

Содержание:

ВВЕДЕНИЕ

Поисковая система - программно-аппаратный комплекс с веб-интерфейсом, предоставляющий возможность поиска информации в Интернете. Под поисковой системой обычно подразумевается сайт, на котором размещён интерфейс (фронт-энд) системы. Программной частью поисковой системы является поисковая машина (поисковый движок) — комплекс программ, обеспечивающий функциональность поисковой системы и обычно являющийся коммерческой тайной компании-разработчика поисковой системы.

Большинство поисковых систем ищут информацию на сайтах Всемирной паутины, но существуют также системы, способные искать файлы на FTP-серверах, товары в интернет-магазинах, а также информацию в группах новостей Usenet.

Актуальность темы. Поиск – один из неперенных атрибутов нашей жизни. Процесс бесконечного поиска, путешествия по ссылкам с сайта на сайт – занятие весьма и весьма захватывающее. Такое развлечение имеет массу поклонников и даже получило свое собственное название – веб-серфинг.

Целью настоящей работы является знакомство с поисковыми системами Интернета.

Задачи работы: рассмотреть понятие и сущность поисковой системы, ознакомиться с первыми поисковыми системами Интернета, дать характеристику основным популярным поисковым системам.

Предмет исследования: поисковая система как программно-аппаратный комплекс.

Объект исследования: современные поисковые системы интернета.

1.

ПОНЯТИЕ И СУЩНОСТЬ ПОИСКОВОЙ СИСТЕМЫ

1.1 Общие принципы работы поисковых систем в сети Интернет

Поисковые системы - это сервисы, предназначенные для поиска информации в мировой сети Интернет. В базе данных поисковых систем находится информация, практически, обо всех сайтах сети Интернет и эта база постоянно обновляется автоматизированными системами, следящими за появлением новых сайтов.

Существует четыре типа поисковых систем: с поисковыми роботами, управляемые человеком, гибридные и мета-системы.

- системы, использующие поисковые роботы

Состоят из трёх частей: краулер («бот», «робот» или «паук»), индекс и программное обеспечение поисковой системы. Краулер нужен для обхода сети и создания списков веб-страниц. Индекс - большой архив копий веб-страниц. Цель программного обеспечения - оценивать результаты поиска. Благодаря тому, что поисковый робот в этом механизме постоянно исследует сеть, информация в большей степени актуальна. Большинство современных поисковых систем являются системами данного типа.

- системы, управляемые человеком (каталоги ресурсов)

Эти поисковые системы получают списки веб-страниц. Каталог содержит адрес, заголовок и краткое описание сайта. Каталог ресурсов ищет результаты только из описаний страницы, представленных ему веб-мастерами. Достоинство каталогов в том, что все ресурсы проверяются вручную, следовательно, и качество контента будет лучше по сравнению с результатами, полученными системой первого типа автоматически. Но есть и недостаток — обновление данных каталогов выполняется вручную и может существенно отставать от реального положения дел.

Ранжирование страниц не может мгновенно меняться. В качестве примеров таких систем можно привести каталог Yahoo[en], dmoz и Galaxy.

- гибридные системы

Такие поисковые системы, как Yahoo, Google, MSN, сочетают в себе функции систем, использующие поисковых роботов, и систем, управляемых человеком.

- мета-системы

Метапоисковые системы объединяют и ранжируют результаты сразу нескольких поисковиков. Эти поисковые системы были полезны, когда у каждой поисковой системы был уникальный индекс, и поисковые системы были менее «умными». Поскольку сейчас поиск намного улучшился, потребность в них уменьшилась. Примеры: MetaCrawler[en] и MSN Search.

Одним из первых инструментов поиска в интернете (до Всемирной паутины) был Archie.

Первой поисковой системой для Всемирной паутины был «Wandex», уже не существующий индекс, который создавал «World Wide Web Wanderer» — бот, разработанный Мэтью Грэм (англ. Matthew Gray) из Массачусетского технологического института в 1993. Также в 1993 году появилась поисковая система «Aliweb», работающая до сих пор. Первой полнотекстовой (т. н. «crawler-based», то есть индексирующей ресурсы при помощи робота) поисковой системой стала «WebCrawler», запущенная в 1994. В отличие от своих предшественников, она позволяла пользователям искать по любым ключевым словам на любой веб-странице — с тех пор это стало стандартом во всех основных поисковых системах. Кроме того, это был первый поисковик, о котором было известно в широких кругах. В 1994 был запущен «Lycos», разработанный в университете Карнеги Мелона.

Вскоре появилось множество других конкурирующих поисковых машин, таких как «Excite», «Infoseek», «Inktomi», «Northern Light» и «AltaVista». В некотором смысле они конкурировали с популярными интернет-каталогами, такими, как «Yahoo!». Позже каталоги соединились или добавили к себе поисковые машины, чтобы увеличить функциональность. В 1996 году русскоязычным пользователям интернета стало доступно морфологическое расширение к поисковой машине Altavista и оригинальные российские поисковые машины «Рамблер» и «Апорт».

23 сентября 1997 была открыта поисковая машина Яндекс.

В последнее время завоёвывает всё большую популярность практика применения методов кластерного анализа и поиска по метаданным. Из международных машин такого плана наибольшую известность получила «Clusty» компании Vivísimo. В 2005 году на российских просторах при поддержке МГУ запущен поисковик «Нигма», поддерживающий автоматическую кластеризацию. В 2006 году открылась российская метамашина Quintura, предлагающая визуальную кластеризацию в виде облака тегов. «Нигма» тоже экспериментировала с визуальной кластеризацией.

Помимо поисковых машин для Всемирной паутины, существовали и поисковики для других протоколов, такие как Archie для поиска по анонимным FTP-серверам и «Veronica» для поиска в Gopher.

Пополнение информационных ресурсов Интернета происходит высокими темпами, и найти необходимую информацию становится всё труднее. Различные печатные справочники устаревают ещё до выхода в свет. Единственным надёжным способом поиска информации является использование различных поисковых систем, которые постоянно отслеживают изменение информации в сети.

За время существования Интернета предпринимались различные попытки организации поисковых средств. Многие из этих попыток оказались неудачными, другие же привели к созданию удобных средств поиска информации. Наиболее удачные проекты появились в последние пять лет. Мы рассмотрим поиск информации во Всемирной паутине с помощью нескольких наиболее распространённых систем поиска. Всего же в мире существуют сотни различных поисковых систем, и выбор той или иной системы зависит только от ваших личных пристрастий. Отметим, что поисковые системы часто называют поисковыми машинами или машинами поиска.

Многие поисковые системы позволяют искать информацию не только в Web-страницах, но и в группах новостей и хранилищах файлов. Таким образом, в результате поиска вы можете найти сообщение в группе новостей или какой-то файл. Поэтому чаще применяют вместо термина страница более общий термин – документ. Под документом подразумеваются Web-страница, сообщение или файл, содержащие различную информацию.

Поисковые системы обычно состоят из трех компонентов:

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

Несмотря на то, что поисковые системы запрограммированы, чтобы оценивать веб-сайты на основе некоторой комбинации их популярности и релевантности, в реальности экспериментальные исследования указывают на то, что различные

политические, экономические и социальные факторы оказывают влияние на поисковую выдачу.

Такая предвзятость может быть прямым результатом экономических и коммерческих процессов: компании, которые рекламируются в поисковой системе, могут стать более популярными в результатах обычного поиска в ней. Удаление результатов поиска, не соответствующих местным законам, является примером влияния политических процессов. Например, Google не будет отображать некоторые неонацистские веб-сайты во Франции и Германии, где отрицание Холокоста незаконно.

Предвзятость может также быть следствием социальных процессов, поскольку алгоритмы поисковых систем часто разрабатываются, чтобы исключить неформатные точки зрения в пользу более «популярных» результатов. Алгоритмы индексации главных поисковых систем отдают приоритет американским сайтам.

Поисковая бомба - один из примеров попытки управления результатами поиска по политическим, социальным или коммерческим причинам.

- 1.

Популярность поисковых систем у пользователей Рунета

Статистика поисковых систем является важнейшим фактором, определяющим стратегию продвижения сайта. Действительно, важно знать, аудиторию пользователей поисковых систем, тем самым понимая, продвижение в какой поисковой системе (Яндекс, Google, Bing и т.д.) более «выгодно», принесет больше посетителей, конвертаций, прибыли. Неверно выбранная стратегия продвижения может ощутимо повлиять на финансовые затраты при продвижении проектов.

Сегодня сырые данные берутся из открытого доступа статистики Liveinternet. К сожалению, число сайтов, размещающих счетчик LI постепенно снижается. Так в 2016 году были данные на 200 млн. уникальных пользователей ежедневно, а в текущем, 2019 году уже на 100 млн. уникальных пользователей. Но тем не менее, данные статистики достаточны.

В статье «Статистика поисковых систем в 2019 году[\[1\]](#)» было проведено исследование долей поисковых систем в генерации трафика на сайты

русскоязычного сегмента в нескольких разрезах.

На [рисунке 1](#) приведены данные о количестве среднесуточных переходов за август месяц каждого года всех русскоговорящих пользователей на сайты с установленной статистикой LI из поисковых систем Яндекс и Google. Эти данные не учитывают региональную принадлежность посетителей.

Как видно из [рисунка 1](#), доля русскоязычных пользователей Google ежегодно растет, на август 2019 года составляет 56,4% и значительно превышает долю Яндекса в 41%. Это обусловлено тем, что в русскоговорящих странах ближнего зарубежья, Яндекс не используется как поисковая система.

В целом, в среде русскоговорящих пользователей более популярным является Google, при этом его популярность возрастает.

Давайте посмотрим, какой поисковой системой пользуются чаще в России. Сначала воспользуемся данными LI с ограничением - регион: «Россия» ([рисунком 2](#)).

Если еще в 2018 году Яндекс был более популярен среди пользователей России, то уже в текущем, 2019 году 50,7% россиян использовали Google в качестве поисковой системы. Доля Яндекса в этом году составила всего 46,6% процентов.

Благодаря данным LI, можно отследить соотношение пользователей Яндекса и Google в России по регионам ([рисунком 3](#)).

Как это не странно, среди пользователей Москвы лидирует Яндекс. Доля использования этой поисковой системы составляет 49,7%, тогда как Google отстает на 2% с долей пользователей в 47,7%.

В Питере, как более прогрессивном городе, уверенно лидирует Google. Их доля составляет 53,2%, в то время как на Яндекс приходится всего 44,6%.

Далее приведем распределение пользователей в субъектах России (первая доля - Google, вторая - Яндекс):

- Казань: 56,1% / 40,6%
- Екатеринбург: 52,9% / 44,4%
- Калининград: 62,1% / 32,3%
- Кемерово: 45,6% / 49,8%
- Новосибирск: 57,8% / 40,7%
- Уфа: 50,2% / 46,5%

- Владивосток: 67,5% / 30,3%

Таким образом, чем дальше от первопрестольной, тем выше доля использования Google и ниже - Яндекса. Удивительным образом из этой статистики выбивается Кемерово. Там в единственном из регионов доля пользователей Яндекса превышает пользователей Google.

Такова статистика запросов Яндекс и Гугл в России на август 2019 года.

Интересно так же проследить мировые тенденции использования поисковых систем[2].

Google является лидером по всем возможным опросам и занимает однозначное первое место во всем мире. Более 70% интернет-пользователей выбирает именно Google. Столь высоких показателей удастся достичь и благодаря тому, что во многих мобильных браузерах именно Google установлена как ПС по умолчанию. Выходя из популярности, можно считать, что это лучшая поисковая система Интернета. Кто-то этот факт подтверждает, кто-то опровергает, но цифры говорят сами за себя.

Bing – продукт компании Microsoft, следовательно, это обязательная составляющая смартфона, работающего на операционной системе Windows. В процентном соотношении больше всего пользователей ПС Бинг в США (31%), Китай (18%), Германия (6%), Украина (0,7%).

Yahoo! В топ 5 поисковых систем вошел и самый старший поисковик. Тут аналогично больше всего пользователей из США. Помимо строки поиска также есть региональный прогноз погоды и строка с самыми популярными новостными сводками.

Baidu.com - поисковая система, используемая в Китае. Учитывая густоту и число жителей страны, ПС удалось занять третье место, если рассматривать рейтинг поисковиков на 2019 год. Интересно, что ПС имеет свою встроенную энциклопедию, которая среди жителей Китая пользуется большей популярностью, нежели известная всем нам Википедия.

Yandex – самый популярный поисковик в России и других странах СНГ. Отличительной чертой является наличие богатого функционала, который включает музыку, карты, транспорт, почту, переводчик и др.

Search.aol.com – пользовалась популярностью на территории США в 90-х годах. Сегодня остается некоторый процент пользователей, которые не любят менять свои привычки и ищут информацию привычным для них поисковиком.

Ask.com – интересный формат поисковой системы, в которой каждый запрос воспринимается, как вопрос. Ответ предоставляется в форме статьи на соответствующую тему. То есть тут не будет ссылок на форумы, а исключительно на сайты с тематическими статьями.

1.

ХАРАКТЕРИСТИКА ОСНОВНЫХ ПОИСКОВЫХ СИСТЕМ

2.1 Google

Google - самая популярная в мире поисковая система. Наглая, как и все американцы, но предоставляющая много возможностей. В общем, бесспорный мировой лидер среди поисковых роботов.[\[3\]](#)

Помимо поиска у Google имеется множество других сервисов различной направленности.

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

Поисковый робот Google имеет User Agent - Googlebot (Поисковый робот), который является основным роботом, сканирующим содержание страницы для поискового индекса. Помимо него существуют ещё несколько специализированных роботов:

- Googlebot-Mobile — робот, индексирующий сайты для мобильных устройств, ноутбуков
- Google Search Appliance (Google) gsa-crawler[17] — поисковый робот нового аппаратно-программного комплекса Search Appliance,
- Googlebot-Image — робот, сканирующий страницы для индекса картинок,
- Mediapartners-Google — робот, сканирующий контент страницы для определения содержания AdSense,

- Adsbot-Google — робот, сканирующий контент для оценки качества целевых страниц Ads.

Также эти роботы помогают определить страницы низкого качества. Критерии определения низкокачественных страниц:

- Уникальность контента
- Поведение пользователя
- Орфографические ошибки
- Ссылки на страницу
- Интуитивно понятный и удобный дизайн

Google учитывает персональные данные пользователей для выдачи только тех результатов, которые ему нужны. Это явление имеет негативные стороны, например, если пользователь критикует президента в социальных сетях, то на запрос о гражданском мнении о президенте он получит только результаты с негативными характеристиками последнего, вообще на любой запрос пользователь будет получать не противоречащую его мнению информацию, что разумеется не очень хорошо влияет на его гражданское мнение. Илай Парайзер назвал это пузырьём фильтров.

Сейчас в Google возможна функция отключения просмотра персональных результатов при просмотре сайтов.

Как правило, самой привлекательной стороной Google является огромный трафик. Недостатком является то, что все рекламодатели и владельцы сайтов хотят заполучить этот трафик. Именно поэтому наиболее конкурентоспособным является органический поиск, а платный поиск стоит дороже, чем на других сайтах.

Кроме того, в последнее время Google начал активно внедрять новую политику: теперь пользователям вовсе не обязательно переходить на сторонние сайты, чтобы получить ответы на свои вопросы. Вся информация появляется прямо на сайте Google. Конкурировать становится сложнее, а шансов получить хороший трафик при этом становится все меньше.

- 1.

Яндекс

Из поисковых систем Яндекс, пожалуй, пользуется наибольшей популярностью среди русскоязычных посетителей Интернета.

«Яндекс» — российская ИТ-компания, владеющая одноимённой системой поиска в Сети и интернет-порталом. Поисковая система «Яндекс» является 5-ой среди поисковых сайтов мира по количеству обработанных поисковых запросов (более 3 млрд, 1,7% от мирового количества, статистика за сентябрь 2011 года).

Яндекс — отечественная поисковая система. Это одно из её немногих достоинств. В остальном имеем медленное обновление поисковой базы, постоянную смену правил игры (как для пользователей, так и для владельцев сайтов), своеобразное понимание качества поисковой выдачи и т.д. В общем, она популярна в Рунете лишь потому, что большинство пользователей не догадывается о существовании других ищалоков.

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.

Компания вышла на самоокупаемость в 2002 году. В 2010 году выручка «Яндекса» составила \$439,7 млн (рост на 43 %), операционная прибыль — \$169,7 млн (рост — 82,8 %), чистая прибыль — \$134,3 млн (рост — 89,9 %). В мае 2015 г. Яндекс провёл первичное размещение акций, заработав на этом больше, чем кто-либо из Интернет-компаний со времён IPO поисковика Google в 2004 году

Основным и приоритетным направлением компании является разработка поискового механизма, но за годы работы «Яндекс» стал мультипорталом. В 2016 году «Яндекс» предоставляет более 30 сервисов. Самыми популярными являются: Яндекс.Картинки, Яндекс.Почта, Яндекс.Карты, Яндекс.Новости, Яндекс.Погода и другие.

Здесь есть список сервисов Яндекса для решения других проблем.

Отличительная особенность Яндекса — возможность точной настройки поискового запроса. Это реализовано за счёт гибкого языка запросов.

- 1.

Bing

В 2009 году Bing заменил MSN Search. Это был своеобразный ответ компании Microsoft на растущую популярность Google. С того момента популярность Bing выросла с 8,4% до 28,1%. В 2016 году Bing добавил специальные опции по типу AOL на все сайты, для которых они предоставляют результаты поиска[4].

Всевозможные дополнения и обновления сделали Bing реальным соперником Google. Согласно последним данным, в настоящее время около 33% поисковых запросов в США делаются именно в поисковой системе Bing.

В последнее время Bing активно занимается улучшением сервиса Bing Ads. Было добавлено множество новых опций, нацеленных на упрощение взаимодействия с рекламодателями, которые привыкли работать в Google AdWords.

Несмотря на то, что у Bing даже приблизительно нет такого трафика, как у Google, эта поисковая система пользуется большой популярностью во многих странах, включая США.

Алгоритм работы Bing намного проще, чем алгоритм работы Google, поэтому разобраться с настройками этой системы гораздо легче.

Из-за того, что трафик Bing меньше, конкуренция среди рекламодателей также существенно меньше, поэтому, если вы грамотно подойдете к SEO-продвижению, у вас будет больше шансов привлечь новых посетителей на свой сайт.

Помимо этого, Bing Ads устроен достаточно проще и удобнее. Более низкий трафик сайта легко компенсируется очень низкой ценой клика.

- 1.

Rambler

Rambler Media Group (AIM: RMG) — интернет-холдинг, включающий в качестве сервисов поисковую систему, рейтинг-классификатор ресурсов российского Интернета, информационный интернет-портал.

Поисковая система Рамблер понимает и различает слова русского, английского и украинского языков. По умолчанию поиск ведётся по всем формам слова. По умолчанию в Рамблере результаты ранжируются по степени соответствия (релевантность) запросу и группируются по сайтам.

Настройки языка поиска: любой, русский, английский, украинский, казахский, белорусский, татарский, французский, немецкий. Учитывается морфология

Поисковая система содержит информацию о более чем 12 миллионах документов, расположенных на серверах России и стран СНГ. Рамблер обрабатывает ежедневно не менее 500 тысяч поисковых запросов (в среднем - 5 запросов в секунду), сканируя 48 тысяч web-серверов и используя несколько одновременно работающих программ-роботов.

Запрос может состоять из одного или нескольких слов, разделенных пробелами. Могут быть использованы как русские, так и английские слова, и словосочетания. По умолчанию находятся только те документы, в которых встретились все введенные Вами слова. Чтобы найти документы, содержащие хотя бы одно слово из запроса, используйте логическую связку Or (см. ниже) или выберите на странице детального запроса: "Слова запроса: любое". Чтобы исключить документы, содержащие те или иные слова, укажите на странице детального запроса: "Исключить документы, содержащие следующие слова ...".

Все равно, с какой буквы написаны слова запроса: с большой или с маленькой. И при построении индекса, и при поиске по запросу все заглавные (большие) буквы "понижаются".

Слова запроса могут быть соединены логическими связками And ("и"), Or ("или"). Вместо связок (или в сочетании с ними) могут использоваться также символы '&', '|'.

Части запроса могут быть сгруппированы с помощью круглых скобок (). Возможна многократная вложенность скобок в сочетании с логическими операторами.

Рамблер умеет искать слова во всех формах (например, аминокислота, аминокислоты, аминокислотой и т. д.). Чтобы слово находилось во всех формах, перед ним надо поставить служебный символ '#'. В меню детального запроса такой режим может быть включен для всех слов: "Расширение запроса: все формы слов". Служебный символ '@' перед словом позволяет находить не только само это слово, но и однокоренные слова. В меню детального запроса символу '@' соответствует режим "Расширение запроса: все однокоренные".

По умолчанию наша система ищет слова запроса так, как Вы их ввели, чтобы уменьшить "шум" в найденных документах. Если Вы не помните, как пишется слово, или хотите расширить запрос, можно использовать метасимволы '*' и '?' для

обозначения произвольной части слова и произвольного символа.

Поиск в части документа

Ограничить поиск частями документов, такими как название документа, его заголовок, URL и т.п., можно через меню детального запроса "Искать в...".

Язык документов

Можно ограничить поиск документами только на русском или только на английском языке. Для этого надо выбрать соответствующий режим в меню детального запроса "Язык документа...". По умолчанию поиск выполняется по документам на всех языках.

Сортировка результатов

По умолчанию найденные документы сортируются по релевантности (соответствию запросу). Однако Вы можете потребовать, чтобы вместо этого в начало списка были помещены самые свежие (или, наоборот, самые старые документы). Для этого надо выбрать соответствующую установку в меню "Сортировать по..." на странице детального запроса. Также можно ограничить поиск документами, созданными в определенный период времени: для этого необходимо на странице детального запроса указать "От даты ... до даты ...".

Расстояние между словами

Можно потребовать, чтобы Рамблер возвращал только те документы, где слова из запроса находятся на минимальном расстоянии друг от друга. Режим "Ограничить расстояние между словами" может быть включен в детальном запросе. Все перечисленные выше правила могут быть использованы совместно друг с другом в необходимой Вам последовательности.

Выдача результатов

По умолчанию результаты поиска выдаются порциями по 15 документов. Меню "Выдавать по..." на странице детального запроса позволяет увеличить это число до 30 или 50. Меню "Форма вывода..." позволяет получать описания документов с увеличенной или уменьшенной подробностью.

2.5 Aport

Особенностью данной системы является то, что в запросах вы можете использовать русские слова в любой грамматической форме. Интеллектуальный механизм преобразования запросов разберётся с самыми сложными фразами

Впервые была продемонстрирована в феврале 1996 года на пресс-конференции «Агамы» по поводу «Русского клуба». На тот момент поисковая машина искала только по сайту russia.agama.com. В дальнейшем был поиск по четырём серверам, потом по шести. В итоге день рождения Апорта и фактический старт системы сильно «размазались» по времени. Официальная презентация Апорта состоялась только 11 ноября 1997 года, тогда машина уже индексировала весь рунет.

На текущий момент Апорт является частью портала РОЛ. Принадлежит Golden Telecom. При поиске учитываются особенности русского языка.

2.6 Yahoo

Yahoo! (произносится как яху!) — американская компания, владеющая второй по популярности (5.88 %) 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. В 2004 году была запущена новая версия почтового интерфейса, основанная на AJAX.

Компания Yahoo! была основана аспирантами Стэнфордского университета Дэвидом Файло (англ. David Filo) и Джерри Янгом (англ. Jerry Yang) в январе 1994 года; стала корпорацией 2 марта 1995 года. Главный офис компании находится в городе Саннивейл (англ. Sunnyvale), штат Калифорния, США.

Самая известная поисковая машина. Её сайты разбиты по категориям и ключевым словам. Она содержит полезную информацию на своей домашней странице. Может подключаться к другим поисковым машинам

Базы данных: в ведении находится служба поиска Internet-ресурсов, новостей, карт, рекламных информации, спортивная информация, бизнес, номера телефонов, персональные WWW-страницы, и email-адреса (отдельная база данных).

Поиск: Все Yahoo страницы предлагают не только простое поисковое окно, но и опции для этого поиска, а также поиск Usenet или Email-адреса. Поиск может ограничиваться указанием определённого промежутка времени. Boolean операторы

(и, или) и последовательный поиск также поддерживаны. Отметим: если поиск в Yahoo! не привёл к положительному результату, то процесс поиска автоматически переходит на Alta Vista, которая продолжает поиск, и в случае положительных результатов автоматически возвращает найденную информацию в Yahoo!.

Если Yahoo! не может установить связь достаточно быстро с Alta Vista, то в этом случае Yahoo! будет обеспечивать страницу связи с набором инструментов поиска. После того как одна из этих связей выбирается, ключевые слова передаются к поисковой машине на ваше усмотрение.

Средством, облегчающим поиск, является наличие "tip search"(TS) - поиск с помощью "намёка": Yahoo! Является подчиненным справочником, что означает, что система не имеет так много страниц, как поисковые машины, однако задание наиболее общих ключевых слов позволит найти необходимую тему на странице высокого уровня (первая страница, которая возникает перед пользователем при посещении сайта) для организации или компании.

Результаты: Связи отображаются в соответствии с очерёдность задаваемых слов последовательностью поиска наряду с их описательным текстом и подчиненной иерархией.

Адрес: <http://www.yahoo.com/>

Принципиальное отличие поисковой системы Yahoo (Яхо) от других систем в том, что вы можете найти необходимую вам информацию без использования запросов, а просто переходя по ссылкам разделов встроенного справочника и последовательно уточняя область ваших интересов. Yahoo является самой популярной системой в США, так как не требует специальной подготовки для поиска информации. Не надо знать правила формирования запросов, достаточно просто переходить по ссылкам в нужные разделы. Однако для русскоязычных пользователей эта система не слишком удобна, так как разбиение на разделы проводилось с учётом американской специфики. Кроме того, в справочнике Yahoo содержится намного меньше ссылок на русскоязычные документы, чем в базе данных Alta Vista. Однако для поиска информации по конкретной тематике данная система может оказаться достаточно полезной.

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.

Систему Yahoo удобно использовать для поиска спортивных событий, новостей, прогнозов погоды, развлечений и тому подобной информации. Это обусловлено особенностями разбиения информации на разделы в данной системе.

3. НАПРАВЛЕНИЯ СОВЕРШЕНСТВОВАНИЯ ПОИСКОВЫХ СИСТЕМ

3.1 Разработка и внедрение элементов искусственного интеллекта

Искусственный интеллект (artificial intelligence, AI) - это свойство технических или программных систем выполнять творческие функции, которые присущи человеку. Одна из основных задач ИИ - понимание человеческого интеллекта.

Выделяют 3 вида искусственного интеллекта:

- Ограниченный (Artificial Narrow Intelligence, ANI) - ИИ, который создан для решения конкретных задач. Например, угадывать изображения, играть в шахматы.
- Общий (Artificial General Intelligence, AGI) - универсальный искусственный интеллект, который находится на одном уровне с человеческим и способен решать множество разнообразных задач.
- Искусственный суперинтеллект (Artificial Superintelligence, ASI) - сверхинтеллект, который превосходит уровень отдельного человека или всего человечества.

На данный момент все существующие формы искусственного интеллекта - ограниченные. То есть способны решать исключительно конкретные, прикладные задачи и не могут соревноваться с человеческим разумом в универсальности.

Между тем, ИИ уже широко используется. Как указывает Google в своем блоге, корпорация применяет искусственный интеллект в следующих сервисах:

- Переводчике Google - для прямого перевода.
- Почте Gmail - в блокировщике спама и при написании текста писем.
- Голосовом помощнике Google Assistant - для распознавания речи.

- Google Photo - для распознавания изображений.

- YouTube - для рекомендаций видеороликов.

Разработки ИИ на основе глубокого обучения начались еще в 2011 году в проекте Google Brain. Например, в 2012 году нейронная сеть из 16 000 компьютеров научилась распознавать на изображениях котиков.

В России искусственный интеллект использует «Яндекс» - для предсказания погоды, управления беспилотными автомобилями, написания стихов и музыки, распознавания изображений и речи; в голосовом помощнике «Алиса», мультимедийном устройстве «Яндекс.Станция», и, конечно, в алгоритмах поиска.

В целом поисковые системы развивались по мере роста количества документов в интернете. Чем больше их становилось, тем заметнее увеличивалась сложность алгоритмов. Сначала поисковики просто искали страницы, затем решали задачи, а теперь становятся постоянными помощниками.

Можно выделить следующие этапы эволюции поисковых систем:

1. Наивный поиск

Первоначально работал только поиск слов, так называемый инвертированный индекс. Но количество страниц увеличивалось и их стало необходимо ранжировать. Пришлось учитывать частоту слов, их важность в контексте документа с помощью статистической меры $tf-idf$.

2. Ссылочное ранжирование

Когда страниц стало еще больше, к системе ранжирования подключили учет важности страницы в зависимости от качества и количества ссылок на нее - PageRank.

3. Машинное обучение

В поисковых алгоритмах машинное обучение начали использовать в начале 2000-х. В «Яндексе» это была система «Матрикснет», которая обучалась на основе образцов, составленных специальными людьми - ассессорами. В 2017 году «Яндекс» перешел на новую систему машинного обучения - Cat Boost, он дает более точные результаты в задачах ранжирования.

Обе системы используют технологию градиентного бустинга - метод машинного обучения для регрессии и классификации проблем. Метод производит модели прогнозирования в виде дерева решений. Это позволяет обрабатывать разнородные данные, поэтому можно использовать системы сразу в нескольких направлениях: например, в прогнозе погоды, навигаторе, блокировке спама и т. д.

У классического машинного обучения есть определенные ограничения - оно эффективно только при большом количестве данных. Метод отлично работает, когда пользователи запрашивают миллионы и тысячи одинаковых запросов, то есть когда сигнал очень явный и мощный. Однако поиск развивается в сторону уникальных, низкочастотных и многословных запросов.

Искусственный интеллект базируется на достижениях машинного обучения. Разработки в этом направлении велись еще с 2013 года, когда были проведены первые исследования возможностей системы Word2Vec для семантического анализа.

На базе этой программы в Google была создана самообучающаяся система с искусственным интеллектом - Rank Brain. Запуск был утвержден 26 октября 2015 года. Цель алгоритма - уяснить смысл текста, отыскивая связи между отдельными словами.

Rank Brain является частью алгоритма Hummingbird («Колибри») в Google. Когда система встречает незнакомые слова, она ищет подсказки и синонимы по запросу. Найденные аналогии становятся основой для фильтрации данных. На данный момент Rank Brain является одним из трех важнейших критериев оценки страницы, наряду со ссылками и текстом.

В 2016 году «Яндекс» объявил о запуске нового алгоритма «Палех», который работает на основе нейросетей. Алгоритм позволил лучше искать страницы, которые соответствуют запросу не только по ключевым словам, но и по смыслу. «Палех» анализирует заголовки страниц и извлекает из них скрытые семантические связи.

Развитием технологии стал алгоритм «Король», о введении которого объявили 22 августа 2017 года. В отличие от «Палеха», «Король» сравнивает семантические векторы запросов и страниц целиком, а не только заголовков. При этом, помимо нейросетей, здесь задействовано машинное обучение на основе поведения людей. В роли ассессоров выступают миллионы обезличенных пользователей.

Все алгоритмы имеют схожую схему работы и 1 задачу - улучшить понимание сложных и многословных запросов.

Внедрение искусственного интеллекта полностью изменило поисковую выдачу и правила игры SEO.

Преимущества:

- повысилась точность выдачи по редким и низкочастотным запросам - поисковики понимают простой человеческий язык.

- в выдаче теперь лидируют более качественные ресурсы - спам и переоптимизация ключевыми словами теперь приводит лишь к попаданию под фильтры.

Отпала необходимость в SEO-текстах - теперь нужно отталкиваться от нужд пользователя. Чтобы оптимизировать текст под запросы, стоит использовать LSI-копирайтинг.

Искусственный интеллект поисковых систем еще развивается и имеет ряд недоработок:

- Размытые результаты поиска - в случае многозначности смысла, робот не может точно определить нужный контекст и поэтому предлагает сразу несколько вариантов.

- Непрозрачная система ранжирования - пользователь не может уточнить область поиска с помощью подбора словосочетаний, поиск все равно покажет то, что считает нужным.

- Нетематические ресурсы в выдаче - зачастую в выдаче присутствуют сайты, которые не относятся к теме запроса или на них размещен контент низкого качества.

Помимо органической выдачи, искусственный интеллект теперь используется для подбора целевой аудитории и поисковых запросов в контекстной рекламе. В «Яндекс.Директ» эта функция получила название «Автотаргетинг». Кроме того, тестируется автоматическое создание текстов объявлений на основе контента посадочных страниц и сайтов.

Подобный механизм есть в Google AdWords Express и Google AdWords - платформа сама подбирает ключевые запросы и генерирует объявления, исходя из текста и

картинок на сайте. Подобные функции были введены для упрощения работы с рекламными системами.

Более того, ИИ уже выходит в офлайн. Недавно «Яндекс» сообщил о запуске продаж рекламы на цифровых билбордах. Алгоритмы распознавания лиц используются, чтобы таргетировать рекламные сообщения на целевую аудиторию.

Новейший пример использования ИИ в рекламе - платформа AstraOne, которая встраивает рекламу непосредственно в изображения на сайте. При этом она учитывает общий контекст страницы: распознает текст и изображения - это позволяет попасть точно в целевую аудиторию.

Искусственный интеллект уже встроен в механизмы поисковых систем и в будущем его роль будет только расти. Помимо оценки текста, ИИ применяется для распознавания речи, изображений и видео, перевода страниц, оценки качества ссылок и изучения поведения пользователей.

3.2 Расширение области индексирования и эволюция поисковых запросов

Ведущая тенденция последнего времени заключается в расширении числа обследуемых документов и одновременном увеличении глубины просмотра источника. Несколько лет назад поисковые средства были способны индексировать только несколько миллионов источников, которые на тот период включали web-страницы, статьи конференций Usenet, файлы Gopher и FTP серверов. Современные системы в среднем индексируют порядка 50 миллионов документов (от 31 миллиона Alta Vista, до 69 миллионов в Lycos). При этом скорость работы не только не снизилась, но даже увеличилась за счет использования новейшего программного и аппаратного обеспечения.

Однако даже более важным является то, что поисковые механизмы последнего поколения индексируют все слова на web-странице или в статье из конференции, в то время как ранее область индексирования ограничивалась как правило названием, заголовками, первыми несколькими строками и адресом документа. Это существенно ограничивало возможность выявления материалов по узкой тематике, поскольку результаты поиска не всегда отражали реально существующие данные. Устранив этот недостаток, современные поисковые системы стали намного более надежными, чем их предшественники.

В ближайшем будущем полнотекстовое индексирование утвердится в качестве неотъемлемой характеристики поисковых средств, которые собираются серьезно конкурировать. На наш взгляд, это один из ключевых параметров будущего развития поисковых инструментов.

Следующая важнейшая черта - совершенствование внутреннего поискового механизма, выражающееся в увеличении числа операторов и других элементов составления запросов. Несколько лет назад применение находили только два, в лучшем случае, три классических булевых оператора: AND (и), OR (или) и NOT (не). Теперь появились NEAR (рядом, около) в Alta Vista и FOLLOWED BY (следует за) в OpenText - в высшей степени полезные операторы расстояния, которые дают возможность в максимальной степени конкретизировать запрос. Многие системы позволяют усекавать окончания терминов, ограничивать поиск по дате создания документа, искать ключевые слова только в обозначенных элементах web-страниц (названии, заголовках, электронном адресе и т.д.), а также вести разыскание на точное словосочетание. Новейшие разработки также позволяют выявлять файлы определенного вида (например, графические или аудио) и обладают чувствительностью к строчным и заглавным буквам. Общепринятой становится возможность искать данные на любых языках. Все это дает возможность составлять поисковое предписание с большой степенью точности, что конечно же повышает релевантность получаемых результатов.

Немаловажное значение имеет также ясно проявляющаяся тенденция совершенствования интерфейса. Графические браузеры типа Netscape Navigator, Internet Explorer и им подобные практически полностью вытеснили текстовой Lynx, что само по себе позволило в значительной степени повысить дружелюбность интерфейса, сделать его намного более сенситивным чем 2-3 года назад. Разработчики первых поисковых систем не слишком заботились о комфорте пользователей. В тот период гораздо важнее было просто заставить систему работать. Поэтому желающих воспользоваться услугами поисковых серверов встречали зачастую лишь строки для ввода запроса, без каких-либо экранов помощи или примеров употребления булевых операторов. По этой причине пользователи часто испытывали затруднения, пытаясь точно выразить свой запрос. Системы последнего поколения во многих случаях имеют детализированное многоуровневое меню, которое делает составление запроса максимально простым. За счет его применения пользователи освобождаются от необходимости знать принципы работы булевой логики - они просто комбинируют искомые термины с имеющимися условиями поиска. Меню Power search (углубленного поиска) в

OpenText и Modified search (модифицированного поиска) в HotBot яркие примеры прогресса в этой сфере.

Нынешние поисковые средства предоставляют также детализированную «помощь», примеры составления запросов, файлы «часто задаваемых вопросов». В зависимости от полученных результатов генерируются подсказки и рекомендации, направленные на устранение типичных ошибок, допускаемых при составлении запроса.

Таким образом сегодняшний пользователь уже не должен обладать какими-либо специальными знаниями и опытом для обращения к поисковым средствам.

Еще одной немаловажной тенденцией является оптимизация выдачи результатов поиска. В данном случае можно с уверенностью утверждать, что поисковые средства становятся более интеллектуальными. Ныне их блок искусственного интеллекта способен ранжировать результаты в зависимости от соответствия конкретному запросу. Программа анализирует расположение слов в документе, их повторяемость и общее количество. На этой основе перечень ссылок выдается в порядке соответствия запросу, причем ссылки на наиболее ценные источники, выдаются в начале списка[5].

В перспективе работа над тем, как сделать поисковые средства еще более интеллектуальными будет продолжаться, поскольку только системы с мощной интеллектуальной базой будут способны переработать стремительно расширяющийся океан полнотекстовых данных. В частности, именно системам искусственного интеллекта предстоит решить проблему информационного шума или "мусорных данных", которая становится все острее в связи с ростом в Интернет материалов низкой содержательной ценности.

Последняя тенденция - появление дополнительных сервисных функций, которые делают пользование поисковыми системами более привлекательным. Некоторые из навигационных средств формируют добавочные базы данных по определенным направлениям (электронные адреса частных лиц, справочники компаний, перечни последних новостей и т.д.).

Другие размещают на своих страницах ссылки на наиболее популярные справочные ресурсы Интернет, как например электронные словари, тезаурусы, энциклопедии, интерактивные карты и т.п. источники, что превращает интерфейс поисковых систем в хорошую отправную точку для любого, работающего с Интернет. В последние месяцы наиболее продвинутые системы предлагают даже

воспользоваться услугами персонального сетевого информатора, который обеспечивает текущий поиск новых материалов по заявленной тематике.

Помимо технологических перемен, стратегические изменения в сфере поискового сервиса также очевидны. Среди них наиболее заметным является дробление поисковых средств, выделение поисковых инструментов региональной или отраслевой направленности, учитывающих материалы по отдельным странам или на определенном языке (только чисто русскоязычных систем насчитывается не менее десятка) или по определенной тематике (пока это в основном юриспруденция). Эти локальные поисковые средства призваны решить задачу исчерпывающего учета информации по отдельным странам и наиболее актуальным темам. Возможно в скором времени мы будем располагать столькими поисковыми системами, сколько сегодня существует печатных справочников и указателей.

Наряду с образованием местных поисковых серверов, ясно обозначилось также стремление распространения вширь гигантов поискового бизнеса. Широко известные Yahoo!, Alta Vista, InfoSeek форсируют создание как "зеркальных", так и специальных профильных серверов в европейском и азиатском регионах, стремясь захватить там рынок сетевой рекламы. Само по себе это еще более ожесточает конкурентную борьбу в области поискового бизнеса.

Нет сомнений, что все перечисленные тенденции эволюции поисковых средств найдут свое продолжение в ближайшем будущем. С ростом ресурсов Интернет их потенциал будет также расширяться. Продолжится совершенствование уже существующих механизмов (примерами постоянного обновления являются WebCrawler, Lycos и Alta Vista), а также вытеснение старых систем новыми под иными именами (типичный пример HotBot вытеснивший Inktomi). С уверенностью можно прогнозировать и появление совершенно новых средств, созданных на базе последних технологических достижений.

ЗАКЛЮЧЕНИЕ

Сегодня трудно представить себе мир без компьютера, и мало кто задумывается, а что же на самом деле мы называем умными машинами. И уж точно никто не знает насколько стали умными данные аппараты. Для многих людей Искусственный интеллект и компьютер, который стоит на вашем столе это одно и то же. Но как люди, просвещенные мы знаем, что до разума человека, или даже собаки любой самой умной машине еще далеко.

В наше время информация играет огромную роль во всех сферах жизнедеятельности. Пользователям каждый день необходимо взаимодействовать с большими объемами данных, будь то новости, Web-страницы или музыка. И без помощи поисковых систем делать это было бы практически невозможно. Но, благодаря технологическому прогрессу и развитию информационных сетей, поисковые системы становятся удобнее и проще в обращении, облегчая работу с ними как новичкам, так и опытным пользователям.

Технологический прогресс не стоит на месте и человечество совершенствуется, помещая во всемирную паутину терабайты информации, и в будущем без поисковых систем обойтись будет совершенно нереально.

Итак, в первую очередь поисковая система должна предоставлять людям именно ту информацию, которую они ищут.

За время существования Интернета предпринимались различные попытки организации поисковых средств. Многие из этих попыток оказались неудачными, другие же привели к созданию удобных средств поиска информации. Наиболее удачные проекты появились в последние пять лет. Мы рассмотрели наиболее распространенные системы поиска. Всего же в мире существуют сотни различных поисковых систем, и выбор той или иной системы зависит только от личных пристрастий.

Реальной базой для дальнейшего усовершенствования поисковых орудий является расширение деловой активности в Интернет. Предприимчивые компании уже осознали, что создание и поддержка поисковых серверов является доходным бизнесом. Так как к их услугам прибегает огромное число пользователей, интерфейсы систем становятся наиболее выгодным местом для размещения рекламы. Поэтому разработчики поисковых орудий постоянно заботятся о том, чтобы сделать свой продукт максимально привлекательным. Добиться же этого можно только путем неуклонного повышения качества работы систем по всем рассмотренным показателям. На благо всех обитателей киберпространства.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Ашманов И. С., Иванов А. А. Продвижение сайта в поисковых системах - М.: Вильямс, 2016. - 304 с.

Поисковая система	2017	2018	2019
Google	59987682	58969944	56012783
Яндекс	55254426	51037788	40756800
Прочие	6048959	3751971	2590116
Всего	121291067	113759703	99359699
Поисковая система	2017	2018	2019
Google	49,5%	51,8%	56,4%
Яндекс	45,6%	44,9%	41,0%
Прочие	5,0%	3,3%	2,6%
Всего	100,0%	100,0%	100,0%

Рисунок 1. Данные о количестве

Поисковая система	2017	2018	2019
Google	43725137	42588995	41217823
Яндекс	50889452	46970659	37928260
Прочие	4716254	2962262	2206799
Всего	99330843	92521916	81352882
Поисковая система	2017	2018	2019
Google	44,0%	46,0%	50,7%
Яндекс	51,2%	50,8%	46,6%
Прочие	4,7%	3,2%	2,7%
Всего	100,0%	100,0%	100,0%

систем

Рисунок 2. Популярность

поисковых систем у жителей России

<< Авг 19

сентябрь 2019 г.

отчет: переходы из поисковых систем

[по дням](#) | [по неделям](#) | [по месяцам](#)

значения:	сентябрь 2019 г.		август 2019 г.		в среднем за 3 месяца	
среднесуточные / <u>суммарные</u>						
<input checked="" type="checkbox"/> Яндекс	12,498,021	49.1%	11,677,322	49.7%	11,923,281	49.9%
<input checked="" type="checkbox"/> Google	12,304,313	48.4%	11,192,431	47.7%	11,365,048	47.5%
<input checked="" type="checkbox"/> Search.Mail.ru	569,472	2.2%	535,695	2.3%	548,008	2.3%
<input checked="" type="checkbox"/> Rambler	41,351	0.2%	39,089	0.2%	42,423	0.2%
<input checked="" type="checkbox"/> Bing	23,065	0.1%	18,770	0.1%	19,684	0.1%

Рисунок 3.

Соотношение пользователей поисковых систем в Москве

1. Статистика поисковых систем 2019 года см.: <https://drmax.ru/statistika-poiskovyh-sistem-2019.html> Дата обращения: 14.11.2019 [↑](#)
2. Популярные поисковые системы в мире: интересные факты см.: <https://project-seo.net/blog-ru/rejting-poiskovyh-sistem-2018-2019/> Дата обращения: 10.11.2019 [↑](#)
3. Колисниченко Д.Н. Часть IV. Поисковая система Рамблер // Поисковые системы и продвижение сайтов в Интернете — М.: Диалектика, 2007. — 272 с. [↑](#)
4. 7 самых популярных поисковых систем в мире см.: <https://hype.ru/@id27/7-samyh-populyarnyh-poiskovyh-sistem-v-mire-0g70cweu> Дата обращения: 10.11.2019 [↑](#)
5. Поисковые системы Internet: эволюция и перспективы см.: <http://www.gpntb.ru/win/inter-events/crimea97/doc/sem4/s4doc3.html> Дата обращения: 13.11.2019 [↑](#)