

Содержание:

Введение

Современный этап развития цивилизации характеризуется переходом наиболее развитой части человечества от индустриального общества к информационному. Одним из наиболее ярких явлений этого процесса является возникновение и развития глобальной информационной компьютерной сети Интернет.

Всемирная сеть очень важна и полезна почти для всех! Любой пользователь Интернета может отыскать в нем много разной и интересной информации, а также использовать все широкие возможности сети. Для меня главными обстоятельствами в выборе темы «Анализ поисковых систем в сети Интернет», для моей курсовой работы, стали актуальность темы на сегодняшний день, а также достаточная открытость и известность мне этой темы, так как я часто пользуюсь всемирными сетями.

Ресурсы Интернета уже давно не просто игрушка, превратившаяся в незаменимый инструмент для каждодневной работы людей различных профессий. Количество данных в сети стремительно растет, и пропорционально им растет и объем. Ученые утверждают, что объем информации, передаваемой по Интернету, увеличивается в два раза каждые шесть месяцев.

В сети каждый день появляются множество новых документов, и что бы быстро и качественно находить нужную информацию появились специальные поисковые средства. Несколько лет назад говорили: в Интернете ничего невозможно найти, но там есть всё. Но когда появились поисковые программы, ситуация в корне поменялась, и сейчас в интернете информацию которая вам нужна, можно найти практически мгновенно.

Наиболее популярным способом поиска в Интернете является использование поисковых систем. Что же такое поисковая система? Поисковая система – портал, осуществляющий поиск, сбор и сортировку информации в сети Интернет. Первоочередная задача любой поисковой системы – доставлять людям именно ту информацию, которую они ищут.

В данной курсовой работе рассматривается поиск информации во Всемирной паутине с помощью нескольких наиболее распространённых систем поиска (Рамблер, Яндекс и т.д.). Всего же в мире существуют сотни различных поисковых систем, и выбор той или иной системы зависит только от ваших личных пристрастий.

Глава 1. Теоретическая часть

1.1 Особенности поисковых систем

Поисковая система – это сайт, к которому пользователь обращается посредством ключевого слова и находит интересующую его информацию. Первоочередная задача любой поисковой системы – доставлять людям именно ту информацию, которую они ищут.

По пространственному масштабу поисковую систему можно разделить на локальные, глобальные, региональные и специализированные. Локальные поисковые системы могут быть разработаны для быстрого поиска страниц в масштабе отдельного сервера. Региональные поисковые системы описывают информационные ресурсы определенного региона, например, русскоязычные страницы в Интернете. Глобальные поисковые системы в отличие от локальных стремятся объять необъятное - по возможности наиболее полно описать ресурсы всего информационного пространства сети Интернет.

Поисковой процесс представляет собой четыре этапа: формулировка (до начала поиска), действие (идет поиск), обзор результатов (выводится на экран), и усовершенствование (возвращаемся к поиску с той же потребностью, но уже иной формулировкой). Более удобная схема поиска состоит из следующих этапов:

- Задание информационной потребности на естественном языке;
- Выбор поисковых сервисов и точная запись информационной потребности;
- Выполнение созданных запросов;
- Предварительная обработка полученных ссылок на документы;
- Обращение за искомыми документами;
- Предварительный просмотр найденных документов;
- Сохранение подходящих документов для изучения;
- Извлечение из подходящих документов ссылок для расширения запроса;

- Изучение всех сохраненных документов;
- Если искомая информация не найдена, то возвращаемся к первому пункту.

1.2 Методы поискового запроса

Процесс поиска довольно прост: задавая ключевые слова, характерные для искомой информации, пользователь надеется максимально просто и быстро найти интересующую его информацию. Это значит, что поисковая система должна думать точно также как думает пользователь, когда ищет ту или иную информацию. Разработчики поисковых систем постоянно стараются совершенствовать алгоритмы и принципы поиска, пытаются всячески ускорить работу системы, добавляя новые функции и возможности, чтобы удовлетворить потребности пользователей.

Поисковая система состоит из поисковой машины, базы данных или индекса и точек входа в систему. Поисковая машина – это аппаратно-программный комплекс, который осуществляет быстрый поиск внутри сервера или Интернет-ресурса необходимой информации.

Поисковая машина состоит из:

- паук или спайдер (spider) - приложение, которое занимается скачиванием страниц интернет ресурсов.
- индексатор (indexer). Индексатор производит первоначальный анализ содержимого скачанной страницы, выделяет основные части (название страницы, описание, ссылки, заголовки и т.д.) и раскладывает все это по разделам поисковой базы данных – помещает в индекс поисковой системы. На основе результатов первоначального анализа индексатор также может принять решение, что страница вообще “недостойна” находиться в индексе. Причины такого решение могут быть разными: страница не имеет названия, является точной копией другой, уже имеющейся в индексе страницы или содержит ссылки на запрещенные законодательством ресурсы.
- краулер (crawler) - программа, являющаяся составной частью поисковой системы и предназначенная для перебора страниц Интернета с целью занесения информации о них в базу данных поисковика. Краулер анализирует пути, ведущие с текущей страницы на другие разделы сайта, или на страницы внешних интернет ресурсов и определяет дальнейший порядок обхода пауком нитей всемирной паутины. Именно краулер находит новые для поисковой

машины страницы и передает их пауку. Работа краулера построена на базе алгоритмов поиска на графах в ширину и глубину.

- подсистема обработки и выдачи результатов (Search Engine and Results Engine) - самая важная часть любой поисковой машины. Алгоритмы работы этой подсистемы компании разработчики хранят в строгой секретности, поскольку они являют собой коммерческую тайну. Именно эта часть поисковой машины отвечает за адекватность ответа поисковой системы на запрос пользователя.

Вне зависимости от того, какой поисковой системой вы пользуетесь, примерный алгоритм поиска информации следующий:

- Вы переходите на начальную страницу поисковой системы или на любую страницу, на которой находятся поля для ввода запроса и кнопка для начала поиска. В последнем случае после нажатия кнопки вы автоматически перейдете на страницу поисковой системы. Отличий между собственно поисковой системой и страницей с полями для ввода запроса, который передается поисковой системе, нет. Главное – перейти на страницу, на которой вы можете ввести ваш запрос;
- Если система позволяет последовательные уточнения области поиска, то вы переходите в разделы, определяющие границы поиска информации (например, поиск только в области науки или поиск среди европейских узлов Интернета). В таких системах, запрос в главном разделе, осуществляет поиск среди всех узлов Интернета. Если же вы, например, перейдете в раздел «Новости», то поиск по запросу будет выполняться только среди узлов, посвященных новостям. Таким образом, вы сначала определяете границы поиска, после этого отправляется запрос на поиск информации. Если выбранная вами система поиска не позволяет уточнить запрос, то это действие пропускается;
- Вводится запрос на поиск в соответствии с правилами, принятыми в выбранной системе поиска. В простейшем случае это одно или несколько слов, например «Анекдоты», но возможны и сложные запросы с логическими операциями «И», «ИЛИ», «НЕТ»;
- Если система поиска позволяет, то уточняются некоторые параметры запроса с помощью дополнительных полей, списков, флажков и переключателей, которые могут находиться на странице рядом с полем запроса. В большинстве случаев уточнение параметров не требуется. Составление сложных запросов чем-то похоже на программирование и пользуется популярностью у программистов, людей, не знакомых с программированием;

- Нажимается кнопка для начала поиска. После некоторой паузы вы автоматически попадаете на страницу со списком документов, удовлетворяющих вашему запросу. Длительность паузы зависит от сложности запроса, скорости работы поисковой системы и качества вашей связи с данной системой;
- Если ссылок много, и они не помещаются на одной странице, то после прочтения первой страницы с результатами поиска вы можете перейти к следующей, и так далее. Таким образом, вы просматриваете все найденные ссылки на документы;
- После нахождения нужной ссылки вы щёлкаете на ней мышью и переходите на страницу, которую вы разыскивали. Если вы захотите посмотреть другие найденные страницы, то можно вернуться назад, к странице со ссылками на найденные документы, и перейти по другой ссылке.

Когда говорят о поисковых запросах, то часто упоминают слово «релевантность». Слово релевантность означает степень соответствия документа запросу. Чем больше релевантность, тем лучше результаты поиска. Релевантность определяется по следующим критериям:

1. Проверяется, сколько раз заданное словосочетание (слово) встречается на отображенных страницах;
2. Если введена фраза, то проверяется расстояние между словами;
3. Индекс цитируемости (количество ссылок на данную страницу);
4. Чем больше на страницу ссылаются, тем больше вероятность, что именно эта страница нужна пользователю;
5. Каким стилем (шрифтом) написана на странице искомая фраза;
6. Если фраза выделена жирным или написана крупным шрифтом; (заголовок), то это признак более высокой релевантности страницы;
7. Возраст сайта (Чем старше сайт, тем лучше это влияет на релевантность).

Таким образом, можно отметить, что уровень релевантности – это целый комплекс параметров.

Итак, для того чтобы найти интересующую вас информацию в кратчайшее время, вам необходимо правильно составлять поисковой запрос.

Наиболее эффективным является следующий метод составления поискового запроса:

1. Слова поискового запроса нужно писать грамотно. Так, например, запрос «пошив пальтов» нужного результата не даст;
2. Если поиск нужных результатов не принес, нужно переформулировать запрос с помощью синонимов;
3. Если результат поиска не удовлетворяет вас, то попробуйте включить в запрос уточняющие слова. Для запроса «стихи Пушкина» слово «стихи» является уточняющим;
4. В сложных случаях нужно использовать язык запросов. С помощью языка запросов можно задавать комбинации разных критериев поиска.

1.3 Поисковая система Yandex

Подробнее рассмотрим язык запросов поисковой системы Yandex. Предположим, что нам нужно найти адрес проживания Мартиросяна С.Т. Если правильно сформулировать поисковой запрос, то это не составит нам труда. В языке запросов есть такая функция, как включение/ исключение определенных слов – знаки «+» и «-».

Попробуем сформулировать поисковой запрос. Для этого используем фамилию человека и пару уточняющих слов. В качестве уточняющего слова возьмем «адрес» и дату рождения человека. Получим:

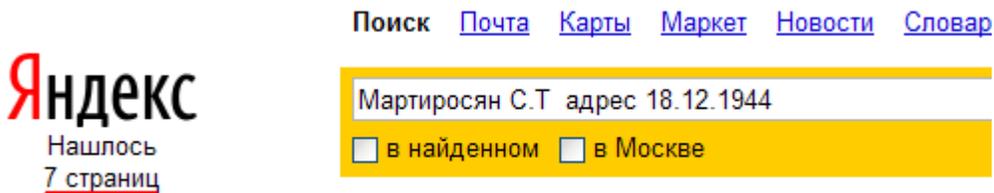


Рис.1

Как видно на рисунке 1 нашлось 7 страниц, но с помощью языка запросов можно сократить их число до 1 страницы. Применим специальный символ «+» к уточняющим словам – это означает, что будут найдены те страницы, на которых в обязательном порядке присутствуют уточняющие слова. Получим:

1. [Ищите старого друга, одноклассника, коллегу? Все люди Москвы у нас! Их адреса...](#) 

Москвич: **МАРТИРОСЯН САШИК ТОРГОМОВИЧ**, **18.12.1944** г.р.: в одноклассниках, в поиске работы живет по **адресу**: Москва, СТАЛЕВАРОВ УЛ. д. 18 19, эта улица: на карте, в недвижимости, в организациях.

www.gorstat.ru/8/542.php [сохраненная копия](#)

Из результатов поиска исключены слишком похожие страницы. [Показать все без исключения.](#)

▶ [Яндекс.Телепрограмма](#)

Не пропустите любимую передачу!

Рис.2

На рисунке 2 видно, что мы нашли 1 страницу, без лишнего мусора. Этот поисковый запрос сформулирован правильно. Еще одним важным символом языка запросов является знак «|». Вертикальная черта позволяет задать альтернативы – система ищет хотя бы одно из перечисленных слов. Например запрос: «Физика лекции | конспекты» будет аналогичен как запросам «Физика лекции» и «Физика конспекты».

Обычно поисковики ищут все формы введенного слова. Например по запросу «Звукоизолирующая» будут выведены страницы со словом «звукоизолирующий», «Звукоизолирующих» и т.д. Рисунок 3.

1. [Акустик.Ру - звукоизоляция квартир, шумоизоляция, студии звукозаписи,](#) 
[Инженерные решения](#)  [Продукция](#)  [Прайс-лист](#)  [Контакты](#)  [Услуги](#) 
эффективным **звукоизолирующим** и звукопоглощающим материалам и технологиям их применения. новейшим разработкам в области архитектурной и строительной акустики. Австрийская фабрика HERADESIGN при содействии Акустик Групп открыла официальное...
+7 (495) 959-29-44 Москва, Новокузнецкая ул., 33, корп.2, офис Акустика
www.acoustic.ru [сохраненная копия](#) [еще с сайта](#)
2. [Шумоизоляция и виброизоляция в Самаре](#) 
Звукоизолирующая панельная система начального уровня ЗИПС®-Вектор. Применяется в строительстве и реконструкции зданий для увеличения звукоизоляции однослойных строительных конструкций: гипсовых, кирпичных и бетонных стен и перегородок.
www.acoustic-stroy.ru/materials/?... [сохраненная копия](#)
3. [Можно ли с помощью **звукоизолирующих** материалов добиться абсолютной тишины...](#) 

Рис. 3

Но чтобы найти именно ту форму слова, которая задана в поисковой строке, в языке запросов предусмотрен символ «!». Применение этого символа видно на рисунке 4.

1. [Звукоизолирующая](#)  [кабина IAC MINI 250](#)
Оборудованная бесшумной системой вентиляции, освещением и панелью разъемов для подключения аудиометров, **звукоизолирующая** камера готова к работе через несколько минут после установки. звукоизолированная дверь с магнитным уплотнителем
www.assomed.ru/good_1894467002.html [сохраненная копия](#) [еще с сайта](#)
2. [Звукоизолирующая](#)  [способность перегородок](#)
Если в перегородке предусмотрена хотя бы одна дверь, то ее **звукоизолирующая** способность должна находиться в пределах 30 дБ. При необходимости повысить ее уровень изменяется конструкция дверного полотна. Другие статьи на эту тему: Ремонт перегородок.
homart.ru/zvukoizoliruyushhaya-sposobnost-peregorodok/ [сохраненная копия](#) [еще с сайта](#)
3. [Stroy.net.ru - купить **Звукоизолирующая** панельная система ЗИПС-Синема цена...](#) 
Чтобы купить "Звукоизолирующая панельная система ЗИПС-Синема", нажмите кнопку "Заказать" или

Рис. 4

Если вам нужно найти точную фразу, то следует её заключать в кавычки: «»

В случае, когда вы хотите найти все варианты фразы: «очень (любое слово) девушка». То нужно поставить запрос так «очень /+2 девушка» , причем 1 – это означает отсутствие слова, а 3 означает, что между «очень» и «девушка» будут два любых слова:

Яндекс

Нашлось
79 млн страниц

Очень /+2 девушка

в найденном в Москве

1. [Очень красивая девушка \(13 фото\) » хуЯндекс.net - уже всё нашлось!](#) 
Обсудить публикацию **Очень красивая девушка (13 фото)** на форуме. Уважаемый посетитель, Вы зашли на хуЯндекс как незарегистрированный пользователь. Мы рекомендуем Вам зарегистрироваться либо войти на сайт под своим именем.
xyyandex.net/.../1649-ochen-krasivaja-devushka-13-foto.html [сохраненная копия](#) [еще с сайта](#)
2.  [Видео@Mail.Ru: ящур : Женский ББ и пауэр : Очень сильная девушка!!](#) 
1 мин. 25 сек., 3,4 МБ
загружено 7 июня 2007
video.mail.ru/bk/yashyr/122/113.html [сохраненная копия](#) [еще с сайта](#)
3. [Очень сексуальная девушка \(12 фото\) на Girlfriends.ru](#) 
Зарегистрированные пользователи могут добавлять новости на сайт и оставлять свои комментарии. **Очень сексуальная девушка (12 фото)** автор: sergio | 24 марта 2009 | Просмотров: 6683 | Раздел: **Девушки-модели**.
girlfriends.ru/.../4271-ochen-seksualnaja-devushka-12-foto.html [сохраненная копия](#) [еще с сайта](#)
4.  [Очень энергичная девушка! :: Видео на RuTube](#) 
2 мин. 14 сек., 5,1 МБ
загружено 9 сентября 2008
RuTube.ru/.../997162.html?... [сохраненная копия](#) [еще с сайта](#)
5. [Очень красивая девушка Наташа »](#)
Если онлайн видео **Очень красивая девушка Наташа** не запускается, то, пожалуйста, Сообщите нам об этом.
Вы можете отправить ссылку на это видео другу
x-cams.ru/.../656-ochen-krasivaya-devushka-natasha.html [сохраненная копия](#)
6. [Очень милая девушка. 18+ | Хохмы.РУ](#) 
Новости от Radeant.com. Главная. **Очень милая девушка. 18+** Опубликовано хохмаш в пт., 11/09/2009 - 04:15.

Рис.5

На рисунке 5 видно, как между словами «очень» и «девушка» вставлено «любое слово».

Следует отметить, что при выполнении запроса поисковая система не ищет документ непосредственно в Интернете. Она обращается к своей базе данных, где

в компактном виде собраны данные об информации в Интернете. Возможно, что реальное состояние Интернета несколько отличается оттого, что храниться в базе данных.

Поисковые системы постоянно пополняют и обновляют свои базы данных, чтобы минимизировать такое несоответствие. Существует способ пополнения информации в базы данных поисковых машин: Разработчик Web-страницы самостоятельно добавляет ссылку на неё в поисковую систему. Большинство систем позволяют это сделать совершенно бесплатно. Специальная программа, называемая роботом, находит ссылку на новую страницу во Всемирной паутине. Роботы запускаются поисковой системой, чтобы проверить, существуют ли ранее найденные документы и не появились ли новые.

Глава 2. Практическая часть.

2.1. Поиск с помощью Alta Vista

Поисковая система Alta Vista была открыта для свободного использования в конце 1995 года. Это система до сих пор является достаточно популярной, хотя в последнее время другие системы составляют ей значительную конкуренцию. Система обеспечивает поиск как во Всемирной паутине, так и в группах новостей. После ввода ключевых слов вы получаете информацию о количестве найденных документов и их краткие описания со ссылками на информацию в Интернете. Расширенный поиск позволяет использовать логические операторы для формирования сложных запросов. Рассмотрим возможности системы Alta Vista на конкретных примерах.

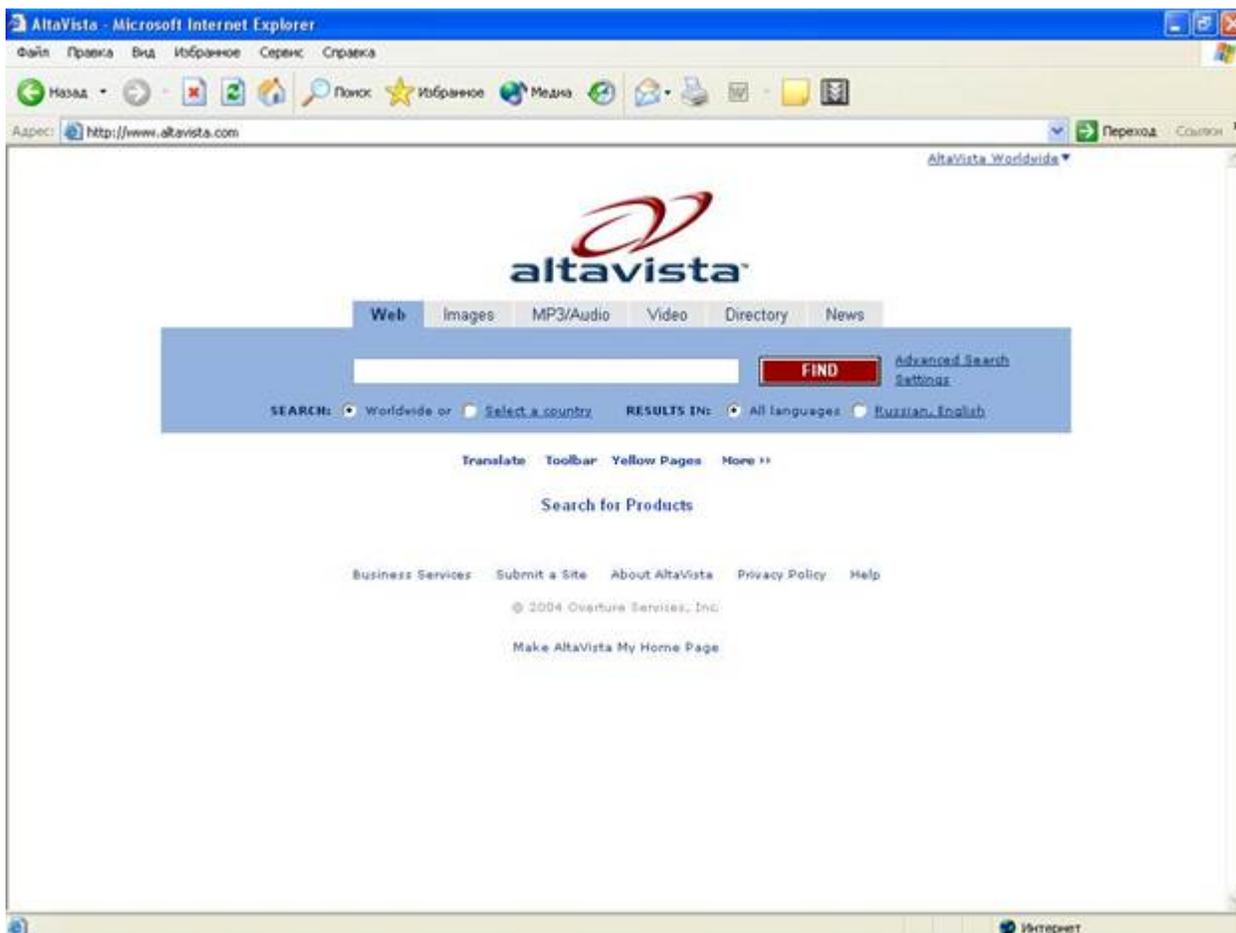


Рис.2.1 Начальная страница системы Alta Vista

На начальной странице (Рис.2.1) кроме поля для ввода запроса расположен открывающийся список. С его помощью можно выбрать интересующий вас язык искомого документа: any language (Любой язык), Russian (Русский) или любой другой язык. По умолчанию ищется информация на любом языке. Выше поля ввода расположены несколько ярлычков, щелчком на которых можно выбрать область поиска: Web (Паутина) – во Всемирной паутине, News (Новости) – в группах новостей и так далее.

Зададим несколько простых вопросов для поиска информации, чтобы продолжить знакомство с возможностями системы Alta Vista. Для поиска документов, содержащих некоторое слово, надо ввести это слово, а для поиска документов, содержащих искомое словосочетание, необходимо заключить несколько слов в двойные кавычки. Если слово содержит только строчные буквы, то ему сопоставляются также и слова, содержащие заглавные буквы.

Например, слову *интернет* будет поставлено в соответствие *интернет*, *Интернет* и *ИНТЕРНЕТ*. То есть, чтобы найти слово, написанное, как большими, так

и маленькими буквами, надо в запросе использовать только маленькие буквы (строчные).

Наличие заглавных букв в слове, заданном в запросе, конкретизирует поиск. В результате поиска будет найдено точно такое слово с учётом регистра. Например, на запрос *интерНЕТ* будет найдено только слово *интерНЕТ*, если оно, конечно, имеется в базе данных. Слова же *интернет*, *Интернет* и *ИНТЕРНЕТ* по такому запросу найдены не будут.

В системе Alta Vista можно задавать только часть слова, используя для этого знак * . Использование этого знака похоже на его использование в шаблонах файлов. Вместо переменной части слова ставится данный знак, например: на запрос *интернет** будут найдены слова *интернет*, *Интернета*, *ИНТЕРНЕТОМ*, и тому подобные. Данную возможность удобно использовать для поиска однокоренных слов.

Другой удобной возможностью является использование в запросах знаков + и -. Добавив такой знак перед словом или фразой, вы требуете обязательное их присутствие или отсутствие в документе. Если вы ставите знак + перед словом, то этим вы указываете, что данное слово обязательно должно присутствовать в найденном документе. Символ - указывает на то, что следующее за ним слово не должно присутствовать. Например: запрос *+интернет* -интернет* выдаст вам только склонения слова Интернет, но не само слово. Мы указали с помощью *+интернет**, что в искомом документе должны быть все слова, начинающиеся с «интернет». С помощью же *-интернет* мы задаём, что само слово «интернет» не должно присутствовать.

Ещё одним удобным способом уточнения поиска является использование специальных ключевых слов. Существуют различные ключевые слова, мы рассмотрим только два наиболее полезных слова. Ключевое слово *link*: позволяет ограничить поиск среди страниц, на которых расположены ссылки на заданную страницу. Например, для поиска страниц со ссылками на узел Microsoft, необходимо в качестве запроса ввести *link:www.microsoft.com*. Ключевое слово *url*: позволяет искать среди страниц, в адресе которых существует заданная в качестве аргумента часть. Если вы хотите найти все страницы, в адресе которых есть название фирмы Intel, то следует ввести запрос *url:intel*. Эти ключевые слова можно использовать вместе. Например, если мы хотим найти все страницы Российской части Интернета, на которых имеются ссылки на систему Alta Vista, следует в качестве запроса указать *link:altavista.digital.com url:ru*.

В подавляющем большинстве случаев вы сможете найти необходимую информацию с помощью простых запросов, но иногда могут потребоваться и более сложные. Система Alta Vista позволяет формировать запросы с использованием логических операций. Они называются сложными, или расширенными.

Для построения сложного запроса используются логические операторы и синтаксические выражения. Попытаемся пояснить, что это такое.

Синтаксические выражения – это любые слова и словосочетания, аналогичные рассмотренные в простых запросах. Логические операторы служат для выполнения операций «И», «ИЛИ», «НЕТ» и «ОКОЛО» над синтаксическими выражениями. Часто синтаксические выражения, над которыми производятся логические операции, называются аргументами. В синтаксических выражениях вы также можете использовать отдельные слова или фразы, заключая несколько слов в двойные кавычки. Правила использования заглавных и прописных букв в сложных запросах так же не отличаются от данных правил в простых запросах.

Главное, что отличает сложный запрос – это использование логических операторов и круглых скобок. С помощью операторов и скобок вы создаёте из отдельных синтаксических выражений необходимый запрос.

Рассмотрим несколько примеров. Перед примерами перечислим допустимые в расширенных запросах логические операторы:

- Оператор AND для логической операции «И». Оператор гарантирует, что в документе присутствуют оба документа.
- Оператор OR для логической операции «ИЛИ». Оператор гарантирует, что в документе присутствует хотя бы один из аргументов.
- Оператор NEAR для операции «ОКОЛО». Данный оператор гарантирует, что аргументы стоят друг от друга в тексте не дальше, чем на десять символов.
- Оператор NOT для логической операции отрицания, то есть для операции «НЕТ». Этот оператор гарантирует, что аргумент не присутствует в документе.

Чтобы в найденном документе отсутствовало какое-либо слово, надо перед ним в запросе поставить оператор отрицания. Полезно для удобства чтения запроса выделить слово с отрицанием круглыми скобками: (NOT описание).

Для задания одновременного присутствия в искомом документе сразу нескольких слов или словосочетаний используется логический оператор «и».

Например, если вы хотите, чтобы в документе было слово «провайдер» и словосочетание «очень дешево», то следует в запросе указать *провайдер AND «очень дешево»*. В качестве аргумента можно использовать не только слова и словосочетания, но и логические выражения, например по запросу *интернет AND (NOT описание)* будут найдены документы, в которых имеется слово «интернет» и нет слова «описание». В простом запросе для этого мы использовали бы *+интернет-описание*. Если вы хотите, чтобы в искомом документе находилось хотя бы одно из нужных слов, то следует воспользоваться логической операцией «ИЛИ». Например, запрос *телевизор OR радио OR видео* найдёт документы, в которых есть хотя бы одно из перечисленных слов.

Более сложный запрос *«дешевый доступ в Интернет» OR (провайдер AND «очень дешево»)* задаст поиск документов, в которых есть словосочетание «дешевый доступ в Интернет» или встречаются одновременно слова «провайдер» и «очень дешево».

Операция «ОКОЛО» не является классической логической операцией. С её помощью можно задать, то, чтобы слова находились рядом друг с другом. Например, *провайдер AND «очень дешево»* указывает, что данные слова просто присутствуют, однако они могут располагаться в разных концах документа. Запрос же *провайдер NEAR «очень дешево»* описывает, что слово «провайдер» и словосочетание «очень дешево» должны находиться друг от друга недалеко, то есть между ними не должно находиться более десяти слов.

С помощью логических операций можно создавать достаточно сложные запросы, например:

((провайдер NEAR московск*) OR ((NOT описание) AND интернет*)) AND «очень дешево»*. Разберём этот запрос по частям. Выражение *(провайдер* NEAR московск*)* описывает, что в искомом документе должно быть слово, начинающееся с «провайдер», то есть «провайдер», «провайдера», «провайдеру» и тому подобное, причём это слово должно находиться рядом со словом, которое начинается с «московск». Этим мы ограничиваем поиск московскими провайдерами. В других скобках выражении *((NOT описание) AND интернет*)* описывает, что в документе должно быть слово, начинающееся с «интернет», и не должно присутствовать слово «описание». Рассмотренные два выражения соединяются логическим

оператором *OR*, то есть мы ищем документ, в котором или описываются московские провайдеры, или имеется информация об Интернете, которая не является описанием. Такой поиск осуществляется по запросу *((провайдер* NEAR московск*) OR ((NOTописание) AND интернет*))*. В нашем же запросе ещё присутствует последняя часть *AND* «очень дешево». С её помощью мы задаём, что в дополнение к предыдущим условиям, необходимо, чтобы в найденном документе присутствовало словосочетание «очень дешево». Таким образом мы сформировали достаточно сложный запрос, хотя, скорее всего, по нему вы не найдёте не одного документа.

Пример: зададим более простой запрос. С помощью части запроса *(интернет* OR internet)* мы описываем русское название Интернета в любых склонениях или его английское название. Чтобы искать провайдеров Интернета, мы описываем, что рядом со словом «Интернет», в любом виде, должно находиться слово «провайдер». Теперь наш запрос выглядит так:

(провайдер NEAR (интернет OR internet))*. По нему мы найдём информацию о провайдерах Интернета. Если же мы хотим ограничить информацию только московскими провайдерами, то следует в запрос добавить *AND моск**. Этим мы описываем, что в документе обязательно должно присутствовать слово, начинающееся с «моск*». В итоге мы получили запрос *(провайдер NEAR(интернет* OR internet)) AND моск**.

В этом примере я искал русские слова в Интернете. Для этого можно воспользоваться системой Alta Vista, но более удобно использовать русские поисковые системы. Если же вы хотите найти зарубежную информацию, система Alta Vista становится просто незаменимой.

Здесь я подробно описал поисковую систему Alta Vista по двум причинам. Во-первых, она является крупнейшей в мире, и вам наверняка придётся воспользоваться её услугами. Кроме того, построение запросов в ней аналогично построению запросов для других систем поиска.

2.2. Использование системы Yahoo

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.

Заглавная страничка Yahoo! (Рис.2.2.) грузится очень быстро - хотя на ней очень много ссылок, но все они текстовые. При вводе ключевых слов с основной страницы Yahoo, запрос обрабатывается по методу "Intelligent default", то есть Yahoo! ищет наиболее подходящие результаты в таких областях: в категориях Yahoo, в Web-сайтах, зарегистрированных на Yahoo; на Altavista (запрос передается при отсутствии результатов); в новостях.

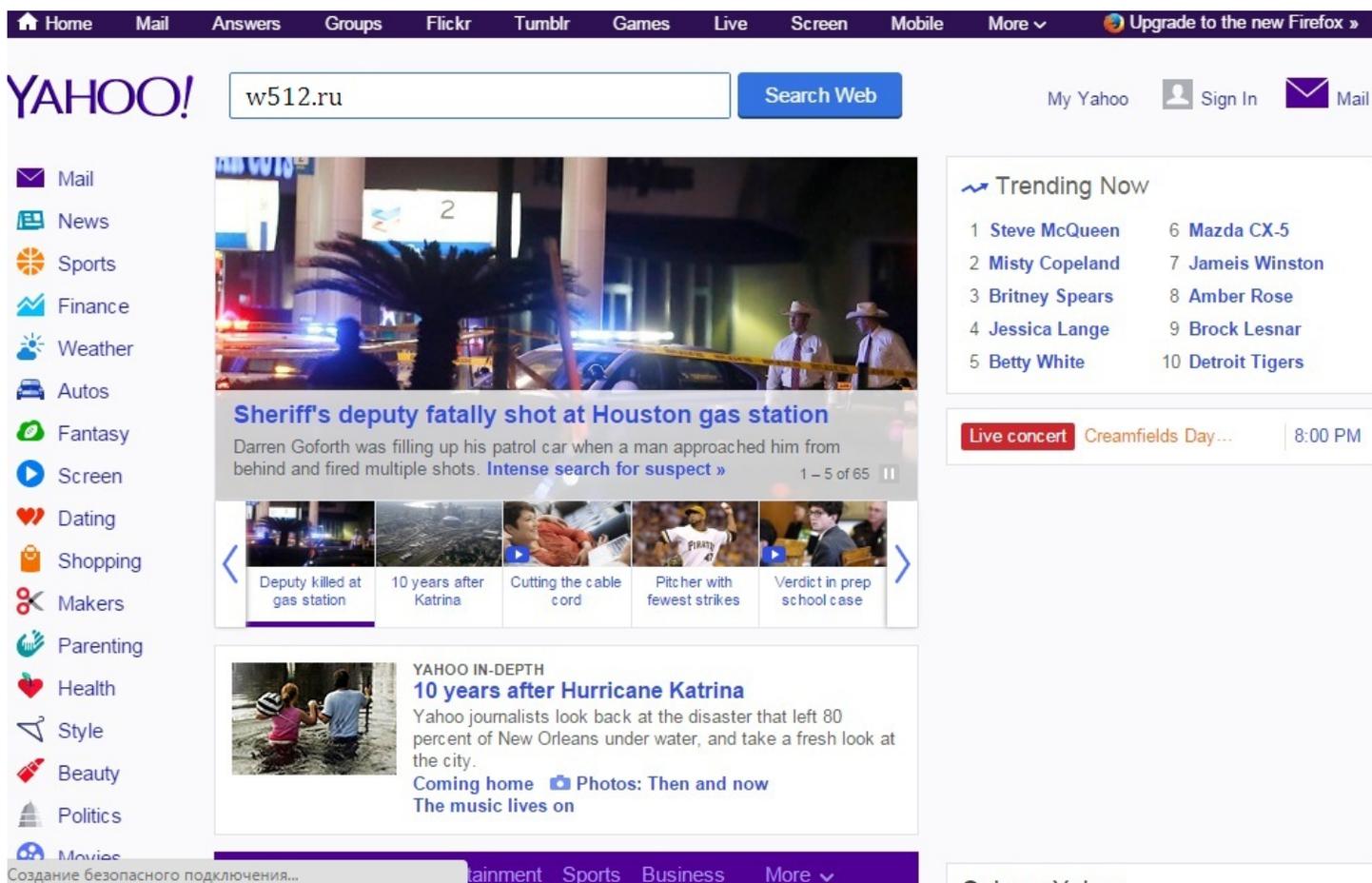


Рис. 2.2 Начальная страница поисковой системы Yahoo

При задании критериев поиска для Yahoo! нужно помнить, что Yahoo! ищет эти слова только в названии и описании страницы, поскольку полнотекстового индекса на Yahoo! нет. Поэтому не следует указывать при поиске слишком много терминов или синонимов - количество результатов с Yahoo! снизится или даже будет нулевым.

Поиск в системе Yahoo реализован с использованием Alta Vista. Все описанные выше приёмы, которые мы применяли при работе с Alta Vista, применимы и к Yahoo. Таким образом, вы можете использовать поиск в Yahoo совершенно так же, как и в системе Alta Vista.

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. Если вам удобнее не составлять запрос, а выполнять последовательные уточнения, то эта система вас должна устроить.

Поисковая система Google

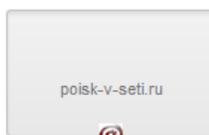
+Сергей Почта Картинки    

Google

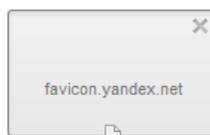
Россия



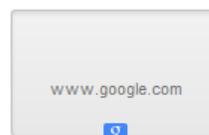
Как искать в интернете.



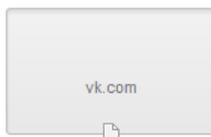
поисковые системы гугл...



poisk-v-seti.ru (1x1)



favicons (16x16)



Сокращение ссылки



Добро пожаловать в Go...



Интернет-магазин Chrome

Рис. 2.3 Начальная страница поисковой системы Google

За почти 20 лет существования корпорации Google, он приобрел статус крупнейшей поисковой системой в мировом масштабе.

Google (Рис 2.3.) осуществляет поиск по документам на более чем 35 языках, в том числе русском. В настоящее время многие порталы и специализированные сайты предоставляют услуги поиска информации в Интернете на базе Google, что делает задачу успешного позиционирования сайтов в Google еще более важной.

Поисковая система Google выглядит стабильнее своих конкурентов в плане смены алгоритма и апдейтов. Информация, только что размещенная на сайте, может в считанные минуты попасть в основную выдачу. Поисковые роботы Google в три раза быстрее, чем роботы других поисковых систем.

Google первым применил новую технологию поиска и выдачи результатов - релевантность интернет-страниц - то, что не лежит на данной конкретной странице, но тем ни менее влияет на ранг ее выдачи. Поисковая система Google учитывает гипертекстовую структуру Интернета, и какие страницы на какие ссылаются. Google сама определяет релевантность интернет-страниц. Страница, на которую ссылаются чаще, более релевантна и значит более популярна.

Особенность Google заключается в том, как и по какому критерию эта поисковая система классифицирует результаты поиска на своей странице. Используемый алгоритм PageRank сортирует все веб-страницы по критерию смыслового соответствия.

Алгоритм PageRank зависит от нескольких факторов:

- От частоты повторения и местоположения ключевых слов на веб-странице - если искомое слово или фраза только один раз встречается на сайте, то страница получает низкий бал.
- От времени существования страницы - каждый день в Интернете появляются новые сайты, но не многие задерживаются надолго. Поэтому преимущество отдается уже зарекомендовавшим себя сайтам, существующим долгое время.
- От количества веб-страниц, связанных с «главной страницей» - Google «смотрит» на количество веб-страниц, относящихся к определенному сайту и определяет ее рейтинг среди всех остальных.

Помимо просто поиска, здесь можно сравнивать цены на товары в интернет-магазинах, читать новости и многое другое. Есть и служба блокировки назойливой

интернет-рекламы

2.3 Поисковая система Rambler

Рамблер — одна из первых поисковых систем российского интернет-пространства. Несмотря на свой солидный «возраст», конкуренции с сильными соперниками она не выдерживает, сегодня система охватывает только 0,4% поиска. (Рис 2.4.)

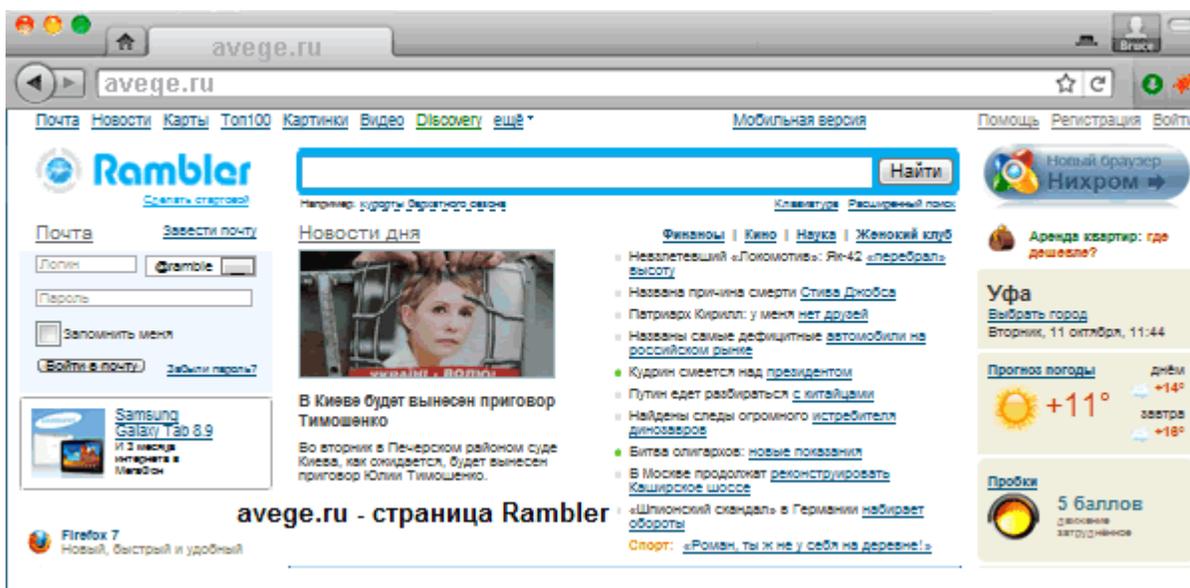


Рис. 2.4 Начальная страница поисковой системы Rambler

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

В июне 2003 года компания запустила новую версию поисковой машины, которая отличается от предыдущей по двум основным параметрам. Первое и главное – значительно увеличилась скорость поиска; второе, благодаря новой архитектуре системы, обновление поискового индекса происходит несколько раз в день.

Теперь поисковик всегда находит самые свежие документы и последние новости. Рамблер знаком с русским языком, понимает термины типа «б/у», «у.е.» и «а/я». Как дань моде, открылась лаконичная версия поисковой системы по адресу «r0.ru».

Содержание ее стандартно – строка поиска, почта и новости.

Еще одна особенность поисковика Rambler, это механизм ассоциаций. Помимо стандартной ответной страницы, в которой найденные документы расположены в порядке убывания релевантности, появляется строка - «У нас также ищут». В ней приведены слова и словосочетания, тематически связанные с запросом. Новый механизм также исправляет ошибки запросов. На введенное gjujlf пользователь получит теперь положительный результат: сведения о погоде (gjujlf – это слово «погода» набранное латиницей).

Необходимо отметить проекты Rambler-Группы и Rambler-Фото. Оба этих проекта предоставляют пользователям сети возможность использования заложенных в проекты программных алгоритмах в собственных целях. Проект Rambler-Группы дает людям возможность начинать обсуждения самых разных тем, которые существуют не только в виде форумов, но также и в виде открытых списков рассылки. Проект Rambler-Фото представляет пользователям сети возможность создания фотоальбомов в объеме до 20 мегабайт, причем это дисковое пространство зарегистрированные пользователи поисковой системы Rambler получают в дополнение к тем мегабайтам почты, на которые они могут претендовать по новым условиям e-mail обслуживания (15 мегабайт).

Rambler Mass Media - первый и по-прежнему лидирующий на рынке информационный дайджест. Помимо текстовых материалов и фоторепортажей, ставших визитной карточкой проекта, Rambler Mass Media представляет пользователям аудио и видеосюжеты. В числе поставщиков - ВГТРК и BBC, Deutsche Welle, радио "Маяк" и др. Заголовки важнейших новостей размещаются на первой странице портала Rambler.ru и в блоке "Главные новости дня" на главной странице Rambler Mass Media. Наиболее актуальные новости могут читать и посетители других сервисов Рамблера - Рамблер Почты, Погоды на Рамблере.

2.4 Поисковая система Aport

Поисковая система Апорт (Рис 2.5.) на сегодняшний день находится на 16 месте по популярности поисковых машин Рунета. Данная система осуществляет полнотекстовый поиск документов с учетом морфологии русского языка в поисковых запросах.

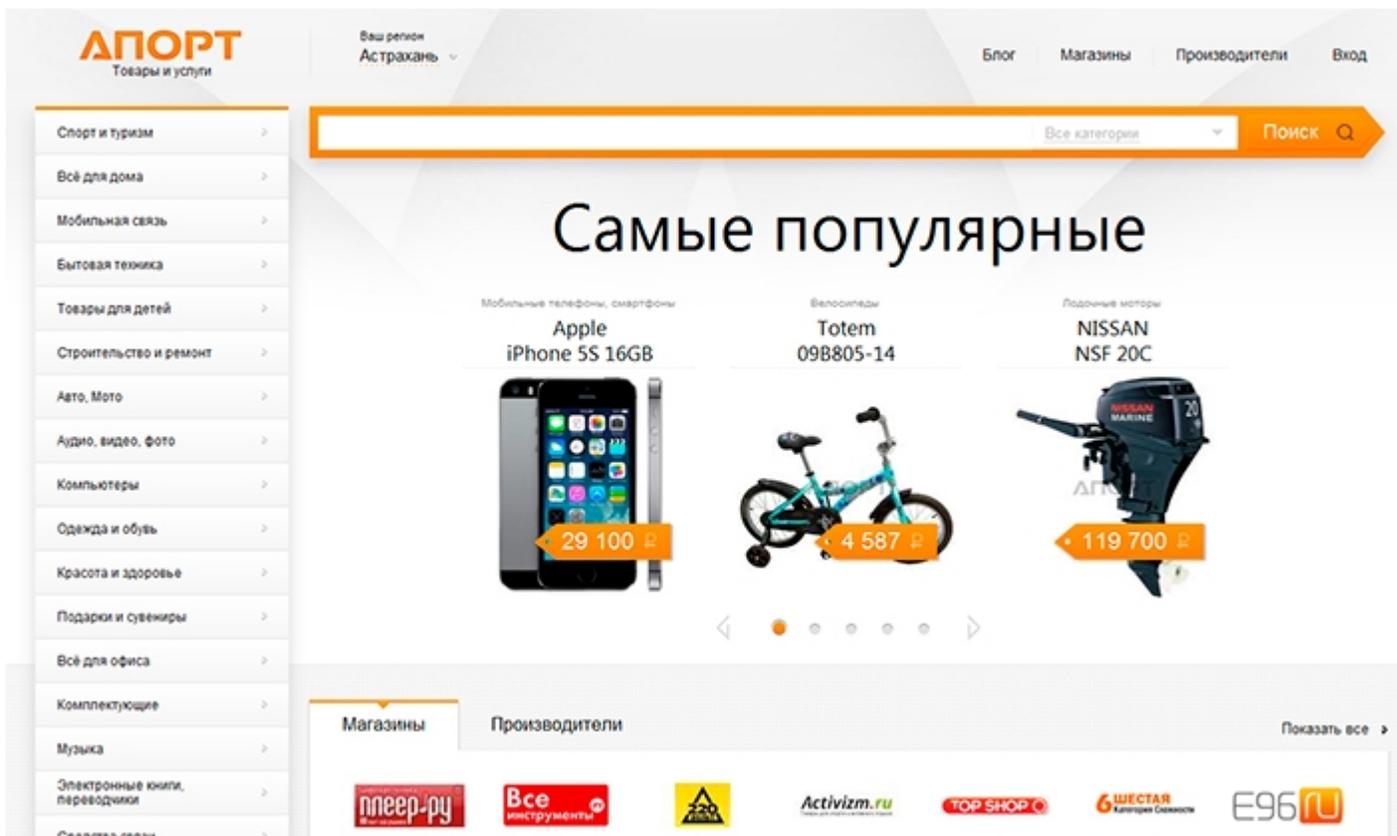


Рис. 2.5 Начальная страница поисковой системы Aport

К важным свойствам первой версии поисковой системы Апорт можно отнести перевод запроса и ответа на английский язык и обратно. Второе свойство - реконструкция проиндексированных страниц из собственной базы. Это дает возможность просмотра уже несуществующих страниц.

www.aport.ru - первым из поисковых систем Рунета реализовал базовые технологии Google. "Page rank" - характеристика популярности ресурса по формуле "обратных ссылок": ссылки с других сайтов на данный ресурс. Причем учитывается не только количество, но и важность ссылок. В поисковой системе Апорт впервые была реализована возможность поиска по новостным лентам.

В настоящее время возможности поисковой системы "Апорт" не так обширны, но все-таки значительны. На официальном сайте поисковика можно найти каталог, охватывающий добрую часть русскоязычных Интернет-ресурсов, с удобным рубрикатором и системой выдачи.

Также, на "Апорт" представлена актуальная информация различного рода: это и программа телевидения, и новости, и прогноз погоды, и котировки валют.

Особый блок сервисов поисковика составляют почта, справочная информация (в разделе "Словари") и поиск объекта по адресу (для Москвы, Санкт-Петербурга, а также для федеральных округов РФ: Дальневосточного, Приволжского, Северо-Западного, Сибирского, Уральского, Центрального и Южного).

С помощью "Доски объявлений" от "Апорт" можно приобрести, продать, подарить или принять в дар все, что угодно: от котят до земельных участков. Данный сервис предлагает еще и поиск работы, однако найти здесь настоящие вакансии среди сомнительных объявлений довольно сложно.

2.5 Поисковая система Яндекс

На сегодняшний день Yandex (Рис 2.6.) это одна из самых востребованных поисковых систем в Рунете. Поиск позволяет искать документы на русском, татарском, украинском, белорусском, казахском, английском, немецком и французском языках с учётом морфологии этих языков и близости слов в предложении.

Новости в Москве 26 марта, четверг 12:12

1. Доллар и евро резко подешевели на фоне роста цен на нефть
2. Власти Франции не комментируют утечку в СМИ данных о крушении А320
3. События в Йемене обеспокоили импортеров нефти
4. Нефть дорожает на опасениях конфликта на Ближнем Востоке
5. Две автоколонны МЧС России с гумпомощью Донбассу пересекли границу

Биржевые курсы: USD -0,87 56,58 EUR -0,71 62,34 Нефть +4,42% 58,81

Расписание транспорта в России и за рубежом
Яндекс.Расписания

Яндекс Поиск Карты Маркет Новости Словари Картинки Видео Музыка Перевод ещё

Яндекс — лидер Рунета

Найдётся всё Например, билеты самолет дешево [Скачайте Яндекс.Браузер](#)

Карта Москвы

Погода ☁ 0 °C
вечером +2, ночью 0

Пробки 🟡 4 балла
Местами затруднения
Прогноз 4 4 5 5 5
13 ч 14 15 16 17

Телепрограмма Афиша

	сегодня	завтра
USD ЦБ	57,3879	-0,9608 56,4271
EUR ЦБ	62,7651	
Нефть	26/03 +4,42% 58,81	

- Яндекс.Браузер — с Умной строкой
- Маркет — распродажи платьев
- Авто — новые до 380 тыс.
- Такси — за 7 минут
- Работа — зарплаты в Москве
- Недвижимость — новостройки в области
- Мастер — мойка окон
- Деньги Фото Диск

Рис. 2.6 Начальная страница поисковой системы Яндекс

Поисковый запрос от пользователя, после анализа загруженности поисковой системы, отправляется на наименее загруженный сервер. Для обеспечения такой возможности сервера Яндекса объединены в кластеры, и даже кластеры кластеров. Затем пользовательский запрос обрабатывается программой под названием «Метапоиск». Метапоиск осуществляет анализ запроса в реальном времени: определяет географическое положение пользователя, проводит лингвистический анализ и т. д. Также, программа определяет, относится ли запрос к категории наиболее популярных или недавно заданных. Выдача на такие запросы некоторое время хранится в памяти (кэше) метапоиска, и в случае совпадения показываються заранее сохранённые результаты. Если запрос является редким и совпадений в кэше не найдено, система перенаправляет его на программу «Базового поиска». Тот анализирует индекс системы, также разбитый по разным дублирующимся серверам (это укоряет процедуру). Затем полученная информация снова попадает на метапоиск, данные ранжируются и показываються пользователю в готовом виде.

Основная особенность системы Яндекс, делающая популярной ее среди русскоязычных пользователей, – это способность определять различные словоформы с учетом морфологических особенностей русского языка.

В 2012 г. Был внедрен новый алгоритм «Калининград», который позволил сделать поиск персонализированным. Это значит, что теперь основным фактором ранжирования является сам пользователь, который формирует запрос для поисковой системы. Другими словами, результаты поиска по одному и тому же вопросу для двух разных пользователей теперь будет различаться в зависимости от нужд и предпочтений самих пользователей. Это новый шаг на пути эволюции поисковых систем.

В 2016 г. компания запустил сервис Яндекс Такси.

Кроме того, среди нововведений Яндекса можно отметить:

- введение геозависимости 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.

2.6 Поисковая система Mail. ru

Поисковик Mail.ru остался позади известных Яндексa и Google. Сначала этот сервис работал, используя поисковые услуги Google, позже — Яндексa, и только в 2013 году Mail.ru стал применять свой поисковой движок. (Рис 2.7.)

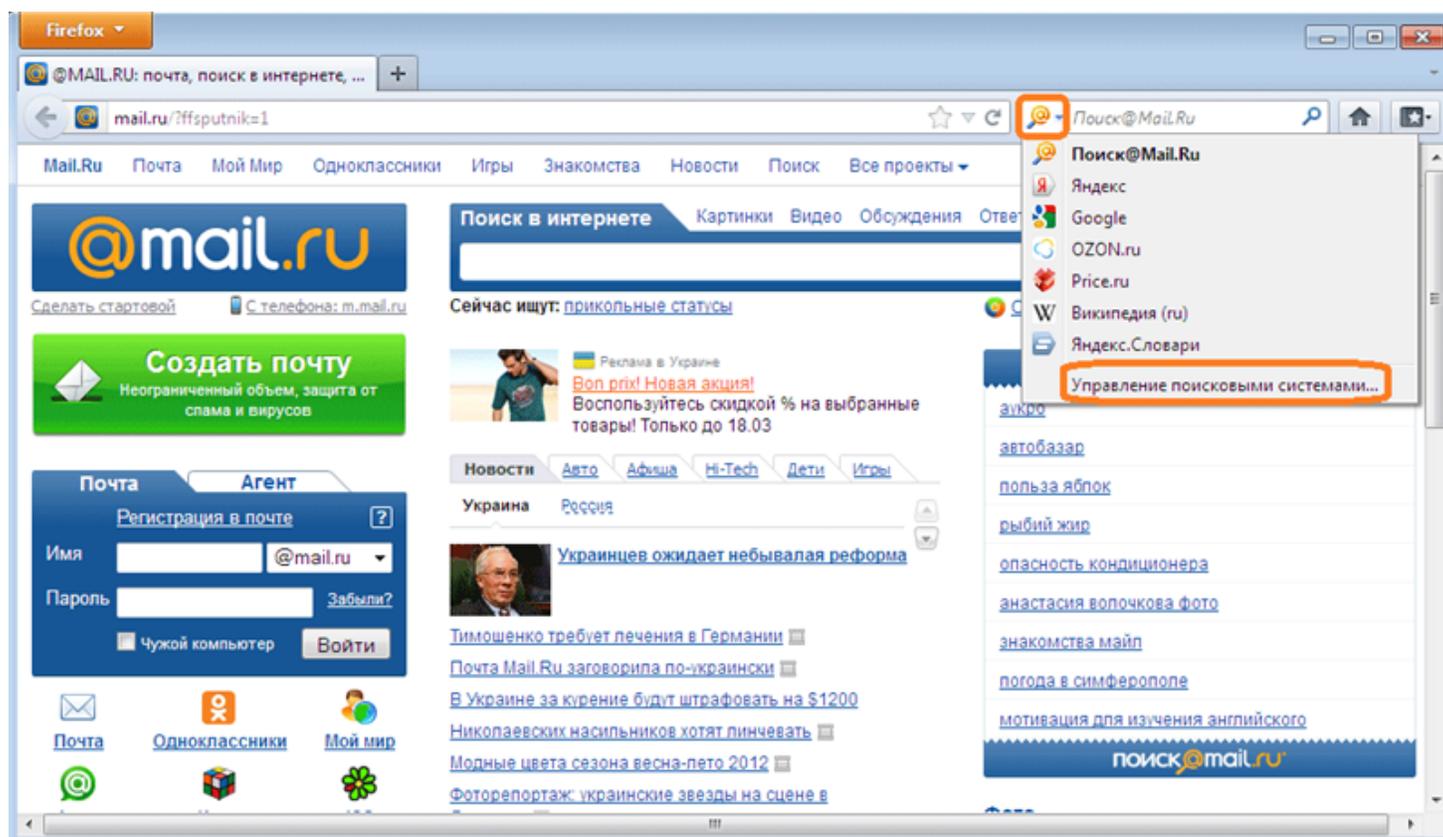


Рис. 2.7 Начальная страница поисковой системы Mail.ru

Национальная почтовая служба Mail. ru – это не только поисковая система но и один из крупнейших порталов российского Интернета. Ежедневная аудитория Mail.ru - более 5 миллионов пользователей. Общее число регистраций со дня основания около 60 миллионов. Mail.ru - самый быстроразвивающийся российский Интернет-ресурс. Через почтовые ящики Mail.ru ежедневно проходит более 25 миллионов писем. Mail.ru занимает лидирующую позицию среди бесплатных почтовых сервисов, предоставляя своим пользователям почтовый ящик неограниченного размера с защитой от спама и вирусов, переводчиком, проверкой

правописания, архивом для хранения фотографий и многое другое.

Нужно сказать, что разработчики поисковой системы изначально делали ставку не на поиск, а на различные сервисы: Агент, Почта, Мой Мир, Одноклассники, и самый главный способ монетизации — игры. Именно за счет этих услуг холдинг, по некоторым данным, является одним из самых посещаемых в России.

Выделить основные приоритеты в работе поисковых алгоритмов довольно сложно, но можно отметить несколько факторов:

- качественный контент;
- поведенческие показатели;
- удобный и интуитивно понятный для посетителей сайт;
- наличие естественных ссылок с других сайтов.

Соблюдение этих очевидных правил поможет сайту уверенно занять позиции в первой десятке. По сравнению с гигантами Яндексом и Google, продвигать сайт на Mail.ru гораздо легче и проще.

Какая из поисковых систем лучше?

Рейтинг мировых поисковых систем по данным 2016 г. представлен в следующей таблице.

<< Ноя 16

декабрь 2016 г.

отчет: переходы из поисковых систем

по дням | по неделям | по месяцам

значения: среднесуточные / суммарные	декабрь 2016 г.		ноябрь 2016 г.		в среднем за 3 месяца	
<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	

Как мы видим лидерами по показателям качества представленной информации оказались Яндекс, Google .

Мною был введен один и тот же «запрос» в несколько поисковых систем.

Yandex оказался одной из наиболее эффективных систем с точки зрения ее релевантности и соответствия выданных результатов заданному запросу. Хотя страниц было много, но нужная информация находилась на самых первых из них. Мало затраченного времени – необходимые результаты. При этом немаловажную роль сыграла также относительная новизна представленной информации.

Google выдавал результаты страниц, на которых не всегда первое место занимали релевантные документы. Зато жалоб на разнообразие просто не было, т.к. в представленном количестве материала можно было найти что угодно (при этом было потрачено времени в два раза больше, по сравнению с поисками в других поисковых системах).

Система Aport оказалась менее эффективной, чем вышеназванные из-за ее чрезмерной ориентации на частные случаи, но результаты, которые она выдавала, значительно отличались от результатов других поисковых систем. Они были единственные в своем роде, не всегда релевантны, но неповторимы.

Rambler, несмотря на прочно занимаемое четвертое место в количественном рейтинге, оказался намного менее эффективной по релевантности системой. Здесь преобладает ориентация на российские источники информации, что снижает ее адекватность в оценке ситуации в других странах. К этой же категории по степени релевантности можно отнести и поисковую систему Google.

Поисковую систему Yahoo можно рассматривать как наиболее эффективную наравне с Yandex, но только в англоязычном поиске. На русском языке в данной системе имеется незначительное количество сайтов и их релевантность минимальна.

Заключение

По итогам сделанной мной работы я могу заключить что; поисковые системы уже давно стали неотъемлемой частью Интернета. Поисковые системы сейчас – это огромные и сложные механизмы, представляющие собой не только инструмент поиска информации, но и заманчивые сферы для бизнеса.

По моему мнению, самой лучшей иностранной поисковой системой является Google, так как для меня основное значение имеет точность и полнота предоставляемых данных. Но можно заключить также что, каждая поисковая система будь то Российская или зарубежная предоставляет различные возможности поиска, из различных баз данных, поэтому сказать точно какой именно лучше пользоваться было бы неправильно. Поэтому для удобства поиска и полноты информации следует пользоваться несколькими поисковиками вводя в них нужные запросы. По моему мнению, из многих Российских поисковиков выделяются Яндекс и Рамблер, для них характерно постоянное обновление баз данных что, обеспечивает именно актуальность и точность предоставляемой информации.

Список литературы

- ○ 1. Экслер, А.Б., "Самоучитель работы в Интернете" - Москва.: NT Press, 2010г.
- 2. Кузьмин А.В. Золотарева Н.Н. Поиск в Интернете - Санкт - Петербург.: Издательство НиТ, 2011г.
- 3. Егоров, А.Б., " Поиск в Интернете" - Санкт - Петербург.: НиТ, 2010г.

4. Павел Храмцов "Поиск и навигация в Internet".
<http://www.osp.ru/cw/1996/20/31.htm>
5. Андрей Аликберов "Несколько слов о том, как работают роботы поисковых машин". http://www.citforum.ru/internet/search/art_1.shtml
6. www.google.ru
7. www.yandex.ru
8. www.rambler.ru
9. www.aport.ru
10. www.mail.ru
11. www.yahoo.ru