. Этот портал имеет наиболее мощный специализированный каталог WhoWhere, содержащий персональные сведения о зарегистрированных пользователях Интернета, а также развитые средства по оказанию услуг электронной почты с заполнением адресных книг и возможностью формирования иерархической структуры подкаталогов для хранения личной корреспонденции. Всем посетителям в виде ответов на запрос сначала предлагаются сайты Open Directory Project, а потом ссылки из базы данных поискового механизма Lycos.

PlanetSearch Networks - портала нового типа. Он основан на онлайн-обществах, в которые люди могли бы объединиться по увлечениям и интересам, а потом пользоваться не только механизмом поиска, но и чатами и тематическими досками объявлений. В итоге получается отличный набор тематических сайтов и соответствующих поисковых систем.

Yahoo - один из наиболее известных порталов Сети. В каталоге есть ссылки, которые максимально полно отвечают выбранной в запросе тематике. Есть интеллектуальные средства «отсечения» пустых, находящихся в разработке или исключительно рекламных сайтов, далеких от выбранной тематики. Во время поиска на Yahoo! особое внимание уделено предварительному предложению отсортированной информации в каталогах, и только если обнаруженная там информация не удовлетворит пользователя, то запрос будет передан метапоисковым машинам.

3. Метапоисковые системы:

All4One (все-в-одном) - предоставляет посетителям возможность получения результатов поиска прямо из поддерживаемых поисковых систем, то есть с привычным интерфейсом. После ввода пользовательского запроса окно браузера делится на несколько фреймов. В каждом из них предлагается перечень ссылок, найденных определённой поисковой машиной, в число которых входят и AltaVista, и Lycos, и Excite. Следует отметить, что подобный интерфейс обладает определёнными недостатками: страницы со ссылками в узких фреймах довольно неудобно просматривать. В то же время тем, кому необходим быстрый и широкомасштабный поиск с применением известных поисковых машин, All4One предоставит весь спектр нужных услуг.

Debriefing - эта метапоисковая система имеет две разные версии пользовательского интерфейса: интернациональный - на английском языке и национальный - на французском. Во время работы с ключевыми словами на английском языке сервер пользуется стандартным набором популярных поисковых систем, а вот запросы на французском языке обслуживают пять поисковых систем и каталогов Франции.

Dogpile - Мощная метапоисковая система Dogpile использует для поиска не только поисковые машины, но и FTP-серверы, а также сайты, на которых собираются последние новости, фондовые котировки и «желтые страницы» Интернета. Среди дополнительных услуг, предоставляемых этим сервером, возможность получения детальных прогнозов погоды и географических карт интересующей пользователя местности.

Google - одна из наиболее популярных метапоисковых машин в сети. Корректное отображение запрашиваемой информации. Представление информации по рейтингу, в основу которого положен индекс цитируемости страниц. Рейтинг страницы определяется по числу ссылок на нее с популярных внешних сайтов и по упоминанию этого адреса в авторитетных источниках информации.

Mamma Systems - Канадская метапоисковая система предоставляет стандартный комплекс услуг, принятый в таких случаях: используя базы данных популярных поисковых машин, она делает собственную работу очень качественно.

MetaCrawler - метапоисковая система принадлежит американской фирме Go2Net. MetaCrawler сначала выполняет поиск нужных сведений по базам данных иных систем, а потом, пользуясь своим алгоритмом, анализирует и сортирует полученные ссылки, ищет похожие, определяет рейтинг и выдает результат

клиенту. Среди других услуг, оказываемых MetaCrawler, необходимо отметить возможность расширенного поиска, довольно интересное приложение MiniCrawler для поиска информации в Сети в обход сайта фирмы и программу MetaSpy, помогающую вести наблюдение за ключевыми словами в системе MetaCrawler. По данным разных информационных агентств, MetaCrawler в последние годы входит в десятку лучших метапоисковых систем в Сети.

OneSeek - метапоисковая система предназначена для тех, кто знает, что конкретно ищет.

Во время использования OneSeek требуется предварительно выбрать нужную категорию, так как поиск интересующей информации станет производиться именно по этой тематике.

Для того чтобы определиться с целями, предлагается весьма удобный интерфейс, позволяющий оптимизировать поиск. Он экономит время и поможет избежать лишних результатов.

ProFusion - метапоисковую систему ProFusion отличает от других аналогичных систем наличие функции автоматического выбора трех самых подходящих для этого запроса поисковых машин.

Это означает, что после ввода ключевого слова ProFusion старается сузить область поиска, определить тематику, к которой относится запрос, и выбрать три оптимальные для этого случая поисковые машины.

Также ProFusion предоставляет персональный сервис по сохранению ключевых слов запроса, а затем регулярно автоматически сканирует выбранные поисковые машины, а если находит новые сведения по интересующей теме, то сообщает об этом пользователю (или создает для него «теневую» базу данных).

Разрабатываются и иные вспомогательные функции системы.

Proteus - представляет собой скорее даже не метапоисковую систему, а средство переадресации запросов на иные поисковые машины.

Здесь отсутствует алгоритм сортировки полученных ссылок - просто на одной странице находится окошко для ввода ключевых слов и большое количество кнопок для поиска во внешних поисковых системах.

SavvySearch - система оказывает услуги расширенного поиска в Сети с 1998 года.

При запросе клиента исследуются 200 внешних баз данных, сборники прайс-листов и целая совокупность специальных электронных библиотек и справочников. SavvySearch имеет простой и продуманный интерфейс, удобные функции выполнения поиска. Также система предлагает много дополнительных услуг, количество которых непрерывно увеличивается.[\[7\]](#)

ГЛАВА 2. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОИСКОВЫХ СИСТЕМ

Сегодня самым мощным и оперативным источником информации является Интернет. Собственный сайт в сети есть почти у каждой крупной фирмы, организации или компании. В сети Интернет есть электронные варианты многих журналов и газет, через Интернет вещают сотни телекомпаний и радиостанций. В современном обществе почти нет области человеческой деятельности, которая не была бы представлена в сети. Умение быстро находить необходимую информацию сегодня так же необходимо как умение читать и писать. Одной из первых проблем, с которыми сталкивается пользователь, подключаясь к сети, является проблема выбора поисковой системы. Каталоги (общие и специальные), поисковые системы, каталоги поисковых систем, тематические коллекции ссылок, рейтинги и т.п., могут стать помощником для эффективного поиска информации, а могут помочь заблудиться в дебрях Интернета. Определённые рекомендации по выбору поискового указателя довольно быстро стареют. Ситуация в Интернете изменяется буквально на глазах. Не проходит и полугодя, чтобы что-либо не поменялось и в поисковых системах. Та система, которая была самой лучшей вчера, может оказаться не наилучшей сегодня и очень плохой завтра.

1. Яндекс

«Яндекс» — российская ИТ-фирма, владеющая одноимённой системой поиска в Сети и интернет-порталом. Поисковая система «Яндекс» считается 4-ой среди поисковых сайтов мира по числу обработанных поисковых запросов (больше 6,3 млрд. в месяц, 0,5% от мирового количества, статистика за июнь 2019 года). Поисковиком в мире пользуются свыше 50 млн. человек. По состоянию на 14 июня 2019 года, в соответствии с рейтингом Alexa.com, по популярности сайт yandex.ru находится на 20-й позиции в мире и 2-й позиции в России. Поисковую систему Yandex.ru официально анонсировали 23 сентября 1997 года, и сначала развивали в рамках фирмы CompTek International. Как отдельная фирма «Яндекс» появилась в

2000 году. В мае 2011 года Яндекс осуществил первичное размещение своих акций, заработав в результате больше, чем какая-нибудь из Интернет-фирм со времён IPO поисковика Google в 2004 году.

Главным и приоритетным направлением фирмы считается разработка поискового механизма, но за годы работы «Яндекс» стал мультипорталом. В 2019 году «Яндекс» предлагает свыше 50 сервисов. Наиболее популярными являются: Яндекс.Картинки, Яндекс.Почта, Яндекс.Карты, Яндекс.Новости, Яндекс.Погода и другие.

Язык поисковых запросов

Характерная особенность Яндекса — возможность точной настройки поискового запроса. Это реализовано за счёт гибкого языка запросов. Так, к примеру, для операции исключения можно указать область действия: запрос $A \sim \sim B$ найдёт документы (страницы), в которых присутствует A , но не присутствует B , а запрос $A \sim B$ — документы, где слово B не присутствует со словом A в одном предложении. По аналогии, оператор $\&$ ищет сочетания ключевых слов в предложении, а $\&\&$ — во всём документе.

Оператор «!» помогает отключить морфологию для определённого слова, а «!!» позволяет указать нормальную форму, что позволяет обойти некоторые проблемы, связанные с омонимией. К примеру запрос !!Петров будет находить Петрова и Петровых, но не Петра.

Результаты поиска.

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

Если по запросу найдено очень много ссылок, страница результатов предлагает ограничить диапазон поиска — по региону (то есть по диапазону IP) или по дате. Если по какому-нибудь слову или словам ничего не найдено, предлагается поменять его/их на похожие (поскольку предлагаемые варианты зависят от частоты нахождения похожих слов, порой возникают забавные ситуации). Также, предлагается исправить слова, набранные не в той раскладке клавиатуры.

Качество поиска.

Время от времени алгоритмы Яндекса, отвечающие за релевантность выдачи, изменяются, что ведёт к изменениям в результатах поисковых запросов. Такие изменения, официально объявленные, происходили, к примеру, в марте 2004 года, августе 2005 года и январе 2007 года; по неофициальным сведениям, их существенно больше (например, в августе-сентябре 2007 года). Крупное изменение произошло в ноябре 2009 года, когда была выложена обновлённая версия поисковой программы «Снежинск». Последнее такое изменение произошло в декабре 2010-го, когда Яндекс добавил новую поисковую технологию «Спектр» (версия «Краснодар»). Она позволяет учитывать потребности пользователей, которые не были явно сформулированы в запросе. К примеру, по запросу [Бетховен] пользователям покажут результаты и про биографию композитора, и его произведения, и фильм «Бетховен».

В частности, указанные изменения направлены против поискового спама, приводящего к нерелевантным результатам по определённым запросам (реже — по целым семействам запросов).

Сайты, которые «Яндекс» не индексирует или ограничивает ранжирование:

1. Копирующие или переписывающие сведения с иных ресурсов и не создающие оригинального контента.
2. Единственной целью которых оказывается перенаправление пользователя на иной ресурс, автоматически (редирект) или добровольно.
3. С автоматически сгенерированным (бессмысленным) текстом.
4. С каталогами (программ, статей, организаций и т.д.), если они выступают лишь агрегаторами контента, не создают тексты и описания самостоятельно и не предлагают никакого уникального сервиса.
5. С невидимым или слабовидимым текстом или ссылками.
6. Отдающие различный контент пользователям и роботам поисковых систем (клоакинг).
7. Предлагающие товары или информацию по партнёрским программам, но не представляющие какой-либо ценности для пользователя.
8. Пользующиеся обманными техниками (к примеру, вредоносный код, скрипты, настройки серверов), перенаправляющие пользователей на сторонние ресурсы либо меняющие окно результатов поиска на страницы иных ресурсов во время перехода из поисковых систем.

9. Содержащие перечни поисковых запросов (многократное повторение и перечисление ключевых слов), используемые только для обмана поисковой системы и манипулирования итогами её работы, в том числе использование компонентов страниц, скрывающих ключевые слова, к примеру, с помощью скроллинга или иных технических приёмов.
10. Группы сайтов одного владельца/компании, предоставляющие пользователю одни и те же товары или услуги, созданные для заполнения нескольких позиций в результатах поиска и сбора трафика.
11. Немодерируемые форумы, доски объявлений, содержащие большое количество ссылочного спама.
12. Ставящие внешние ссылки только для обмана поисковых систем и «накачивания» релевантности и не оказывающиеся рекомендацией автора посетить ресурс.
13. Сайты или группы сайтов, интенсивно ссылающиеся друг на друга (линкфармы).
14. Страницы сайта с результатами поиска.

Образовательная деятельность Яндекса.

В 2004—2005 и 2006—2007 годах фирма финансировала гранты на исследования молодых учёных по тематике информационного поиска в форме конкурса «Интернет-математика». На конкурс 2006—2007 года поступило 156 заявок, из которых 31 отобранному заявкам было выделено финансирование на общую сумму 5 млн. рублей.

Конкурс «Интернет-математика» проводился в третий раз в 2009 году, но в значительно изменённом формате и с урезанным финансированием. Все участники решали одну общую задачу, поставленную Яндексом. По выданным наборам данных, состоящих из оценок релевантности некоторых документов поисковым запросам, требовалось подготовить ранжирующую формулу методами машинного обучения ранжированию.

В четвёртый раз конкурс «Интернет-математика» проводился в 2010 году. К 16 мая предлагалось предсказать замеры скорости на дорогах Москвы в период с 18 до 22 часов, исходя из данных замеров с 16 до 18 часов и из статистики 30 предшествующих дней.

С октября 2009 года Яндекс вместе с Microsoft Research проводит в своём московском офисе цикл научных семинаров по информационному поиску и анализу

данных для всех желающих.

1 апреля 2010 года состоялся первый, так называемый, Студень — студенческий день Яндекса, организованная для студентов однодневная конференция, на которой можно было выяснить о поисковых технологиях и компании.

24 сентября 2010 года состоялся второй Студень, он прошел в Новосибирске и собрал студентов из Барнаула, Омска, Новосибирска и Томска.

Школа анализа данных.

В сентябре 2007 году была открыта Школа анализа данных Яндекса — двухгодичные очные вечерние курсы для подготовки специалистов в прикладных областях, связанных с обработкой крупных массивов данных (в частности, полученных из интернета).

С 2008 года работу школы поделили на два отделения — отделение анализа данных и отделение computer science (информатики). Основной контингент школы — студенты старших курсов, аспиранты московских вузов и недавние выпускники.

Школа сотрудничает с магистратурой Высшей школы экономики и МФТИ, в которой при содействии Яндекса открыли новую базовую кафедру «Анализ данных» и кафедру Дискретной математики. Занятия проводятся в Московском корпусе МФТИ. Среди лекторов школы — известные учёные, такие как Алексей Червоненкис и Альберт Ширяев.[\[8\]](#)

1. Google

Google — самая крупная в мире поисковая система интернета, принадлежащая корпорации Google Inc.

Основана в 1998 году Сергеем Брином и Ларри Пейджем.

Первая по популярности система (77,05%), обрабатывает 41 млрд. 345 млн. запросов в месяц (доля рынка 62,4%), индексирует свыше 25 миллиардов веб-страниц (на закрытой конференции в начале мая 2014 представитель Google упомянул, что на текущий момент проиндексировано 60 триллионов документов, и как можно заметить в результате тестов, счётчик в поиске Google ограничен числом 25 270 000 000, также на это число при выдаче влияют фильтры, добавленные в алгоритм ранжирования выдачи).

Название Google произошло от слова Гугол (Googol) намеренно искажённого Сергеем Брином, которое означает десять в сотой степени — 10^{100} .

Синтаксис запросов.

Интерфейс Google содержит довольно сложный язык запросов, помогающий ограничить область поиска отдельными доменами, языками, типами файлов и т.п. К примеру, поиск «intitle:Google site:wikipedia.org» выведет все статьи Википедии на всех языках, в заголовке которых есть слово Google. Мощный язык запросов в руках хакеров может использоваться для изучения веб-сайтов на уязвимости.

Поиск в найденном.

Для результатов поиска Google ранее предоставлял возможность повторного поиска, что позволяло производить поиск более детально. Для более детального поиска пользователям необходимо было указать дополнительные параметры, по которым осуществлялся отбор результатов, что позволяло сразу отобразить не только запрос, но и контекст, где он используется. Эта возможность упрощала процедуру поиска, исключив необходимость в открытии каждого результата.

Википоиск.

Поисковая технология, помогающая пользователю настроить результаты выдачи по поисковым запросам. Пользователь может удалить результаты из списка и поднять вверх списка. Технология была запущена фирмой Google весной 2009 года и проработала до осени. В настройках поиска сохранилась настройка для включения «википоиска», но в выдаче соответствующие компоненты управления отсутствуют. Другие поисковые системы такой функционал пока не предоставляли.

Голосовой поиск.

22 сентября 2010 года фирма запустила голосовой поиск в России. Чтобы выполнить поиск, следует нажать в телефоне кнопку рядом со строкой поиска и произнести свой запрос, телефон направит ваш голос на сервер, и браузер выдаст строку с распознанным вашим запросом и итогами поиска по нему.

Google Doodle.

По случаю праздника или круглой даты какой-либо широко известной личности, привычный логотип Google у некоторых или, реже, — у всех региональных доменов может меняться на праздничный, имеющий конкретную тематику, смысл, но в

стиле Google (англ. Holiday and Events — Google style!):

- 6 июля 2008 года — логотип в форме коллажа из фрагментов работ Марка Шагала в честь его 121-летия.
- 11 февраля 2010 года — на логотипе белорусского домена Google появились акварели Наполеона Орды по случаю его дня рождения.
- 21 мая 2010 года — интерактивный логотип в честь 30 лет с выхода игры Pac-Man. Логотип позволяет управлять Pac-Man'ом мышкой или клавишами со стрелками клавиатуры.
- 19 августа 2010 года — логотип был посвящён 50-летнему юбилею полёта в космос собак Белки и Стрелки.
- 9 июня 2011 года — логотип, посвящённый 96-й годовщине со дня рождения Леса Пола. Логотип предоставляет возможность генерировать звуки (с изображением колеблющейся гитарной струны), записывать и проигрывать мелодии, передавать записанное (как URL).
- 12 июля 2011 года — логотип с изображением Собора Василия Блаженного в честь 450-летия этого памятника культуры.
- 5 сентября 2011 года — логотип ко дню рождения Фредди Меркьюри, в котором показан мультклип, созданный фирмой Google на песню Don't Stop Me Now группы Queen.
- 11 ноября 2011 года — логотип с изображением Фёдора Достоевского в честь 190-летия со дня его рождения.
- 19 ноября 2011 года — логотип с изображением Михаила Васильевича Ломоносова к 300-летию со дня его рождения.
- 23 ноября 2011 года — к 60-летию первой публикации Станислава Лема (по мотивам иллюстраций польского художника Даниэля Мроза к Кибериаде).
- 18 декабря 2011 года — логотип к 90-летию со дня рождения Юрия Никулина.
- 25 января 2012 года — логотип к 74-летию со дня рождения Владимира Высоцкого.
- 17 февраля 2012 года — логотип к 106-летию со дня рождения Агнии Барто.
- 9 апреля 2012 года — к 182-летию со дня рождения Эдварда Мейбриджа. Логотип был представлен в виде 21 сектора разного цвета, в каждом из которых бежала скаковая лошадь.
- 23 апреля 2012 года — к 99-летию со дня получения Гидеоном Сундбэком патента на застёжку-молнию.
- 31 мая 2012 года — логотип к 100-летию Государственного музея изобразительных искусств имени А. С. Пушкина.
- 7 июля 2012 — логотип к 130-летию со дня рождения Янки Купалы.

- 23 августа 2012 года — логотип с изображением Ассоль (герой повести «Алые паруса»), ожидающей корабль, в честь 132-летия со дня рождения Александра Грина.
- 17 сентября 2012 года — дудл к 155-летию со дня рождения К. Э. Циолковского.
- 15 октября 2012 года — дудл к 107-летию со дня создания комикса «Маленький Нимо в стране снов».
- 16 января 2013 года — интерактивный дудл к 112-летию со дня рождения Фрэнка Замбони.
- 18 сентября 2013 года — 128-летию со дня рождения Узеира Гаджибекова.
- 4 ноября 2013 года — дудл к 84-летию со дня рождения Шакунталы Деви.
- 8 ноября 2013 года — интерактивный дудл к 129-летию со дня рождения Германа Роршаха.
- 23 ноября 2013 года — 50-летие со дня первой серии Доктора Кто.
- 21 апреля 2015 года — в честь Лохнесского чудовища в логотип вместо буквы L оказалась голова чудовища.
- 27 августа 2015 — дудл к 99-летию со дня рождения турецкого археолога Халет Чамбел.
- 9 ноября 2015 года — в честь дня рождения Хеди Ламарр.
- 17 декабря 2015 года — в честь 245-летия со дня рождения Людвиг Ван Бетховена.[\[9\]](#)

Вышеописанная теоретическая часть о двух самых популярных поисковых систем в России разрешает нам перейти к их сравнительному анализу на основе введения любого поискового запроса и сравнения количества найденных результатов.

Например, в поисковой строке набираем «Владимир Путин» и нажимаем на кнопку поиска.

Яндекс Найти Будьте в Плюсе Регистрация Войти

Поиск Картинки Видео Карты Маркет Новости Эфир Коллекции Знатоки Услуги Ещё

[instagram.com](#) > [president_vladimir_putin/](#)
699 публикаций · 243,6 тыс. подписчиков · 0 подписок

Владимир Путин — смотрите картинки
[Яндекс.Картинки](#) > владимир путин

молодец портрет двойники фото молодой биография в молодости

Владимир Владимирович Путин — ВКонтакте
[vk.com](#) > [vladimir_vladimirovichp](#)
Перейдите на страницу пользователя, чтобы посмотреть публикации или отправить сообщение.

Википедия
Российский государственный деятель, действующий президент Российской Федерации с 7 мая 2012 года.
Родился: 7 октября 1952 г. (66 лет), Санкт-Петербург, РСФСР, СССР
Родители: Мария Ивановна Шеломова, Владимир Спиридонович Путин
Рост: 170 см

Смотрите также

Людмила Александровна Путина
Дмитрий Анатольевич Медведев
Рамзан Кадыров
Собчак Ксения Анатольевна
Жириновский Владимир Вольфович
Барак Обама

Нашлось 5 млн результатов
236 тыс. показов в месяц
[Дать объявление](#)

Чаты

Рисунок 1 - «Владимир Путин» в Яндексе

Google Войти

Все Картинки Новости Видео Карты Ещё Настройки Инструменты

Результатов: примерно 78 000 000 (0,56 сек.)

Главные новости

Владимир Путин отправит в отставку замглавы Росгвардии Сергея Захаркина
МК
8 часов назад

Путин прибыл на заседание совета глав государств ШОС в Бишкеке
РИА Новости
29 минут назад

Путин прибыл на заседание лидеров стран ШОС
Life.ru
2 часа назад

[Больше результатов по запросу "владимир путин"](#)

Путин, Владимир Владимирович — Википедия
https://ru.wikipedia.org/wiki/Путин,_Владимир_Владимирович
Владимир Владимирович Путин (род. 7 октября 1952, Ленинград, РСФСР, СССР) — российский государственный и политический деятель, ...

Владимир Владимирович Путин
Президент Российской Федерации

Владимир Владимирович Путин — российский государственный и политический деятель, действующий президент Российской Федерации, Верховный главнокомандующий Вооружёнными силами Российской Федерации с 7 мая 2012 года. Википедия

Родился: 7 октября 1952 г. (возраст 66 лет), Санкт-Петербург
Рост: 1,7 м
Супруга: Людмила Путина (в браке с 1983 г. до 2014 г.)
Дети: Екатерина Тихонова, Мария Владимировна Путина
Сроки президентства: 7 мая 2000 г. – 7 мая 2008 г., 7 мая 2012 г. –
Образование: Санкт-Петербургский Горный Университет (1997 г.), ЕЩЁ

Похожие запросы Ещё 15+

Рисунок 2 - «Владимир Путин» в Google

Как можно увидеть на рисунках 1 и 2, Яндекс нашел около 5 млн. результатов, а Google – примерно 78 млн., т.е. практически в 16 раз больше.

Теперь попробуем скорректировать запрос и ввести в поисковую строку «Владимир Владимирович Путин».

The screenshot shows the Yandex search engine interface. The search bar contains the text "владимир владимирович путин". Below the search bar, there are several search filters: "Поиск", "Картинки", "Видео", "Карты", "Маркет", "Новости", "Эфир", "Коллекции", "Знатоки", "Услуги", and "Ещё". The main search results are displayed in two columns. The left column features a link to a VK.com profile for "vladimir_vladimirovichp", a section titled "Владимир владимирович Путин — смотрите картинки" with a grid of images and their dimensions (1000x744, 800x600, 773x1160, 1200x801), and an Instagram profile for "@president_vladimir_putin" with 699 publications and 243.6 thousand followers. The right column features a Wikipedia entry for "Владимир Владимирович Путин", a section titled "Смотрите также" with a row of six small portraits of other public figures, and a "Чаты" (Chats) button. The bottom of the page shows the search statistics: "Нашлось 6 млн результатов" and "113 тыс. показов в месяц".

Рисунок 3 - «Владимир Владимирович Путин» в Яндексе

The screenshot shows the Google search engine interface. The search bar contains the text "владимир владимирович путин". Below the search bar, there are several search filters: "Все", "Карты", "Картинки", "Новости", "Видео", "Ещё", "Настройки", and "Инструменты". The main search results are displayed in two columns. The left column features a link to the Wikipedia article for "Путин, Владимир Владимирович", a link to the official website "putin.kremlin.ru/bio", and a section titled "Главные новости" with two news items: "Путин пообщался с президентом Киргизии" and "Сотрудника «Севмаша», вывесившего баннер во время телемоста с Путиным, восстановили...". The right column features a large portrait of Vladimir Putin, a section titled "Владимир Владимирович Путин" with a sub-header "Президент Российской Федерации", and a detailed biographical summary including birth date (7 октября 1952 г.), birthplace (Санкт-Петербург), height (1,7 м), spouse (Людмила Путина), children (Екатерина Тихонова, Мария Владимировна Путина), and education (Санкт-Петербургский Горный Университет). The bottom of the page shows the search statistics: "Результатов: примерно 5 560 000 (0,43 сек.)".

Рисунок 4 - «Владимир Владимирович Путин» в Google

Добавив дополнительное слово в поисковый запрос, Яндекс выдал на 1 млн. результатов больше (6 млн.), а Google, наоборот, в 15 раз уменьшил количество страниц, так как чаще всего он ищет не каждое отдельное слово, как это делает Яндекс, а сразу полное словосочетание, что уменьшает количество запрашиваемых страниц вне зависимости от их уровней необходимости, качества выданной информации и интересности.

ЗАКЛЮЧЕНИЕ

Как говорилось ранее, поисковая система – это компьютерная система, используемая для поиска необходимой информации. Её главная задача – минимизировать время, затрачиваемое пользователем на поиск релевантной запросу информации. Релевантность – одно из наиболее запутанных и субъективных определений в науке информационного поиска. Чаще всего говорят о релевантности с позиции пользователя, и в этом случае «релевантная запросу информация» и «нужная пользователю информация» – одно и то же. Как раз о такой релевантности мы говорим в этом разделе. Вопрос заключается в том, какие данные пользователь посчитает нужными? В некоторых обстоятельствах релевантную информацию можно определить как всю информацию из базы, относящуюся к запросу. Так, например, если пользователю следует узнать всё об определённой компании, то он заинтересован в нахождении всех документов, в которых упоминается об этой компании. В других обстоятельствах релевантная информация – это только та информация, которая достаточна для выполнения конкретной пользовательской задачи, к примеру, поиска ответа на определённый вопрос. Если в последней ситуации в результатах поиска окажется большое количество избыточных данных, т.е. данных, имеющих отношение к запросу, но не нужных для выполнения этой задачи, то выборка нужной/релевантной информации займет у пользователя дополнительное время.

Итак, обычно к поисковой системе применяют две главные характеристики: точность и полнота, а точнее, их зависимость. Каждый раз, когда пользователь задает системе свой запрос, тем самым инициализируя поиск, все документы в коллекции поисковой системы подразделяются на четыре части. Точность определяет один аспект поиска, а именно, насколько хорошо поисковая система может минимизировать время, затрачиваемое пользователем на поиск релевантной этому запросу информации. В то время как полнота определяет ещё

один аспект – насколько хорошо система может найти релевантную этому запросу информацию. Можно выбрать оптимальный запрос, когда каждый найденный документ окажется релевантным, и каждый релевантный документ будет найден.

Проведя сравнительный анализ двух крупных и популярных поисковых систем, мы пришли к выводу, что нету идеальной системы по нахождению релевантной информации. И, скорее всего, это зависит не только от начинки самой поисковой системы, но и от правильности формулирования запрашиваемой информации.

СПИСОК ИСТОЧНИКОВ

1. Google (поисковая система) [Электронный ресурс] – М.: Википедия, 2019. Режим доступа: [https://ru.wikipedia.org/wiki/Google_\(поисковая_система\)](https://ru.wikipedia.org/wiki/Google_(поисковая_система)) (дата обращения: 01.08.2019).
2. Давлетов, З.Х. Основы современной информатики: Учебное пособие [Текст] / З.Х. Давлетов. - СПб.: Лань КПТ, 2016. - 256 с.
3. Дискуссия на тему: «Какая из поисковых систем лучше?» [Электронный ресурс] – М.: Readera, 2019. Режим доступа: <https://readera.ru/diskussija-na-temu-kakaja-iz-poiskovyh-sistem-luchshe-140129861> (дата обращения: 01.08.2019).
4. Жаров, М.В. Основы информатики: Учебное пособие [Текст] / М.В. Жаров, А.Р. Палтиевич, А.В. Соколов. - М.: Форум, 2017. - 512 с.
5. Информатика. Введение [Электронный ресурс] – М.: StudWood.ru, 2017. Режим доступа: <https://m.studwood.ru/1938320/informatika/vvedenie> (дата обращения: 01.08.2019).
6. История развития поисковых систем [Электронный ресурс] – М.: SEOkleo, 2019. Режим доступа: <http://seokleo.ru/istoriya-razvitiya-poiskovykh-sistem/> (дата обращения: 01.08.2019).
7. Классификация поисковых систем [Электронный ресурс] – М.: InteWiki, 2008. Режим доступа: https://wiki.iteach.ru/index.php/Классификация_поисковых_систем (дата обращения: 01.08.2019).
8. Кудинов, Ю.И. Основы современной информатики: Учебное пособие [Текст] / Ю.И. Кудинов, Ф.Ф. Пашенко. - СПб.: Лань, 2018. - 256 с.
9. Ляхович, В.Ф. Основы информатики (СПО) [Текст] / В.Ф. Ляхович, В.А. Молодцов, Н.Б. Рыжикова. - М.: КноРус, 2018. - 264 с.
10. Матросов, В.Л. Теоретические основы информатики: Учебник [Текст] / В.Л. Матросов. - М.: Academia, 2017. - 832 с.

11. Поисковые системы [Электронный ресурс] – М.: ЮниВеб, 2016. Режим доступа: https://uniofweb.ru/wiki/poiskovy_e_sistemy/ (дата обращения: 01.08.2019).
 12. Поисковый робот Яндекс и Google. История развития поисковых систем [Электронный ресурс] – М.: КНЕР, 2019. Режим доступа: <https://kner.ru/tech/poiskovyj-robot-yandeks-i-google.html> (дата обращения: 01.08.2019).
 13. Сети ЭВМ и технологии распределенной обработки данных. Основные принципы поиска в информации [Электронный ресурс] – М.: Кафедра экономической кибернетики и экономико-математических методов, 2019. Режим доступа: http://ecocyb.narod.ru/410-417/inrs5_6.htm (дата обращения: 01.08.2019).
 14. Стариченко, Б.Е. Теоретические основы информатики: Учебник [Текст] / Б.Е. Стариченко. - М.: ГЛТ, 2016. - 400 с.
 15. Яндекс [Электронный ресурс] – М.: Википедия, 2019. Режим доступа: <https://ru.wikipedia.org/wiki/Яндекс> (дата обращения: 01.08.2019).
-
1. Поисковые системы [Электронный ресурс] – М.: ЮниВеб, 2016. Режим доступа: https://uniofweb.ru/wiki/poiskovy_e_sistemy/ (дата обращения: 01.08.2019) [↑](#)
 2. Информатика. Введение [Электронный ресурс] – М.: StudWood.ru, 2017. Режим доступа: <https://m.studwood.ru/1938320/informatika/vvedenie> (дата обращения: 01.08.2019) [↑](#)
 3. Сети ЭВМ и технологии распределенной обработки данных. Основные принципы поиска в информации [Электронный ресурс] – М.: Кафедра экономической кибернетики и экономико-математических методов, 2019. Режим доступа: http://ecocyb.narod.ru/410-417/inrs5_6.htm (дата обращения: 01.08.2019) [↑](#)
 4. Дискуссия на тему: «Какая из поисковых систем лучше?» [Электронный ресурс] – М.: Readera, 2019. Режим доступа: <https://readera.ru/diskussija-na-temu-kakaja-iz-poiskovyh-sistem-luchshe-140129861> (дата обращения: 01.08.2019) [↑](#)
 5. История развития поисковых систем [Электронный ресурс] – М.: SEOkleo, 2019. Режим доступа: <http://seokleo.ru/istoriya-razvitiya-poiskovykh-sistem/> (дата обращения: 01.08.2019) [↑](#)

6. Поисковый робот Яндекс и Google. История развития поисковых систем [Электронный ресурс] – М.: КНЕР, 2019. Режим доступа: <https://кнер.ru/tech/poiskovyj-robot-yandeks-i-google.html> (дата обращения: 01.08.2019) [↑](#)

7. Классификация поисковых систем [Электронный ресурс] – М.: InteWiki, 2008. Режим доступа: https://wiki.iteach.ru/index.php/Классификация_поисковых_систем (дата обращения: 01.08.2019) [↑](#)

8. Яндекс [Электронный ресурс] – М.: Википедия, 2019. Режим доступа: <https://ru.wikipedia.org/wiki/Яндекс> (дата обращения: 01.08.2019) [↑](#)

9. Google (поисковая система) [Электронный ресурс] – М.: Википедия, 2019. Режим доступа: [https://ru.wikipedia.org/wiki/Google_\(поисковая_система\)](https://ru.wikipedia.org/wiki/Google_(поисковая_система)) (дата обращения: 01.08.2019) [↑](#)