

Содержание:

ВВЕДЕНИЕ

На сегодняшний день, каждый человек в мире начинает свое утро с поиска какой-либо информации. В мире где технологии так быстро развиваются и являются уже неотъемлемой частью нашей жизни, очень важно правильно и своевременно ее находить. На примере данной работы я хотел бы провести анализ поисковых систем, так как умение быстро находить точную и полную информацию и знание как работают поисковые системы, есть ключ к многим возможностям

Глава 1: История возникновения сетей интернет

В начале развития сети Интернет, в сети было крайне мало информации, а круг пользователей, которые имели к ней доступ так же был ограничен. В основном доступ был у сотрудников крупных компаний, госслужащих, а также научных сотрудников и сотрудников лабораторий.

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.

Таким образом, 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.

На ранних этапах развития сети интернет каталоги ссылок использовались очень активно, но постепенно утрачивали свою популярность. Причина проста: даже при наличии множества ресурсов в современных каталогах, они все равно показывают

только малую часть имеющейся в сети Интернет информации. К примеру, самым большим каталогом сети является — DMOZ (Open Directory Project). Он содержит информацию о чуть более пяти миллионах ресурсов, что несоизмеримо с поисковой базой Google, содержащей более восьми миллиардов документов.

Самым крупным русскоязычным каталогом является Яндекс-каталог. Он содержит информацию о чуть более ста четырёх тысячах ресурсов.[1]

1.1 Хронология развития поисковых систем

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 таким образом хранимой информации (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. Увы 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. 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. 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. В этом же году были запущены первые отечественные поисковые системы – Rambler.ru и Aport.ru. Появление первых отечественных поисковых систем ознаменовало новый этап развития Рунета, позволяя русскоязычным пользователям осуществлять запрос на родном языке, а также оперативно реагировать на изменения, происходящие 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. В октябре 2001 Дэни Салливан написал статью под названием «База данных спам сайтов Inktomi открыта для публичного пользования», в которой рассказывалось о том, как Inktomi случайно сделал свою базу данных спам сайтов, которая к тому времени насчитывала уже около 1 миллиона URL, доступной для всеобщего использования.

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). 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.

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.

На сегодняшний день на мировом рынке обосновались три лидера – Google, Yahoo и Bing. Они имеют свои собственные базы, и свои алгоритмы поиска. Многие другие поисковые системы используют результаты этих трех основных поисковых систем. Например, AOL использует базу данных Google в то время как AltaVista, Lycos и AllTheWeb используют базу данных Yahoo. Все остальные поисковые системы в различных комбинациях, используют результаты (выдачу) перечисленных систем.

Если же провести аналогичный анализ поисковых систем, популярных в странах СНГ, то мы увидим, что mail.ru транслирует поиск Google, при этом накладывая свои новые наработки, Rambler, в свою очередь, транслирует Яндекс. Поэтому весь рынок рунета можно разделить между этими двумя гигантами. [2]

Глава 2: Развитие поисковых систем

2.1 Развитие Google

Компания Гугл на сегодняшний день является мировым лидером в сфере поисковых систем. Но более 20 лет назад все начиналось с простой дружбы двух энергичных студентов.

Сергей Брин (со-основатель Гугл) будучи студентом в 1995 году согласился провести экскурсию в Стэнфордском университете для студентов. Там он познакомился с другим студентом и так же со-основателем Гугл Ларри Пейджом. С этого момента они становятся друзьями и начинают работу над поисковой системой, которая в дальнейшем выльется в такого медиа гиганта каким мы знаем сегодня Гугл.

Хронология развития Гугл:

1996-1997 Брин и Пейд работают над созданием поисковой системы BlackRub. Она работает на серверах Сендфордского университета, но ей перестает хватать ресурсов.

Создатели решают переименовать систему и 15 сентября 1997 года, появляется google.com.

1998 год. Компания начинает стремительно развиваться. Появляются первые инвесторы. Брин и Пейдж закупают новое оборудование и открывают офис в гараже. В сентябре 1998 года они регистрируют компанию в Калифорнии и нанимают своего первого сотрудника, их бывшего однокурсника Крейга Сильверштейна. В декабре 1998 по версии журнала PC Magazin Google вошел в ТОП100 лучших сайтов за год.

1999 год. Гугл стремительно растет. В связи с этим принимается решение о переезде в другой офис. Набираются все новые и новые сотрудники.

2000 год. Появляется сервис MentalPlex, способной читать мысли пользователей. Так зарождается новая традиция первоапрельских розыгрышей в Кремниевой долине. Появляется первая серия знаменитых дудлов, не связанная с какими-то важными событиями. Появляются зарубежные версии Гугл, на сегодняшний день Гугл можно найти на 150 языках мира. В 2000 году появляется отделение Гугл в Нью-Йорке. Появляются плагины для браузеров. В свет выходят все новые и новые функции.

2001 год. Гугл может себе позволить уже приобретать другие компании. Открывается офис в Токио. Запускается сервис Гугл картинки.

2002 год. Запускается Гугл Покупки, Гугл Новости и Гугл Лаборатория, где можно посмотреть новейшие разработки компании. Открывается офис в Сиднее. В феврале Гугл выпускает свое первое устройство Google Search Appliance, которая подключается к корпоративной компьютерной сети и существенно упрощает поиск внутренних документов.

2003 год. Гугл приобретает компанию Pyra Labs, разработчика Blogger. Этот сервис – почти ровесник Google. Сейчас его страницы посещают более 300 млн человек в месяц.

Появляется сервис Книги. Гугл организует конкурсы для программистов и рекламные гранты.

2004 год. Запускается Гугл Почта, Gmail. Организуется сервис по локации, который в дальнейшем выльется в Гугл Карты. Публикуется блог компании в Twitter и Google+. Объявляются первые получатели стипендии Гугл. Открывается офис в Гонконге. В 2004 году Гугл выходит на биржу. Акции ценятся довольно таки высоко. Появляется знаменитый сервис Планета Земля.

2005 год. Активно развивается сервис Карты, появляется множество возможностей и функций. Происходят оптимизации для мобильных устройств.

2006 год. Появляются все новые и новые сервисы. Гугл Финансы, Переводчик, Календарь, Кошелек, Таблицы и Документы, теперь все можно найти на Гугл. В конце года Гугл решает купить видео хостинг YouTube.

2007 год и настоящее время. Гугл продолжает улучшать свои сервисы и развиваться во всех направлениях. Выходят много мобильных устройств от компании с собственной операционной системой. Происходят покупки перспективных стартапов и многое другое.[3]

2.2 Развитие Яндекс

Выпускник и аспирант Института нефти и газа им. Губкина Аркадий Волож трудился над обработкой и анализом больших объемов данных в начале 90х годов и так же был генеральным директором ComТек. В процессе работы у программиста возникла идея создания компьютерной программы, которая позволила бы находить в больших текстах информацию с учётом морфологии русского языка(то есть по запросу слова в любой форме). Вместе с Аркадием Борковским, специалистом по компьютерной лингвистике Академии наук СССР, Волож в 1989 году открыл компанию «Аркадия». В 1990 году к команде программистов ComТек присоединился программист Илья Сегалович, бывший одноклассник и друг Воложа. Специалисты «Аркадии» в течение двух лет создали по заказу Центрального научно-исследовательского института патентной информации и технико-экономических исследований (ЦНИИПИ) две информационно-поисковые системы: Международную классификацию изобретений и Классификатор товаров и услуг. Впоследствии программы, записанные на дискеты, в течение трёх лет продавали научно-исследовательским институтам и организациям, занимавшимся патентованием. К 1993 году ComТек занималась сбытом сетевых технологий, а чтобы не оставлять технологию поиска, «Аркадию», деятельность которой стала менее востребованной, сделали одним из департаментов ComТек. В 1993 году была написана первая рабочая версия локального поисковика для локальных компьютеров который назвали Yandex.

В 1995 году, когда компания ComТек вышла в сеть Интернет, появилось первое решение настроить поисковую систему для новых условий, чтобы пользователем было удобно ориентироваться в сети и проще искать информацию.

1997-2000 год

В сентябре 1997 года на выставке Softool была представлена поисковая система Яндекс, а через два месяца уже был реализован естественно-языковой запрос. Яндекс индексировал сайты в расширении .su и .ru. Программа учитывала специфику русского языка и умела находить начальную форму слова. Язык запросов позволял искать запросы в документах, абзацах, заголовках и тд.

В конце 1997 года Яндекс стал официальным поисковиком в браузере Internet Explorer для наибольшей локализации на Российском рынке.

В 1998 году появилась функция «найти похожий документ». По состоянию на 1998 год Яндекс работал на трех серверах, одна обрабатывала Интернет и индексировала документы, вторая была поисковая и третья дублировала вторую.

В 1999 года произошло обновление поисковой системы, улучшены фильтры анти-спам, появилась возможность поиска по категориям и введен тематический поиск.

В дальнейшем на развитие системы требовались средства и Волаж стал искать инвесторов. В итоге в конце 1999 года, капитализация компании оценивалась в 72 000 долларов в год, хотя по оценки самого Яндекс сумма была около 15 миллионов в год.

2.2.1. Начало 2000х годов

Весной 2000 года ComTex учредили ООО «Яндекс», для управления сайтом. Компании перешли права на одноименные продукты и созданных проектов. В этот период времени инвесторы со всего мира стали интересоваться развитием интернета и интернет-фирмами. Волажу удалось привлечь инвесторов на довольно значительную часть акций компании. В том же году поисковик получил обновление, были открыты сервисы Почта, Новости, Товары, Открытки и тд. В 2001 году Яндекс становится лидером среди поисковых систем в Рунете обогнав Рамблер и всех остальных.

В 2002 году доходы сервиса начали превышать расходы, а летом того же года компания вместе с российской PayCash запускает сервис ЯндексДеньги.

В 2003 году компания начинает выплату первых дивидендов своим инвесторам.[4]

2.2.2. 2004 год и настоящее время

В 2004 году Яндекс объявляет об открытии образовательных программ. При поддержке сервиса избранным стипендиатам была оказана поддержка в написании научных статей. В том же году были открыты новые сервисы ЯндексКарты, ЯндексБлог.

В 2005 году был открыт офис на Украине, а уже в 2006 году запустили первый удаленный офис разработки, который находился в Санкт-Петербурге. С 2007 года компания Яндекс продолжила улучшать свои старые сервисы, а также выпускать новые. В 2011 году для компании происходит зачинаемое событие и Яндекс выходит на биржу, а годом позже они выпускают свой собственный браузер, так же продолжается стремительное улучшение своих старых сервисов и создание новых.

На сегодняшний день Яндекс занимает 4е место в мире среди поисковых систем в Интернете.[5]

2.3 Развитие Baidu, Yahoo! и всех остальных

2.3.1 Yahoo!

В 1994 году аспиранты Стэнфордского университета Дэвид Фило и Джерри Янг создали свой поисковый сервис «Путеводитель Джерри во всемирной паутине». В дальнейшем оценив перспективу данного сервиса сайт был переименован в Yahoo! и была открыта одноименная корпорация. В конце 90х годов XX века Yahoo! был мировым лидером в среде поисковых систем интернета. Активно развивались сервисы, появился первый бесплатный почтовый сервис, было объявлено о слиянии с eBay, которое правда произошло лишь в 2006 году и то с частичным объединением.

В начале 2000х годов лидерство Yahoo! было не так заметно, и развитие пошло на спад. К настоящему времени сервис собирается быть продан двум телекоммуникационным компаниям США, в связи со своей нерентабельностью.

2.3.2 Baidu

В 2000 году в Китае появилась компания Baidu, которая занималась в основном поисковыми системами. Основана в 2000 году, основатели — Робин Ли и Эрик Сю, получившие высшее образование в США. \$1,2 млн стартового капитала привлекли от американских венчурных компаний. Название компании взяли из поэмы времен династии Сун, оно означает «поиск мечты». Через год получили от венчурных компаний ещё \$10 млн. инвестиций. В 2004 году Baidu стала лидирующей поисковой системой в Китае.

В 2005 году компания вышла на биржу, при размещении акции Baidu установили рекорд десятилетия — к концу торговой сессии они подорожали более чем в четыре раза. В декабре 2007 года Baidu стала первой китайской компанией, акции которой включены в индекс NASDAQ-100.

2.3.3 Прочие поисковые системы

В России в середине 90х годов были запущены поисковые системы Рамбер и Апорт, в мире были распространены MSN Search, Ask, Bing и многие другие. На сегодняшний день многие из сервисов не так популярны как раньше и большую долю рынка занимают Google, Baidu, Bing и в России Яндекс.

3. Типы поисковых систем

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

Системы, использующие поисковых роботов

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, 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 и MSN Search.

4. Алгоритмы работы поисковых систем

Работа поисковых систем осуществляется по определенным формулам. Каждый алгоритм основан на поиске и отображении определенных запросов.

Использование определенных алгоритмов позволяет поисковым системам лучше отбирать запросы, отсеивать не нужную информацию и оптимизировать работу. Алгоритмы учитывают плотность материалов на сайте, как долго пользователь проводит на нем и многое другое. Но самое главное, что учитывает поисковая система, это уникальность самого контента на сайте.

Рассмотрим алгоритмы работ поисковых систем на основе популярных систем в нашей стране.

4.1 Алгоритм работы Google

Первые поисковые алгоритмы Google были основаны на ранжировании страниц по PageRank. Чем большее количество внешних ссылок вело на ресурс, тем более авторитетный показатель ему присваивался.

Изначально поисковый алгоритм Google уделял максимум внимания внутренним характеристикам страницы. Теперь же учитываются и множество других показателей, в том числе и географическая принадлежность.

С 2000 года Google запустил алгоритм под названием **Hilltop (Хилтоп)**, который лучше просчитывал значения PageRank.

В 2003 году поисковая система ввела алгоритм под названием **Florida (Флорида)**, который поставил под сомнение используемые на тот момент методы продвижения и понизил в результатах поиска сайты, материал которых был сильно переспамлен ключевыми словами.

Три года спустя, в 2006 году, в работу поисковой машины был внедрен новый алгоритм, разработанный Ори Алоном, под названием **Orion (Орион)**. Он значительно улучшил качество поиска. При ранжировании сайтов алгоритм учитывал их качество.

В 2007 г. Google вводит новый алгоритм **Austin (Остин)**. В результате его работы большинство ресурсов, занимающих первые позиции, потеряли свои «места под солнцем».

В 2009 году Google ввел алгоритм **Caffeine (Кофеин)**, который существенно увеличил скорость поиска и индексации страниц.

4.2 Алгоритм работы Яндекс

Руководители системы Яндекс предоставляют пользователям значительно больше информации о принципе работы поисковой машины.

С 2007 года обо всех изменениях в работе системы пользователи могут узнавать из официального блога Яндекс.

Первый алгоритм, о котором Яндекс заявил публично, стал **8 SP1**. Все последующие алгоритмы ПС были названы в честь различных городов. Следующий внедренный алгоритм получил название **Магадан**. Его анонс состоялся 16 мая 2008 года. Возможности алгоритма:

количество критериев при ранжировании увеличилось в два раза; внедрена система классифицирования сайтов в зависимости от их содержания и тематики

ссылок;

возросла точность и скорость поиска;

система «научилась» распознавать аббревиатуры.

Помимо вышеперечисленных улучшений поисковая система Яндекс начала индексировать зарубежные ресурсы. Следующим поисковая система внедряет алгоритм **Находка**. Он включил в себя следующие изменения:

расширение словарной базы системы;

улучшение способов обучения поисковой машины.

Также пользователи заметили, что по многим запросам лидирующие позиции занимают старые сайты, из чего следует, что возраст ресурса, играет определенную роль при его продвижении.

Алгоритм Арзамас результаты выдачи по запросу стали отличаться в зависимости от географического расположения пользователей, всего было создано 19 зональных принадлежностей; были внедрены фильтры, которые понижали позиции сайтов, использующих «жесткие» рекламные методы, в том числе popunder, а затем и clickunder (bodyclick);

Алгоритм Снежинск осенью 2009 года Яндекс заявил о внедрении алгоритма **Снежинск**, основной частью которого стал Матрикснет (MatrixNET). Результаты стали еще более релевантными поисковым запросам, так как система смогла переобучаться и делать вычисления точнее. Были введены следующие изменения в работе поисковой системы:

начали работать фильтры АГС (АГС 17 и АГС 30);

количество факторов при ранжировании страниц возросло в разы;

была внедрена система распознавания первоисточника информации и обнаружения сайтов, копирующих материалы. В отношении последних начали применяться меры для понижения их в результатах поиска. В дальнейшем поисковая система вводила новые алгоритмы: **Конаково, Обнинск, Краснодар, Калининград, Дублин**. Каждый последующий алгоритм совершенствовал работу поисковой машины. Наиболее значимыми улучшениями стали:

совершенствование ранжирования для геодеpendимых поисковых запросов;

индексация информации из социальной сети ВК;

начала работать партнерская программа «Оригинальные тексты», в которой автор может подтвердить свои права на материалы. Также было внедрено большое количество мелких доработок. 15 апреля 2015 года Яндекс объявил о внедрении нового алгоритма ранжирования Минусинск, старт которого запланирован на 15 мая 2015 года. Новый алгоритм направлен на борьбу с закупкой ссылок с целью быстрого выведения сайта в ТОП. Теперь сайты, злоупотребляющие закупкой ссылок, будут пессимизироваться в поисковой выдаче Яндекс.

5 Архитектура поисковых систем

Рассмотрим классическую архитектуру, которая чаще всего реализована на корпоративных сайтах и информационных порталах. Такая архитектура изображена на рисунке 1.

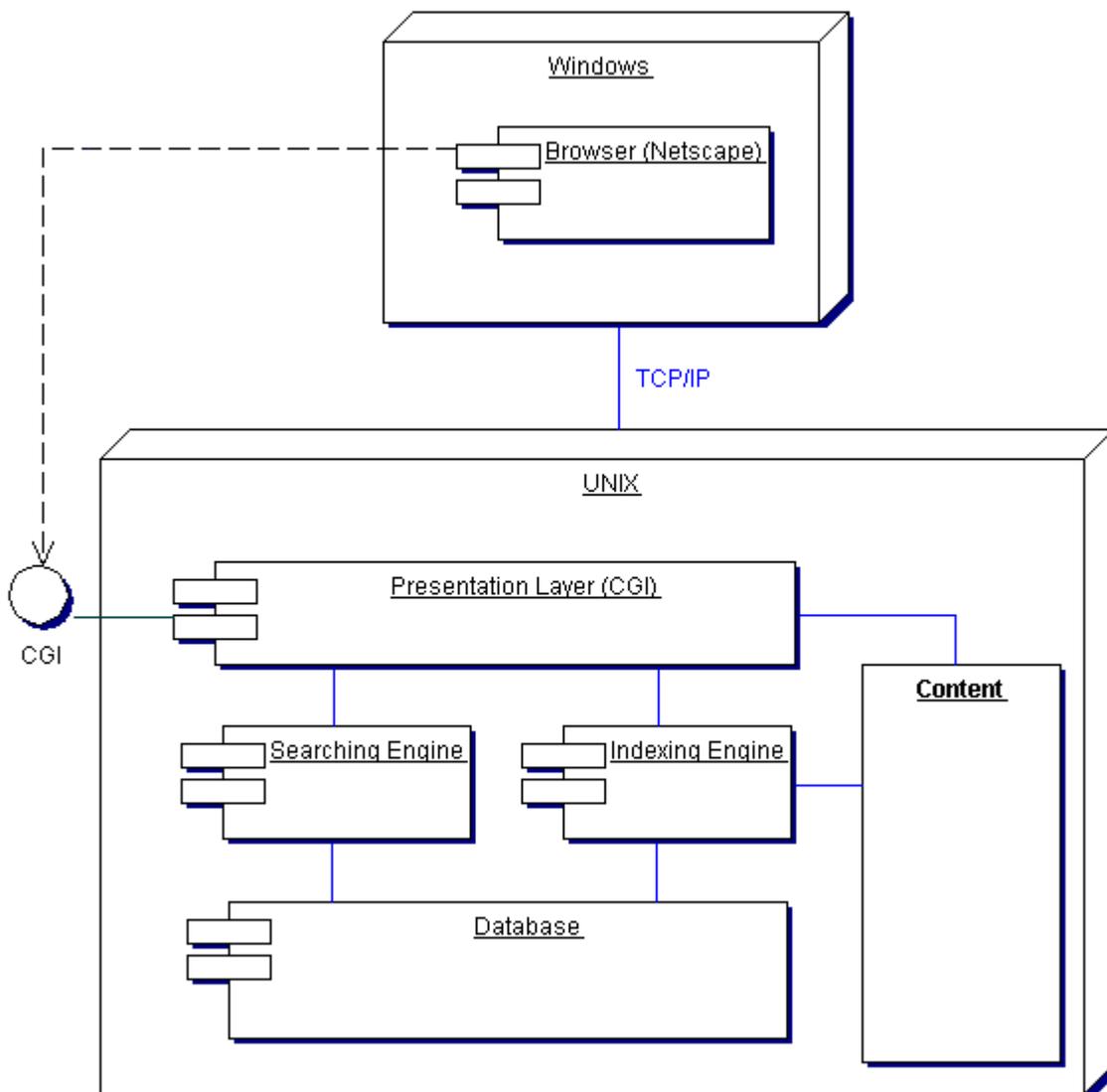


Рисунок 1. Архитектура поисковой системы

Разберем по частям то, что изображено на рисунке. Существует клиентская вычислительная машина под управлением ОС Windows и существует Web-сервер под управлением UNIX-подобной ОС. На стороне клиента запущен типичный браузер, такой как Netscape. На стороне сервера запущен web сервер, который обслуживает запросы от браузера, передавая запросы презентационному слою понимающему CGI. Презентационный слой передает запросы к поисковому механизму в случае вызова услуги поиска или отображает наполнение (content) сайта. При работе администратора презентационный слой также может передавать запросы на инициализацию механизма индексации нового контента, который еще не индексируется. Это необходимо по той причине, что пока текст не индексируется, поиск в нем с помощью поисковой машины невозможен. Идея заключается в следующем. Существуют мегабайты текстовой информации, и скорость поиска документов содержащих заданные ключевые слова отнимает очень много процессорного времени. Предположим, в 10 мегабайтах текстовой информации ключевое слово будет находиться в течение 10 секунд. И вот заходит посетитель на Ваш сайт, задает ключевые слова, вызывает услугу поиска и ждет 10 секунд, пока ваш сервер не выдаст ему результат. Предположим, случилось так, что одновременно запросило поиск 5 человек. Естественно, время ответа увеличится в 5 раз. Получается, что в среднем по 50 секунд пользователь будет ждать ответа от вашего сервера. Это не приемлемо, особенно если у Вас сотни мегабайт текстовой информации и время реакции системы будет катастрофически велико. Необходимо использовать другой подход при поиске ключевых слов в текстовой информации - время ответа сократить до миллисекунд.[6]

6 Защита поисковых систем

ЗАКЛЮЧЕНИЕ

В своем заключении хочу сказать, что поисковые системы важны в наше время, так как важно применять умение нахождения информации, да и не возможно знать все и где что находится.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Закер К. Компьютерные сети. Модернизация и поиск неисправностей[6]

2. Соколов-Митрич Д. В. Яндекс.Книга. — М.: Манн, Иванов и Фербер, 2014. — 368 с[6]
3. Справочная информация по ЭВМ и телекоммуникациям.
4. Таненбаум Э., Уэзеролл Д. Компьютерные сети — 5-е изд. — СПб: Питер, 2012.[6]
5. Яндекс:Найдётся всё// COMPUTER BILD — 2012. — вып. 10 (162). [4]
6. Google.com [3]
7. Jawadekar, Waman S. 8. Knowledge Management: Tools and Technology // Knowledge management: Text & Cases. — New Delhi: Tata McGraw-Hill Education Private Ltd, 2011. — С. 278. — 319 с. — ISBN 978-0-07-07-0086-4.[1]
8. Pariser E. The Filter Bubble: What The Internet Is Hiding From You. — NY: Penguin Group, 2011. — 257 с. — ISBN 978-0-14-196992-3.[2]
9. Zhang, Séaghdha, Quercia, Jambor Auralist: introducing serendipity into music recommendation (англ.) // ACM WSDM. — 2012. — P. 13-22. [2]
10. Tarakeswar M. K., Kavitha M. D. Search Engines: A Study (англ.) // Journal of Computer Applications (JCA): journal. — 2011. — Vol. 4, no. 1.[2]
11. Yandex.ru [5]