

Содержание:

ВВЕДЕНИЕ

Стремительный рост популярности интернета в последние десятилетия обусловлен возможностью простого и быстрого поиска информации. Обеспечивают такую возможность поисковые системы.

Поисковые системы – это сервисы, которые сканируют все сайты, находящиеся в сети, анализируют информацию на страницах и выдают пользователям список страниц подходящих под запрос. Для сбора и обработки данных с сайтов используются различные программные алгоритмы (роботы).

В настоящее время, мы не представляем себе интернет без поисковой системы. Каждый из нас пользовался или пользуется поисковыми системами, такими как Google, Yandex, Bing, Yahoo и т.д. И наверное, каждый пользователь хоть раз задавался вопросом: «А правилен ли мой выбор?» или «Сколько еще человек пользуется такой же поисковой системой, как я?».

Целью данной курсовой работы является анализ поисковых систем в сети Интернет. Для этого необходимо рассмотреть принципы работы и функционирования поисковых систем, схему обработки запросов; определить наиболее популярные поисковые системы мира и Рунета.

ГЛАВА 1. ПОИСКОВЫЕ СИСТЕМЫ: ПОНЯТИЕ, ФУНКЦИИ, ПРИНЦИПЫ РАБОТЫ

1.1 Понятие и функции поисковой системы

Поисковая система - это программно-аппаратный комплекс, предназначенный для осуществления поиска в сети Интернет и реагирующий на запрос пользователя, задаваемый в виде текстовой фразы (поискового запроса), выдачей списка ссылок на источники информации, в порядке релевантности (в соответствии запросу). Наиболее крупные международные поисковые системы: «Google», «Yahoo», «MSN».

В русском Интернете это – «Яндекс», «Рамблер», «Апорт».

Рассмотрим понятие поискового запроса. Поисковый запрос должен быть сформулирован пользователем в соответствии с тем, что он хочет найти, максимально кратко и просто. Для этого открываем главную страницу поисковой системы и вводим текст поискового запроса. Далее, задача сводится к тому, чтобы открыть предоставленные по запросу ссылки на источники информации в Интернет.

Однако, возможна ситуация, когда нужная информация не найдена. Если это произошло, то нужно либо перефразировать свой запрос, либо в базе поисковой системе действительно нет никакой актуальной информации по нашему запросу.

Основная 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 создают такие алгоритмы и принципы работы поисковых систем, которые бы позволяли находить пользователям искомую ими информацию (рис.1).

Это означает, поисковая система должна «думать» так же, как думает пользователь при поиске информации.

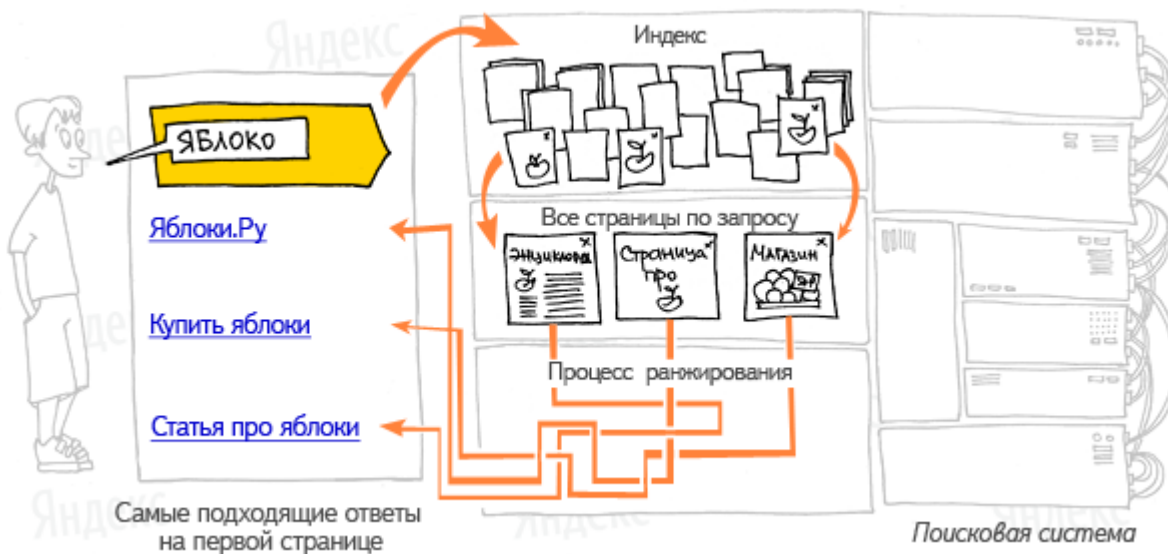


Рисунок 1. Пример поиска информации

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 [2]:

- Нашел ли он то, что искал?
- Если не нашел, то сколько раз ему пришлось перефразировать запрос, чтобы найти искомое?
- Насколько актуальную информацию он смог найти?
- Насколько быстро обрабатывала запрос поисковая машина?
- Насколько удобно были представлены результаты поиска?
- Был ли искомый результат первым или же сотым?
- Как много ненужного мусора было найдено наравне с полезной информацией?
- Найдется ли нужная информация, при обращении к поисковой системе, скажем, через неделю, или через месяц?

Для того, чтобы получить правильные ответы на подобные вопросы, разработчики поисковых систем постоянно совершенствуют алгоритмы и принципы поиска, добавляют новые функции и возможности, всячески пытаются ускорить работу системы.

1.2 Основные характеристики поисковой системы

Рассмотрим основные характеристики поисковых систем [3]:

- *Полнота*

Полнота представляет собой отношение количества найденных по запросу информационных документов к общему числу документов в сети Интернет, относящихся к данному запросу. Например, если в Интернете имеется 100 страниц, содержащих словосочетание «как выбрать автомобиль», а по соответствующему запросу было найдено всего 60 из них, то полнота поиска будет 0,6. В этом случае, чем полнее поиск, тем меньше вероятность того, что пользователь не найдет нужный ему документ, при условии, что он вообще существует в Интернете.

- *Точность*

Точность определяется степенью соответствия найденных документов запросу пользователя. Например, если по запросу «как выбрать автомобиль» находится 100 документов, в 40 из них содержится словосочетание «как выбрать автомобиль», а в остальных просто наличествуют эти слова («как правильно выбрать магнитолу и установить в автомобиль»), то точность поиска считается равной $40/100 (=0,4)$. Чем

точнее поиск, тем быстрее пользователь найдет нужные ему документы, тем меньше различного рода «мусора» среди них будет встречаться, тем реже найденные документы не будут соответствовать запросу.

- *Актуальность*

Актуальность характеризуется временем, проходящим с момента публикации документов в сети Интернет, до занесения их в индексную базу поисковой системы. Например, на следующий день после появления интересной новости, большое количество пользователей обратились к поисковым системам с соответствующими запросами. В большинстве случаев информация об этой новости уже доступна в поиске, хотя с момента публикации новостной информации на эту тему прошло меньше суток. Это происходит благодаря наличию у крупных поисковых систем так называемой «быстрой базы», которая обновляется несколько раз в день.

- *Скорость поиска*

Скорость поиска тесно связана с так называемой «устойчивостью к нагрузкам». Например, по данным ООО «Рамблер Интернет Холдинг», на сегодняшний день в рабочие часы к поисковой машине Рамблер приходит около 60 запросов в секунду. Такая загруженность требует сокращения времени обработки отдельного запроса. Интересы пользователя и поисковой системы в этом случае совпадают: посетитель желает получить результаты как можно быстрее, а поисковая машина должна обрабатывать запрос максимально оперативно, чтобы не притормозить обработку следующих запросов.

- *Наглядность*

Наглядность представления результатов является важнейшим компонентом удобного поиска. В результате большинства запросов поисковая машина находит сотни, а то и тысячи документов. Из-за нечеткости составления запросов или неточности поиска, даже первые страницы выдачи не всегда содержат нужную информацию. Это означает, что пользователю часто приходится производить свой собственный поиск среди предоставленных результатов. Различные элементы страницы выдачи поисковой системы помогают ориентироваться в результатах поиска.

1.3 История развития поисковых систем

В начальный период развития Интернет, число его пользователей было небольшим, а объем доступной информации сравнительно невеликим. В тот период доступ к ней имели, в основном, работники различных крупных учебных заведений и лабораторий, и полученные данные использовались в научных целях.

Использование Сети не имело такой актуальности, как сейчас.

В 1990 году британский ученый Тим Бернерс-Ли (который также является изобретателем URI, URL, HTTP, World Wide Web) создал сайт **info.cern.ch**, который является первым в мире доступным каталогом интернет-сайтов. С этого момента Интернет начал набирать популярность не только среди научных кругов, но и среди простых обладателей персональных компьютеров [1].

Так, первым способом облегчения доступа к информационным ресурсам в Интернете стало формирование каталогов сайтов. Ссылки на ресурсы в них были сгруппированы по тематике.

Первым проектом такого рода принято считать Yahoo, открытый **в апреле 1994 года**. В связи со стремительным ростом количества сайтов в нём, вскоре появилась возможность поиска необходимой информации по запросу. Конечно же, это ещё не было полноценной поисковой системой. Поиск был ограничен только данными, которые находились в каталоге.

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 известный и большой каталог сети DMOZ (его еще называют Open Directory Project) содержит информацию о 25 миллионах страниц [2].

Хронология развития поисковых систем [1]

1945 год – американский инженер Ванневар Буш опубликовал записи идеи, которая привела в дальнейшем к изобретению гипертекста, и рассуждение о необходимости разработки системы быстрого извлечения данных из хранимой информации таким образом (эквивалент сегодняшних поисковых систем).

Введённое им понятие устройства-расширителя памяти содержало оригинальные идеи, которые, в конце концов, воплотились в Интернете.

1960-е — Джерард Сэлтон и его группа в Корнелльском университете разработали «Остроумную систему извлечения информации» (SMART information retrieval system). SMART — аббревиатура от Salton's Magic Automatic Retriever of Text, то есть «Волшебный автоматический извлекатель текста Сэлтона». Джерард Сэлтон считается отцом современной поисковой технологии.

1987-1989 – разработана **Archie** — поисковая система для индексации FTP архивов. Archie представлял из себя сценарий, автоматизирующий внедрение в листинги на ftp-серверах, которые затем переносились в локальные файлы, а уже потом в локальных файлах осуществлялся быстрый поиск необходимой информации. Поиск основывался на стандартной grep-команде Unix, а доступ пользователя к данным осуществлялся на основе telnet.

В следующей версии данные были разбиты на отдельные базы, одна из которых содержала только текстовые названия файлов; а другая — записи со ссылками на иерархические директории тысячи хостов; и еще одна, соединяющая первые две. Эта версия Archie была эффективней предыдущей, так как поиск производился только по именам файлов, исключая множество существующих ранее повторов.

Поисковая система становилась всё популярнее, и разработчики задумались, как ускорить её работу. Упомянутая выше база данных была заменена на другую, основанную на теории сжатого дерева. Новая версия, по существу, создала полнотекстную базу данных вместо списка имен файлов и была значительно быстрее, чем раньше. В дополнение, второстепенные изменения позволили системе Archie индексировать web-страницы. К сожалению, по различным причинам, работа над Archie вскоре прекратилась.

В 1993 году была создана первая в мире поисковая система для Всемирной сети **Wandex**. В её основу был заложен World Wide Web Wanderer бот, разработанный Метью Греем из Массачусетского технологического института.

1993 год – Мартин Костер создаёт **Aliweb** – одну из первых поисковых систем по World Wide Web. Владельцы сайтов должны были сами их добавлять в индекс Aliweb, чтобы они появлялись в поиске. Поскольку слишком мало вебмастеров это делали, Aliweb не стал популярным

20 апреля 1994 г. – Брайан Пинкертон из университета Вашингтон выпустил **WebCrawler** — первого бота, который индексировал страницы полностью. 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. К сожалению, пропускная способность была невысокой и в дневное время система часто была недоступной.

20 июля 1994 г. – открылся **Lycos** — серьезная разработка в технологии поиска, созданная в университете Карнеги Мелон. Майкл Малдин был ответственен за эту поисковую систему и до сих пор остаётся ведущим специалистом в Lycos Inc. Lycos открылся с каталогом в 54,000 документов. И в дополнение к этому результаты, которые он предоставлял, были ранжированными, кроме того он учитывал приставки и приблизительное совпадение. Но главным отличием Lycos был постоянно пополняемый каталог: к ноябрю 1996 было проиндексировано 60 миллионов документов — больше, чем у любой другой поисковой системы того времени.

Январь 1994 г. — был основан **Infoseek**. Он не был по-настоящему инновационным, но имел ряд полезных дополнений. Одним из таких популярных дополнений была возможность добавления своей страницы в реальном времени.

1995 год – запустилась **AltaVista (рис.2)**. Появившись, поисковая система AltaVista быстро получила признание пользователей и стала лидером среди себе подобных. У системы была практически неограниченная на то время пропускная способность, она была первой поисковой системой, в которой было возможно формулировать запросы на естественном языке, а также формулировать сложные запросы. Пользователям было разрешено добавлять или удалять их собственные URL в течение 24 часов.



Рисунок 2. **AltaVista**

Также AltaVista предлагала много советов и рекомендаций по поиску. Основной заслугой системы AltaVista считается обеспечение поддержки множества языков, в том числе китайского, японского и корейского. Действительно, в 1997 году ни одна поисковая машина в Сети не работала с несколькими языками, тем более с редкими.

1996 год — поисковая машина AltaVista запустила морфологическое расширение для русского языка.

Также, в 1996 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 внутри Сети.

20 мая 1996 г. — появилась корпорация Inktomi вместе со своим поисковиком **Hotbot**. Его создателями были две команды из калифорнийского университета. Когда сайт появился, то он быстро стал популярным.

1997 год – в западных странах наступает переломный момент в развитии поисковых систем, когда С. Брин и Л. Пейдж из Стэнфордского университета основали **Google** (первоначальное название проекта BackRub) (рис.3). Они разработали собственную поисковую машину, которая дала пользователям возможность осуществлять качественный поиск с учетом морфологии, ошибок при

написании слов, а также повысить релевантность в результатах выдачи запросов.



Рисунок 3. С. Брин и Л. Пейдж

23 сентября 1997 года – анонсирован **Yandex**, который быстро стал самой популярной у русскоязычных пользователей Интернета системой поиска.

С запуском в 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

Таким образом, развитие поисковых систем и их становление можно охарактеризовать перечисленными выше этапами.

1.4 Состав и принципы работы поисковой системы

Практически все крупные поисковые системы имеют свою собственную структуру, отличную от других. Однако можно выделить общие для всех поисковых машин основные компоненты [2]. Различия в структуре могут быть лишь в виде реализации механизмов взаимодействия этих компонентов.

- **Модуль индексирования**

Модуль индексирования состоит из трех вспомогательных программ (роботов):

1. Spider (паук) – программа, предназначенная для скачивания веб-страниц (рис.4). «Паук» обеспечивает скачивание страницы и извлекает все

внутренние ссылки с этой страницы. Скачивается html-код каждой страницы. Для скачивания страниц роботы используют протоколы HTTP.

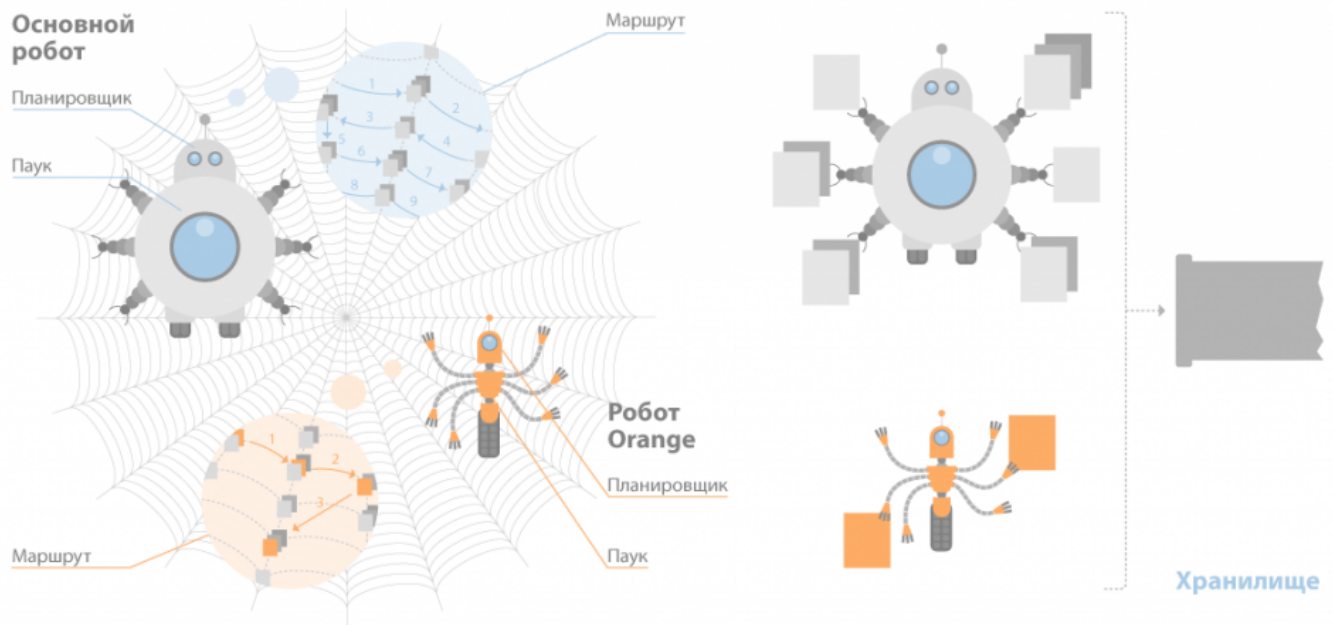


Рисунок 4. Spider

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).

1. 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.

2. Indexer (робот-индексатор) - программа, которая анализирует веб-страницы, скаченные пауками (рис.5).

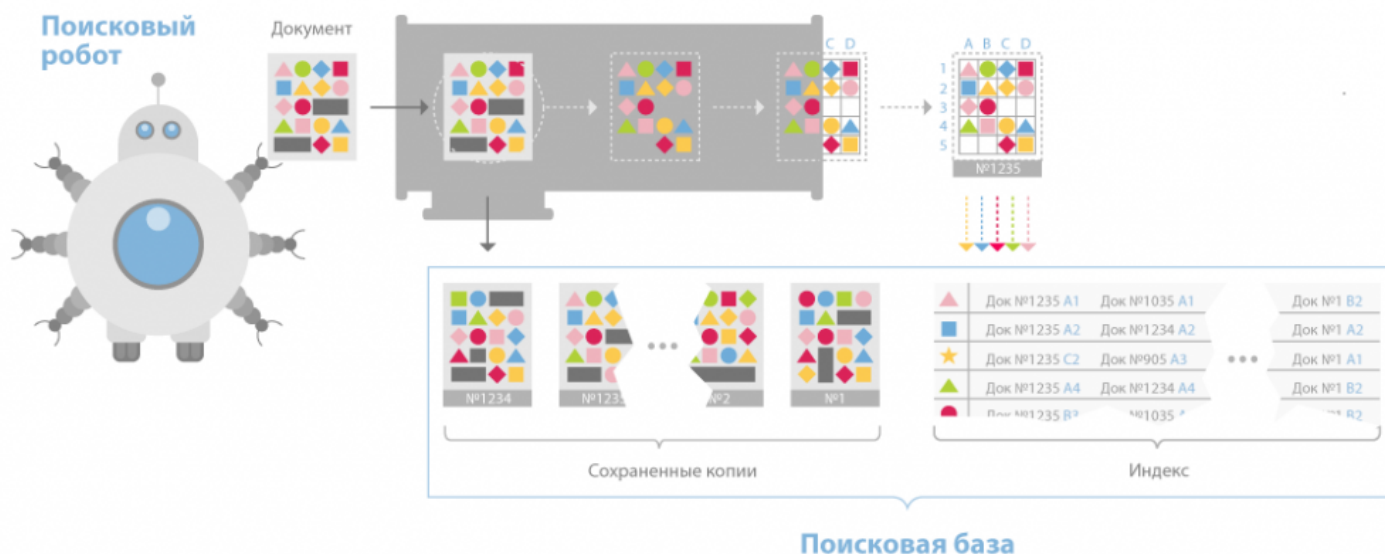


Рисунок 5. Indexer

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 скаченных и обработанных модулем индексирования документов.

- **Поисковый сервер**

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

Поисковый сервер работает следующим образом:

1. Запрос, полученный от пользователя, 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).
2. 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.
3. В зависимости от выбора пользователя этот рейтинг может быть скорректирован дополнительными условиями (например, так называемый «расширенный поиск»).
4. Далее генерируется сниппет, то есть, для каждого найденного документа из таблицы документов извлекаются заголовок, краткая аннотация, наиболее соответствующая запросу и ссылка на сам документ, причем найденные слова подсвечиваются.
5. 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.

Например, по информации ООО «Рамблер Интернет Холдинг» обработка поискового запроса в системе «Рамблер» происходит, так, как это изображено на рисунке 6 [3].

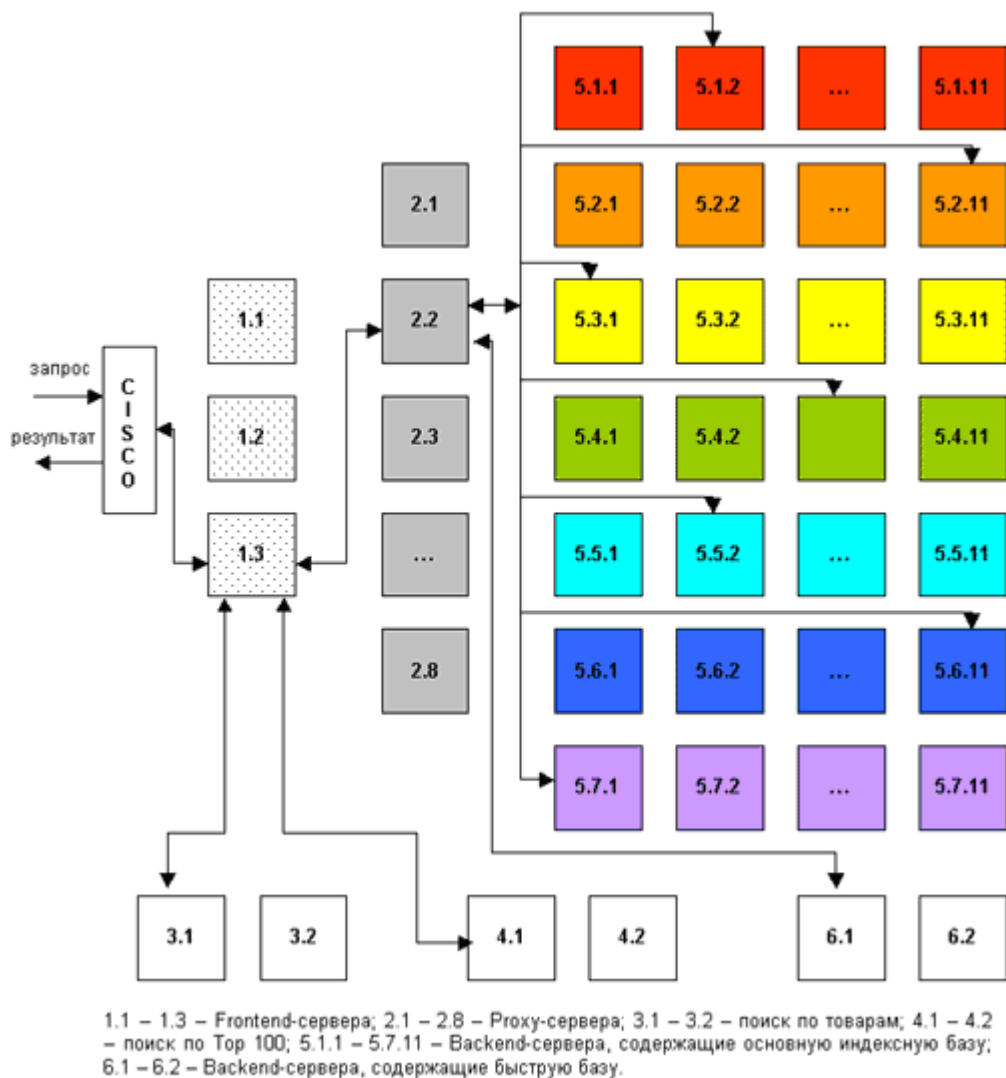


Рисунок 6. Обработка поискового запроса в системе «Рамблер»

Запрос поступает в поисковую систему через маршрутизатор Cisco 6000 series. Cisco передает его наименее загруженной машине первого уровня - frontend (1.1 - 1.3, на рис. машине 1.3). Frontend, в свою очередь, отправляет запрос дальше, на один из восьми proxy-серверов, также выбирая наиболее свободный сервер (2.1 - 2.8, на рис. машине 2.2). Одновременно frontend отправляет запрос на машины, осуществляющие поиск по товарам (3.1 - 3.2, на рис. машине 3.1) и по базе Top 100 (4.1 - 4.2, на рис. машине 4.1). На proxy проводится поиск по ссылочному индексу, и его результаты вместе с поисковым запросом передаются на машины, которые содержат основную индексную базу, - backends (5.1.x - 5.7.x, на рис. машинам 5.1.2, 5.2.11, 5.3.1 и т.д.) Та же информация отправляется на машины с "быстрой базой" (6.1 - 6.2).

На текущий момент в поиск включено 77 backend'ов. Они сгруппированы по 11 машин, и каждая группа содержит копию одной из частей поискового индекса.

Таким образом, информация о сайтах, условно входящих в красный сектор Интернета, находится на backend'ах первой группы (5.1.1 - 5.1.11 на рис), оранжевый сектор - на backend'ах второй группы (5.2.1 - 5.2.11) и т.д. Проху-сервер выбирает наименее загруженный backend в каждой группе машин и отправляет на него поисковый запрос с результатами ссылочного поиска. На backend'ах осуществляется поиск по частям индексной базы и ранжирование с учетом результатов поиска по ссылочному индексу. При ранжировании для всех найденных документов высчитываются веса по конкретному запросу.

После того, как запрос обработан на backend'ах, информация о результатах и ранжировании отдается обратно на проху-сервер. Туда же поступают отсортированные результаты с машин "быстрой базы". Проху интегрирует данные, полученные с восьми машин: клеит дубли, объединяет зеркала сайтов, переранжирует документы в общий список по весам, рассчитанным на backend'ах. Так, первым в списке найденного может быть документ с машины 5.3.1, вторым и третьим - с 6.1, четвертым - с 5.5.2 и т.д. На проху-сервере также реализуется построение цитат к документам и подсветка слов запроса в тексте. Полученные результаты отдаются на frontend.

Помимо информации с проху-сервера, frontend получает результаты из поиска по товарам и из базы Top 100, отсортированные, с цитатами и подсветкой слов запроса. Frontend осуществляет окончательное объединение результатов, генерирует html со списком найденного, вставляет баннеры и перевязки (ссылки на различные разделы Рамблера) и отдает html Cisco, который маршрутизирует информацию пользователю.

ГЛАВА 2. РЕЙТИНГ ПОИСКОВЫХ СИСТЕМ ИНТЕРНЕТА

2.1 Популярные поисковые системы мира

Рейтинги поисковых систем строятся, основываясь на двух критериях:

- качество поиска;
- популярность.

Напрямую оценивать качество получается не всегда, так как этот параметр имеет субъективный характер – одним кажется, что лучше ищет Яндекс, другим – Google, третьи отдадут предпочтение интеллектуальной поисковой системе Нигма и т.д.

Более менее объективную информацию может дать только статистика посещений. Именно эта статистика и будет использована при анализе популярных поисковых систем.

В настоящее время существуют три основные поисковые системы (международные) – Google, Bing и Yahoo, имеющие собственные базы и алгоритмы поиска. Остальные поисковые системы (большинство) используют в том или ином виде результаты трех перечисленных. Например, поиск AOL (search.aol.com) использует базу Google, а AltaVista, Lycos и AllTheWeb – базу Yahoo.

По данным на ноябрь 2016 года, доли поисковых систем в мире распределены следующим образом (рисунок 7) [5]:

- Google — 69,24 %;
- Bing — 12,26 %;
- Yahoo! — 9,19 %;
- Baidu — 6,48 %;
- AOL — 1,11 %;
- Ask — 0,23 %;
- Excite — 0,00 %.

Доли поисковых система в мире

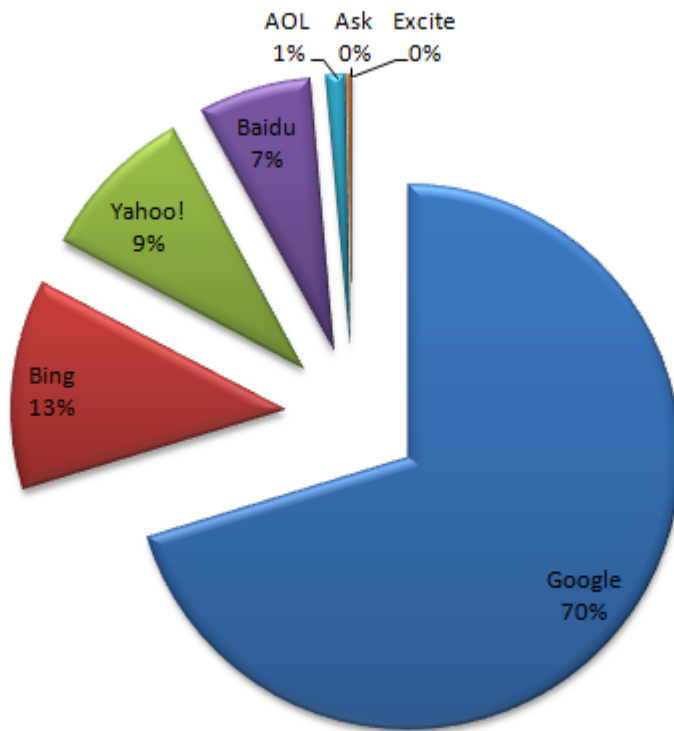


Рисунок 7. Доля поисковых систем в мире (на ноябрь 2016года)

Вот данные исследовательской компании NetMarketShare на декабрь 2016 года [8]. В таблице собраны самые популярные поисковики, и процент от общего числа запросов пользователей ПК, который они обрабатывают:

Таблица 1

Рейтинг поисковых систем мира (декабрь 2016)

Поисковая система Доля поисковых запросов

Google 75,97%

Bing 8,28%

Baidu 7,54%

Yahoo 6,56%

AOL 0,10%

Для наглядности, представим данные таблицы 1 в виде диаграммы (рисунок 8):

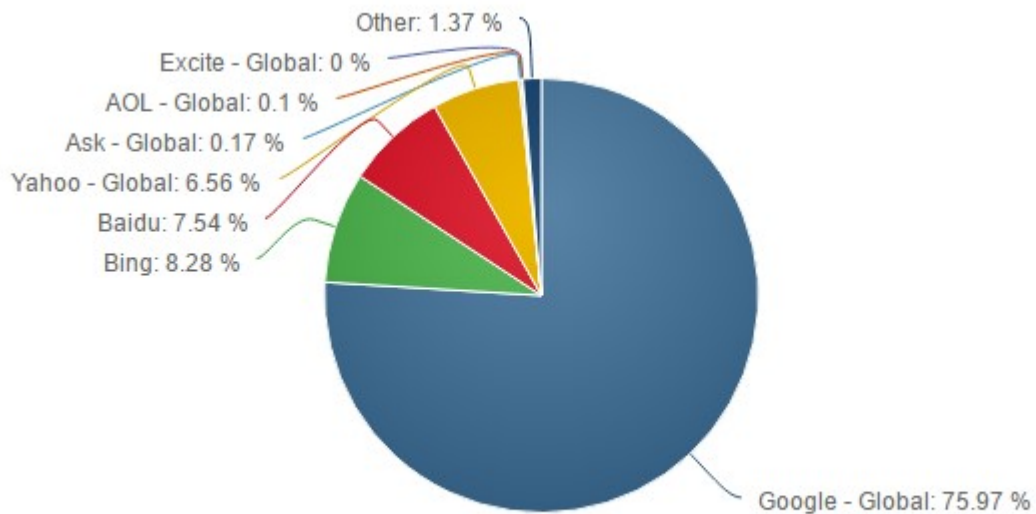


Рисунок 8. Доля поисковых систем в мире (на декабрь 2016 года)

Как видно из рисунков 7 и 8, популярность Google выросла на 6%, а Bing и Yahoo! уменьшилась на 4% и 5% соответственно.

Для мобильного поиска рейтинг выглядит по-другому (таблица 2) [8]:

Таблица 2

Рейтинг поисковых систем на мобильных устройствах (декабрь 2016)

Поисковая система Доля поисковых запросов

Google 94,27%

Yahoo	3,33%
Bing	1,11%
Baidu	0,49%
ASK	0,04%
AOL	0,01%

Очевидно, самая популярная поисковая система в мире на данный момент – Google. Она занимает очень значительную долю на рынке десктопов, а для мобильных устройств Google, по сути, является монополистом.

Для наглядности, представим данные таблицы 2 в виде диаграммы (рисунок 9):

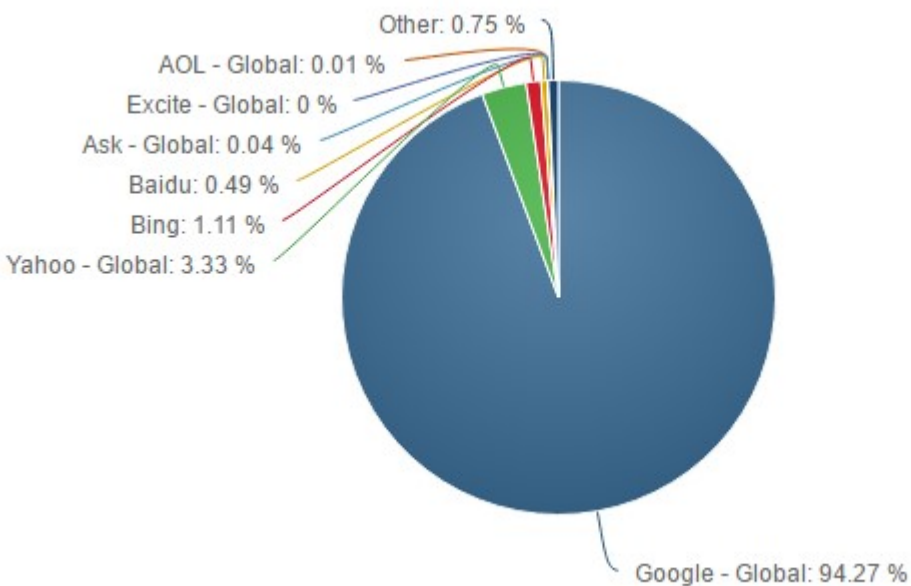


Рисунок 8. Доля поисковых систем на мобильных устройствах (декабрь 2016года)

лей моПри этом за последний год Google только укрепил свои позиции на ПК с 65,44% в начале года до 75,97% в настоящий момент (рисунок 9) [5].

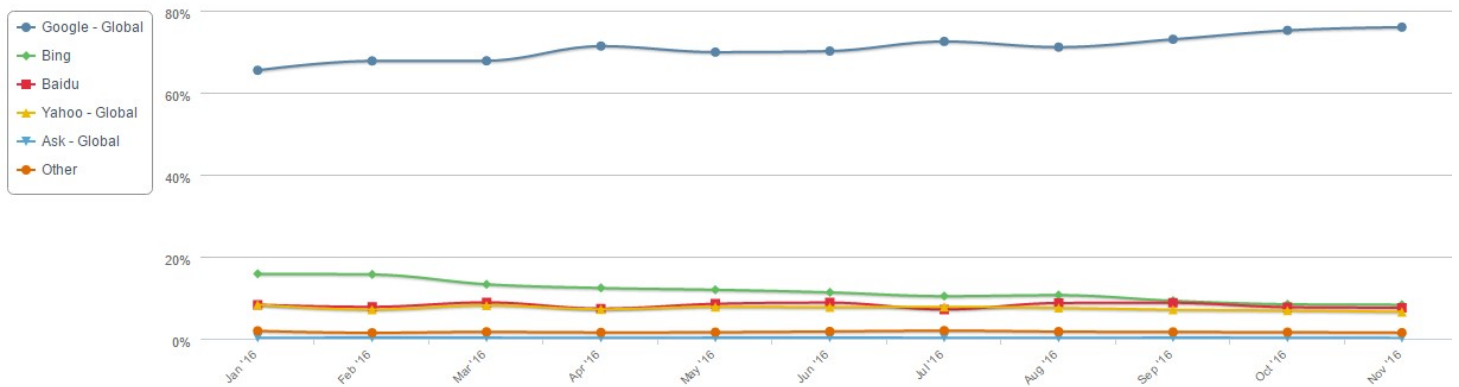


Рисунок 9. Динамика использования поисковых систем на ПК за 2016 год

Предпочтения пользователей мобильных устройств за этот год почти не менялись. В начале года у Google было 94,13%, а сейчас 94,27% (рисунок 10) [5]:

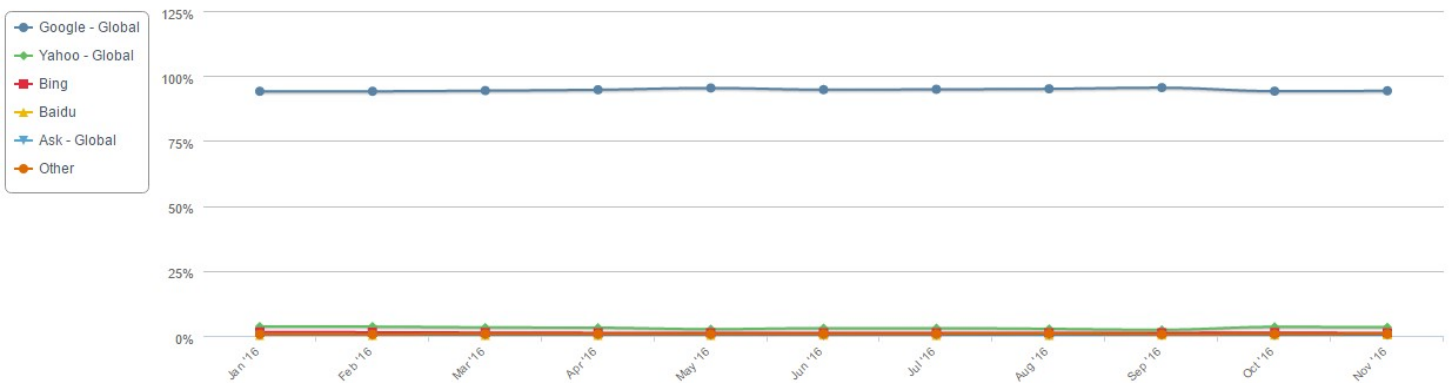


Рисунок 10. Динамика использования поисковых систем на мобильных устройствах за 2016 год

С 2000 года Google является самой крупной поисковой системой в мире. Однако не все страны и континенты пользуются одинаковыми поисковиками. Так, в странах Восточной Азии Google не в фаворитах.

В Китае популярны поисковики Soso и Baidu. Причём, последняя поисковая система ворвалась в десятку сайтов, лидеров по посещаемости, и продолжает там находится по сегодняшний день. Baidu — 3-ой поисковый сайт в мире по посещаемости.

В Тайване и Японии используют Yahoo! Taiwan и Yahoo! Japan.

В Южной Корее большинство жителей пользуются «отечественной» разработкой Naver.

В странах Ближнего Востока существуют поисковые системы, выдающие только «дозволенную» информацию с точки зрения религии. Это либо такие «молодые» системы, как Halalgoogling, либо уже знакомые нам Yahoo!, Google и Bing с обусловленной системой фильтрации.

Конечно, эти списки не являются окончательными, так как разные источники на основе своих критериев оценки формируют перечни популярных поисковиков, включая такие порталы, как: Infoseek, HotBot, Teoma, Exite, Galaxy, Microsoft MSN, AltaVista и др.

16 год.

2.2 Популярные поисковые системы Рунета

По данным на декабрь 2016 года, доли поисковых систем в Рунете по данным Liveinternet (рисунок 11) [7]:

- Яндекс - 48,40%
- Google - 45,10%
- Search.Mail.ru - 5,70%
- Rambler - 0,40%
- Bing - 0,30%
- Yahoo - 0,10%

В статистике учитываются персональные компьютеры и мобильные устройства вместе.

Доля (%)

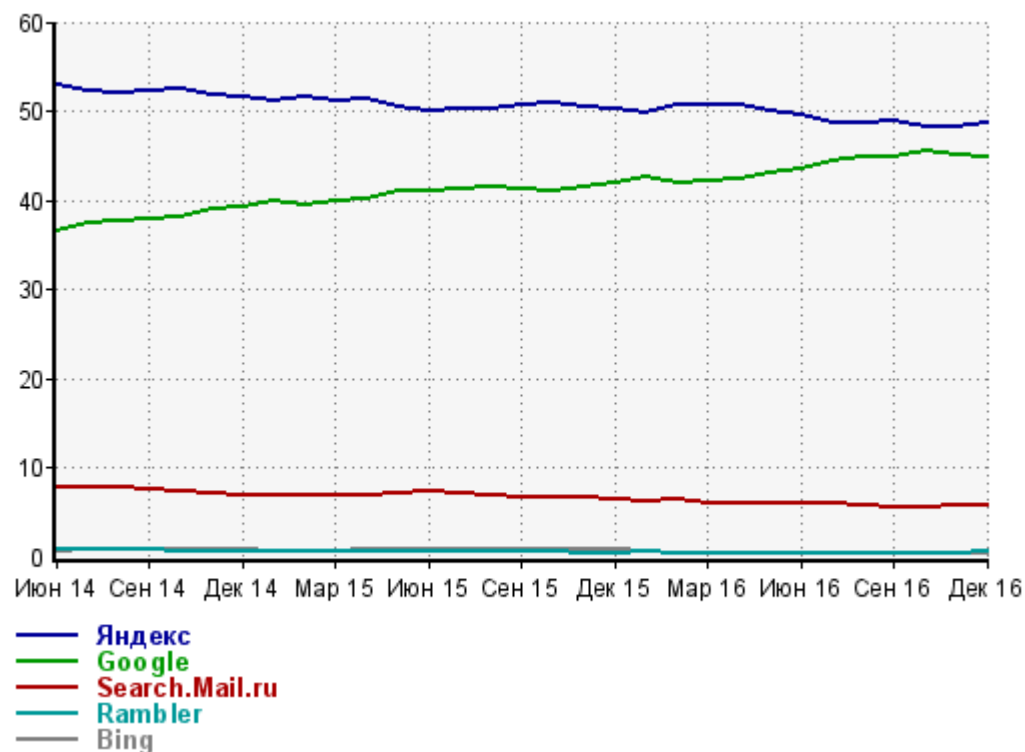


Рисунок 11. Доля поисковых систем Рунета (на декабрь 2016года)

Таким образом, Яндекс это одна из немногих поисковых систем, которая устояла под экспансией Google и не потеряла своих позиций на отечественном рынке. Хотя, следует отметить, что доля использования Google возросла и практически приблизилась к Яндекс.

Наблюдая данную тенденцию, стоит сказать, что Google забирает огромную долю в России благодаря мобильным устройствам, Яндекс пытается противостоять, но пока не очень успешно (рисунок 12).

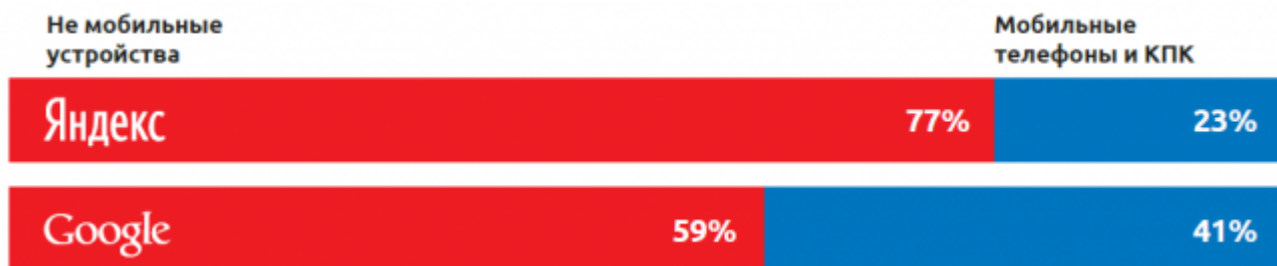


Рисунок 12. Использование Яндекс и Google

В пятерку лучших попали еще две российские компании – Mail и Rambler. Они занимают 3 и 4 места соответственно.

Вот расширенный перечень наиболее распространенных русскоязычных поисковых систем (рисунок 13).

<< Ноя 16		декабрь 2016 г.					
отчет: переходы из поисковых систем		по дням по неделям по месяцам					
значения:		декабрь 2016 г.		ноябрь 2016 г.		в среднем за 3 месяца	
среднесуточные / суммарные							
<input checked="" type="checkbox"/>	Яндекс	98,703,569	48.7%	98,535,628	48.3%	97,544,434	48.3%
<input checked="" type="checkbox"/>	Google	90,509,817	44.6%	92,113,817	45.1%	91,093,102	45.1%
<input checked="" type="checkbox"/>	Search.Mail.ru	11,677,604	5.8%	11,774,194	5.8%	11,490,203	5.7%
<input checked="" type="checkbox"/>	Rambler	1,029,342	0.5%	736,876	0.4%	712,079	0.4%
<input checked="" type="checkbox"/>	Bing	520,024	0.3%	524,236	0.3%	522,690	0.3%
<input type="checkbox"/>	Yahoo	166,662	0.1%	170,403	0.1%	172,576	0.1%
<input type="checkbox"/>	Google (картинки)	71,003	0.0%	69,970	0.0%	68,623	0.0%
<input type="checkbox"/>	Nigma	59,462	0.0%	60,895	0.0%	59,276	0.0%
<input type="checkbox"/>	Tut.by	29,870	0.0%	30,288	0.0%	29,731	0.0%
<input type="checkbox"/>	QIP.ru	28,217	0.0%	26,374	0.0%	28,664	0.0%
<input type="checkbox"/>	сумма выбранных	202,440,359	99.8%	203,684,754	99.8%	201,362,510	99.8%
<input type="checkbox"/>	всего	202,833,686		204,076,886		201,756,760	

Рисунок 13. Перечень наиболее распространенных русскоязычных поисковых систем

Как видно, с огромным отрывом в Рунете лидируют Яндекс и Google, существенно отстает поиск от mail.ru, остальные системы имеют незначительные объемы трафика. ТОП 10 поисковых систем Рунета собирают 99,8% всего трафика.

При этом, следует учитывать, что одна и та же поисковая машина может давать лучше результаты по одним категориям запросов и хуже по другим.

Например, для русскоязычной части интернета, весь коммерческий трафик собирает Яндекс. Он подходит как для совершения покупок, так и продажи товаров или оказания услуг.

Другой пример – Нигма – поисковик, который позиционирует себя как система для студентов и школьников. Встроенные в него модули помогают решать задачи по математике и химии, есть немало и других возможностей (поиск по торрентам, музыке и др.).

Рассмотрит также график динамики изменения количества запросов, обрабатываемых основными поисковыми системами (рисунок 13):

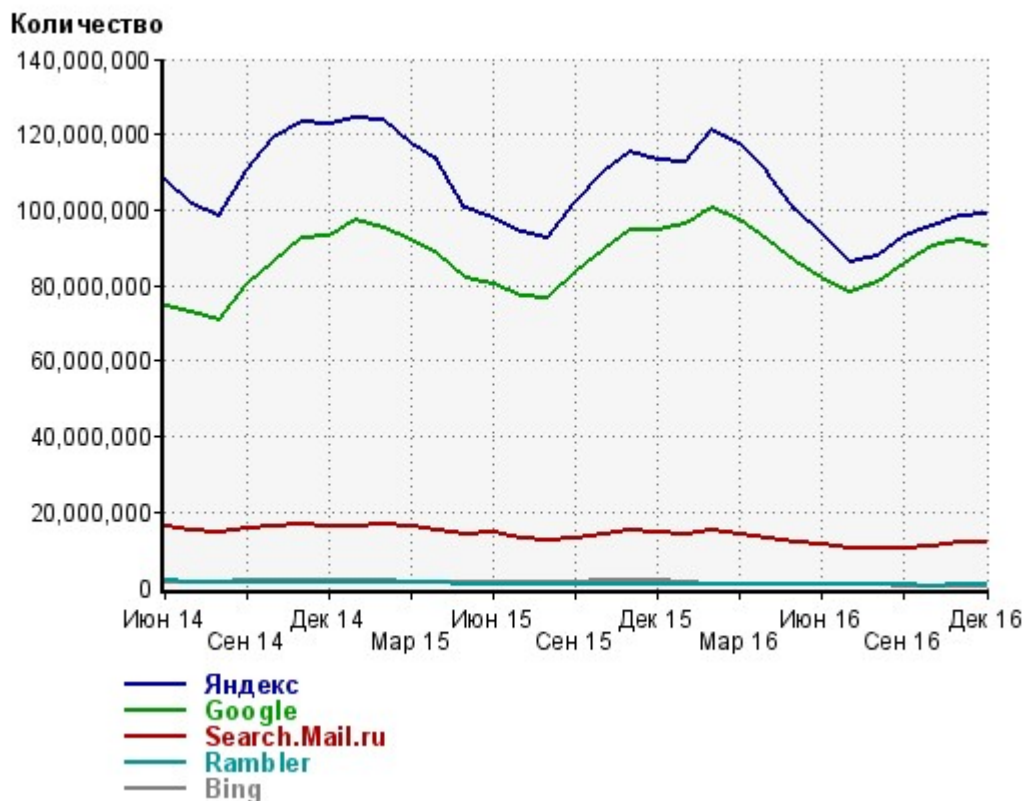


Рисунок 13. Динамика изменения количества запросов, обрабатываемых основными поисковыми системами

Из рисунка видно, что также безоговорочными лидерами являются Яндекс и Google, и разница между ними становится менее заметной.

Следует отметить, что последнее время Яндекс активно выходит на международный уровень и уже имеет локализованные версии сервиса в Украине, Казахстане, Беларуси и Турции.

2.3 Особенности популярных поисковых систем

Рассмотрим некоторые особенности популярных поисковых систем [6]:

Yandex

Самый популярный поисковик России. Большинство людей считает его лучшим благодаря множеству сервисов (Яндекс почта, Яндекс деньги, Яндекс карты и пр.),

наличию коммерческих модулей (Яндекс Маркет, Яндекс Директ), строгой системе модерации рекламных площадок, жесткому контролю за рекламодателями и многим другим особенностям – Яндекс идеально подходит для работы с коммерческими запросами. А геотаргетинг, который учитывает местоположение пользователя с точностью до населенного пункта, превосходит всех конкурентов.

Учет сотен факторов, включая поведение пользователей на страницах сайтов, делает не менее качественным и информационный поиск.

Модуль Яндекс Маркет, вообще, можно считать отдельным поисковиком по интернет магазинам, так как сам он ничего не продает, а объединяет в каталог множество отдельных проектов с их товарами.

Также, среди нововведений Яндекса можно отметить:

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

Google

Крупнейший поисковик в мире. Лишь некоторые страны могут похвастаться наличием реальных местных конкурентов, так что рынок Google огромен. Пока в России этот сервис в роли догоняющего, но отставание от лидера сокращается очень быстро, в основном, благодаря мощной монополии на мобильный поиск (все устройства на ОС Android используют Google поиск по умолчанию).

В коммерческом плане Google значительно уступает Яндексу, так как геотаргетинг у него работает только в масштабах страны.

Между тем, информационный поиск здесь реализован прекрасно, особенно новостной. Google быстрее индексирует новые статьи на сайтах (они быстрее попадают в поиск). Один из самых удобных поисков картинок и фотографий.

Mail

Данный проект в большей степени ориентирован на развлекательные сервисы, поэтому, do.mail.ru (собственно поисковик) не является его главным модулем. По качеству он сравним со старшими конкурентами, но качественных особенностей (которые бы помогли существенно увеличить долю на рынке), конкурирующих с Яндекс и Google, у него нет.

Nigma

Популярность поисковика очень мала, так как это бесплатный проект, создаваемый на общественных началах студентами и преподавателями МГУ. Коммерческой цели он не преследует, соответственно и в раскрутку не вкладывает.

Основное назначение поисковика – это решение задач по химии и математике, поиск по аудио файлам, торрентам, книгам. Кроме того, имеется возможность уточнять запросы через систему фильтров, получая идеальные результаты выдачи.

Rambler

Один из старейших российских поисковиков, появился раньше Яндекса. Но, в какой-то момент проиграл битву за рынок, не выдержав конкуренции. Сейчас Рамблер больше напоминает информационный портал со множеством сервисов (включая электронную почту).

Для поиска Ramler.ru использует технологии Яндекса, поэтому говорит о том, что это самостоятельный поисковый алгоритм не приходится.

Bing

Пожалуй, единственный из западных поисковиков, который хоть как-то известен в русскоязычной части интернета, да и то, благодаря тому, что это детище Microsoft (разработчика Windows), который интегрирует его во все свои продукты.

В англоязычном интернете качество его поиска на уровне, для нас он недостаточно адаптирован и не составляет конкуренцию.

Рассмотрим некоторые правила пользования поисковыми системами. Почти любая поисковая система способна давать качественные ответы, главное уметь у нее спрашивать. К сожалению, машины плохо понимают человеческую речь, особенно вырванную из контекста, поэтому нам приходится опускаться до их уровня, чтобы быстро найти желаемое.

Существует определенный синтаксис поисковых запросов, помогающий в этом. Основной синтаксис у всех поисковиков совпадает (бывают исключения).

Наиболее популярные операторы:

«!» – восклицательный знак перед словом – поиск слова в точной форме (без словоформ и падежей);

«+» — знак плюс перед словом – плюсы ставятся перед словами во фразе, которые обязательно должны присутствовать в найденных документах;

«» — фраза в кавычках – в результатах будут только страницы, содержащие слова фразы подряд (цитата). Если неизвестно или неважно какое-либо слово в цитате, то вместо него можно поставить звездочку (*);

«&» — поиск по нескольким словам, которые обязательно должны встречаться в одном предложении. Двойной знак && — в одном документе;

«|» — вертикальная черта – поиск по любому из слов, разделенных чертой (например, синонимами);

«-» — знак минус перед словом – указанное слово не должно присутствовать в документе.

Это основные операторы, которых будет достаточно для эффективной работы.

ЗАКЛЮЧЕНИЕ

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

С момента появления системы поиска огромное количество поисковых систем совершенствуют свои алгоритмы с целью улучшения качества выдачи, чтобы, именно их результаты были лучшими и люди выбирали их поисковик для работы в интернете. Общие принципы работы у всех поисковиков схожи, различается лишь степень учета различных факторов.

Напрямую оценивать качество получается не всегда, так как этот параметр имеет субъективный характер. Более объективную информацию дает статистика

посещений. Согласно ей, самыми популярными поисковыми система являются Google (мировой лидер) и Яндекс (лидер Рунета).

Яндекс отлично определяет города, регионы пользователя, тогда как Google не настолько уверенно это делает. Основная аудитория Google: ИТ сообщество - программисты, системные администраторы. Google хорошо заточен для поиска всякого рода документации.

Яндекс уверенно держится на первом месте Рунета, Google растёт и забирает рынок. За 2016 год статистика говорит: -5,6% у Яндекса, и около +7% у Google в России. Из-за увеличения роста мобильных устройств использование Google будет увеличиваться, так как популярность данной поисковой системы в мобильных устройствах выше, чем у того же Яндекса.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. История развития поисковых систем [Электронный ресурс]. URL: <http://sirblond.crimea.com/issledovaniya/istoriya-razvitiya-poiskovyx-sistem.html> (Дата обращения: 03.01.2017).
2. Поисковые системы [Электронный ресурс]. URL: http://uniofweb.ru/wiki/poiskovye_sistemy/#full_version (Дата обращения: 30.12.2016).
3. Поисковые системы Интернета: Яндекс, Google, Rambler, Yahoo. Состав, функции, принцип работы [Электронный ресурс]. URL: <https://www.seonews.ru/masterclasses/poiskovye-sistemy-interneta-yandeks-google-rambler-yahoo-sostav-funktsii/> (Дата обращения: 02.01.2017).
4. Рейтинг поисковых систем [Электронный ресурс]. URL: <http://www.gs.seo-auditor.com.ru/sep/> (Дата обращения: 05.01.2017).
5. Рейтинг топ 5 самых лучших отечественных и мировых поисковых систем [Электронный ресурс]. URL: <http://www.bestseoblog.ru/rejting-poiskovyx-sistem/> (Дата обращения: 02.01.2017).
6. Поисковики интернета – рейтинг, критерии оценки, описание [Электронный ресурс]. URL: <http://biznessystem.ru/2016/05/poiskoviki-interneta-rejting-kriterii-ocenki-opisanie/> (Дата обращения: 05.01.2017)
7. <http://www.liveinternet.ru/stat/ru/searches.html?period=month>
8. <https://www.netmarketshare.com>