



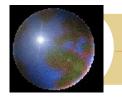
Поиск информации в интернете



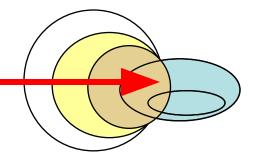
• Это занятие посвящено поисковым системам.



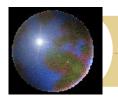
Классификация поисковых средств по В.В. Дудихину



Поисковые системы

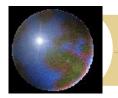


- Как видно на схеме, поисковые системы (ПС) имеют пересечение с «продвинутыми» каталогами.
- Многие современные поисковые ресурсы совмещают в себе возможности каталога и ПС, позволяя искать информацию не только в содержимом каталога, но и во всем интернете.



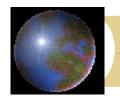
Поиск с помощью ПС

- Работа пользователя с ПС основывается на формировании запроса, по которому происходит отбор нужных документов из базы данных.
- Запрос формируется с помощью ключевых слов (одного или нескольких).
- Результаты поиска выдаются пользователю в виде списка адресов (гиперссылок) и краткой аннотации к ним.



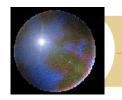
Ключевое слово

- Это лексическая единица, являющаяся существительным, прилагательным, глаголом, числительным, наречием или местоимением, которая в наибольшей степени отражает содержание всего искомого документа.
- При формировании запроса могут использоваться не только отдельные ключевые слова, но и словосочетания, состоящие из нескольких ключевых слов.



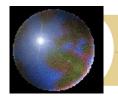
3anpoc

- Запрос это набор соединенных операторами ключевых слов, с помощью которых поисковая система автоматически ведет поиск и отбор необходимых документов.
- Другими словами, запрос это инструкция (команда) для ПС на поиск нужных документов.



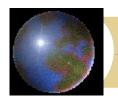
Виды запросов

- Запросы бывают двух типов: простые и сложные (или расширенные, advanced).
- Простые запросы состоят из отдельных ключевых слов или словосочетаний.
- Сложные запросы, кроме ключевых слов, содержат логические и другие операторы.



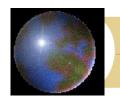
Работа ПС

- Когда идет обслуживание конкретного запроса, ПС сравнивает ключевые слова, введенные пользователем, с ключевыми словами, полученными в процессе индексации и хранящимися в базе данных.
- При совпадении этих слов пользователю выдается адрес данного документа.
- Процедура поиска по ключевым словам очень напоминает работу с обычной книгой, в которой есть алфавитный указатель.



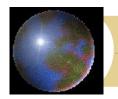
Поисковые системы

- Для реализации поиска ПС проводит индексацию сайтов интернета.
- Индексация это процедура автоматического создания базы данных, в которой хранятся ключевые слова, аннотации документов и адреса, по которым размещены эти документы.
- В базе данных каждому документу ставится в соответствие свой набор ключевых слов и адресов.



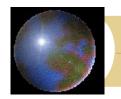
Роботы на службе ПС

 Индексация документов, размещенных на различных серверах, производится поисковыми системами автоматически с помощью специальных программ – роботов.



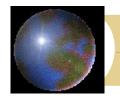
Роботы на службе ПС

- Программа-робот автоматически периодически «посещает» сайты и собирает сведения о их содержимом и местоположении.
- Эта программа постоянно сканирует сеть, запоминает расположение информации на сайтах, чтобы потом показать пользователю точное место хранения документа.



Скорость поиска

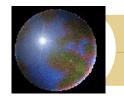
- Фактически роботы начинают поиск задолго до обслуживания конкретного запроса.
- Этим объясняется потрясающе высокая скорость поиска нужной информации.



Алгоритм работы ПС

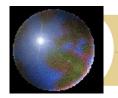
• Таким образом, можно схематически изобразить алгоритм работы поисковых систем:





Как работает ПС

• Более детально работу поисковой системы мы рассмотрим на следующем занятии.



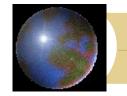
Тенденции развития ПС

- Современные наиболее быстродействующие ПС стремятся произвести индексацию всего документа, а не только его названия и первых предложений текста.
- Наиболее совершенные роботы при индексации сканируют не только главную страницу, но и по гиперссылкам заходят вглубь сайта.





Зарубежные лидеры поисковых систем

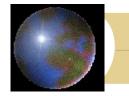


Зарубежные лидеры ПС

- Google www.google.com
- Yahoo! Search search.yahoo.com
- AltaVista www.altavista.com

www.av.com

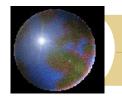
Alltheweb
 www.alltheweb.com



Статистика (кол-во документов)

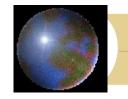
- Google более 8 млрд.
- Alltheweb более 600 млн.
- AltaVista более 550 млн.
- Yahoo!
 более 1.5 млн.

(в каталоге)



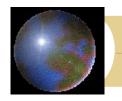
Google www.google.com

- Название поисковой системы происходит от измененного слова googol (гугол), которое означает величину 10^{100} , т.е. единицу со ста нулями.
- По мнение автора термина, математика Э. Каснера, во всей Вселенной нет ни одного материального объекта в количестве гугола единиц.



Google www.google.com

• Назвав свою систему Google и явно ассоциируя это название с термином «гугол», основатели ПС хотели выразить необъятный масштаб интернета.



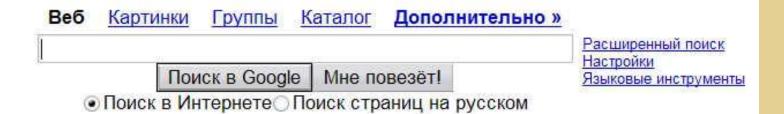
Google w

www.google.com

- ПС автоматически определяет из какого региона поступил запрос и в соответствии с этим отображается на одном из 26 языков.
- Так, существует сайт www.google.ru, куда автоматически отсылаются пользователи из России.

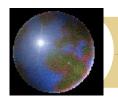
Google www.google.ru



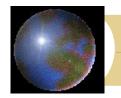


Рекламные программы - Bcë o Google - Google.com in English

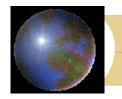
©2006 Google



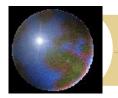
- Один из основателей Google Сергей Брин.
- Его семья уехала из СССР в 1979 году, когда Сергею было шесть лет.
- Закончив школу, в 1990 году он поступил в университет Мэриленд.
- Досрочно получив «красный» диплом бакалавра по математике и компьютерным системам, он продолжил учебу в Стэнфордском университете.



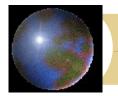
- Научная работа Сергея Брина, как и его друга Ларри Пейджа, касалась поиска информации.
- «Мы не читали в интернете гороскопов и объявлений. Нас интересовал поиск та информация, которая по-настоящему влияет на жизнь людей», рассказывает Брин.



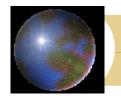
- Одну работающую программу поиска Сергей написал в 1994 году.
- Она автоматически искала на сайте «Плейбоя» новые картинки и закачивала их на рабочий стол в компьютере Брина.



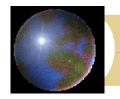
- Осенью 1996 года на сайте Стэнфордского университета заработала поисковая страница. Это была практическая часть научной работы Брина и Пейджа.
- А необходимый для работы поисковика сервер объемом 1024 Гб располагался прямо в комнате Брина в университетском общежитии.



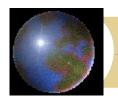
- В основе поисковой системы лежал запатентованный в 1996 г. алгоритм PageRank.
- Принцип действия: все многочисленные найденные страницы ранжировались по числу ссылающихся на них других страниц.
- Таким образом наверху оказывались самые востребованные страницы.



- Поисковая система оказалась настолько удобной, что к лету 1998 года к ней обращалось около 10 тысяч пользователей.
- Брин и Пейдж взяли академический отпуск и 7 сентября 1998 года основали компанию Google Inc.

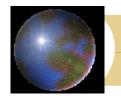


- В 1999 году Google стала самой популярной ПС и с тех пор ее рейтинг не уменьшался.
- В языке американцев появилось новое слово «гуглить», т.е. искать в интернете (как само собой разумеющееся) с помощью Google.



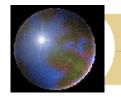
Поисковая система Google

- В США Google предпочитают 34,7% пользователей.
- В мире доля Google для англоязычного поиска достигает 43,3%!
- Google позволяет производить поиск web-документов, изображений, новостей, сообщений форумов, а также сайтов из собственного каталога



Yahoo! Search search.yahoo.com

- Самый близкий преследователь Google в рейтинге ПС компания Yahoo!
- Традиционно служба Yahoo! позиционировалась как развитый каталог ресурсов интернета, но возможности поиска все же присутствовали: до 2004 года в каталоге использовалась поисковая технология Google.



Yahoo! Search search.yahoo.com

- С 2004 года начала работу глобальная ПС Yahoo! Search.
- ПС позволяет искать web-документы, изображения, новости, сообщения форумов. Возможет поиск и в собственном каталоге.
- Существует также уникальная технология борьбы с поисковым мусором и фильтрация избыточных ссылок.

>> Try the ALL-NEW Yahoo! home page YAHOO! Music !* My Yahoo! Finance !* Messenger Web Images | Video | Audio | Directory | Local | News | Shopping Select Category: Advanced Yahoo! Search My Web Ad Programs Search the Web: Vahoo! Answers: Ask a question | Answer questions Yahoo! Travel: Flights, Hotels, Cars, Vacations, Try Trip Planner Check your mail status: Sign In

Free mail: Sign Up 360° Horoscopes Shopping Music Autos HotJobs My Yahool Sports Finance Kids News Tech Games Local People Search Travel GeoCities Mail Personals TV Groups Maps Photos Yellow Pages Health Movies Real Estate All Y! Services...

Yahoo! Movies » Showtimes Clerks II **Now Playing** Also in Theaters Coming Soon You, Me and Dupree Pirates of the Caribbean Clerks II Little Man Superman Returns Lady in the Water Mini's First Time The Devil Wears Prada My Super Ex-Girlfriend

Yahoo! Small Business

Web Hosting Sell Online Domain Names Search Listings

Yahoo! Featured Services

Personal Website Downloads Fantasy Football Mobile

Entertainment

» More Entertainment

Eager to tour

- Britney Spears . Britney Spears 'can't wait' to get onstage
 - Comedian Red Buttons dies at age 87
 - · Rivers Cuomo says Weezer is 'done'

 - · Primetime Emmy Awards Who should win?

Buzz Log - What the world is searching for

» More Buzz



Gruseome new video emerging from Iraq sparks a new wave of searches on the bloody conflict. More... Popular News Searches

- 1. Iraq War
- 2. Israel

- 3. Lebanon
- 4. Valerie Plame

Yahoo! Web Directory

» More Yahoo! Web Directory

Puppy L Yahoo! Games

Play Puppy Luv, Poker Superstars II, Plantasia, Magic Match, and more.

In the News



Israel batters Lebanon

- Iran warns against attacking Syria
- Tourists flee Lebanon en masse Full Coverage - Photos - Video
- Big Dig possible defect count quadrupled Former CIA officer sues Cheney over leak
- Bush to tell Putin of democracy concerns
- Activists sue to block electronic voting
- Teenagers need specialized cancer care
- Naomi Campbell faces new employee lawsuit
- Tour de France · MLB · NBA · Soccer · Golf

News · Popular · Sports · FIFA World Cup · Stocks

Weather

Enter City or U.S. Zip Code



Go

✓ Save location on this page

Marketplace



Free credit score in seconds The average U.S. credit score is 678- See your score: \$0.

Countrywide Home Loans® - Credit problems? Low refinance rates - 4 out of 5 approved.

Degrees in as fast as 1 year: - Bachelor's. Master's, Associate's - Top schools,

Find great deals on all your Golf needs this summer at Yahoo! Shopping.

Yahoo! www.yahoo.com

Yahoo! Search search.yahoo.com



Web Images Video Audio Directory Local News	Shopping More »	
	Search the Web	Advanced Search My Web

Mail Status: Sign In

New User: Sign Up *

In the News

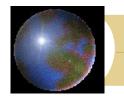
7:10am, Mon Jul 17 🗶

- Israel hammers at Lebanese infrastructure
- Quake triggers Indonesia tsunami, killing five
- Oil prices plunge on Iran report
- Storms may hinder fight against California fires
- · Fort Collins tops 'best places to live' list
- Everest remains a deadly draw for climbers

Markets: Dow +0.13% - Nasdaq -0.08%

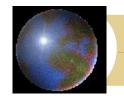
Advertising Programs - Search Services

Copyright © 2006 Yahoo! Inc. All rights reserved. Privacy Policy - Terms of Service - Copyright/IP Policy - Submit Your Site



Компания Yahoo!

• Компании Yahoo! с 2003 года принадлежат поисковые службы Inktomi, Fast и AltaVista.



AltaVista

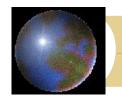
www.av.com

- Служба AltaVista появилась в 1995 г.
- Преимущество системы развитые, мощные средства сложного поиска
- Сервис осуществляет поиск на 25 языках.
- ПС не имеет собственного каталога и производит поиск web-документов, новостей, изображений, музыки и видео.

AltaVista

www.av.com

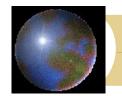




AltaVista

www.av.com

- Поисковая система AltaVista в отличие от других ПС не занимается морфологической обработкой текстов.
- Все слова для нее, независимо от языка, лишь последовательности символов. Это привлекает к AltaVista многих пользователей.
- ПС AltaVista принадлежит компании Yahoo!



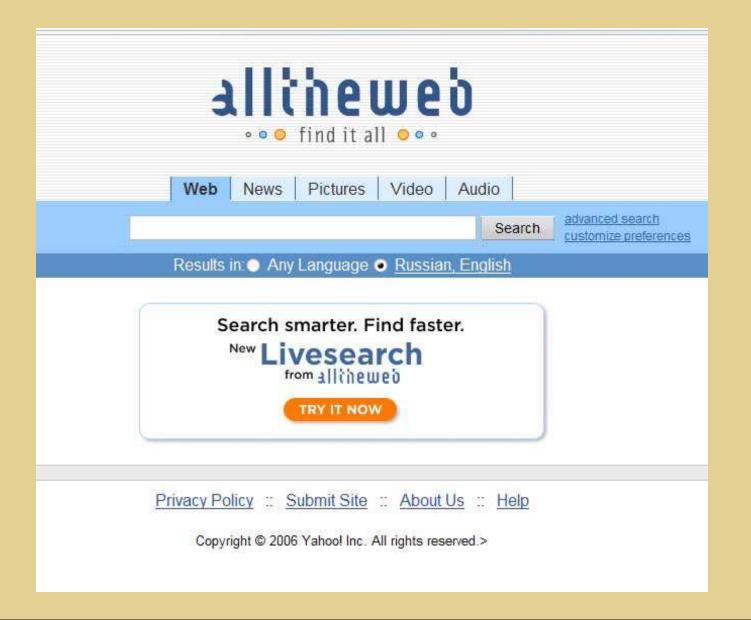
Alltheweb

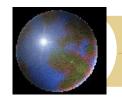
www.alltheweb.com

- Основана в Норвегии в 1997 году.
- В 2002 году на некоторое время ПС Alltheweb стала лидером по количеству документов в базе данных более 2 млрд.
- Сегодня поисковая система Alltheweb получила название Fast. Она считается наиболее близкой к Google по своим возможностям.

Alltheweb

www.alltheweb.com





Alltheweb

www.alltheweb.com

- ПС отличается высокой скоростью.
- Время ответа на поисковый запрос не превышает 0,05 секунды!
- Alltheweb обеспечивает поиск webдокументов, новостей, изображений, музыки, видео.
- ПС Alltheweb принадлежит компании Yahoo!





Российские лидеры поисковых систем



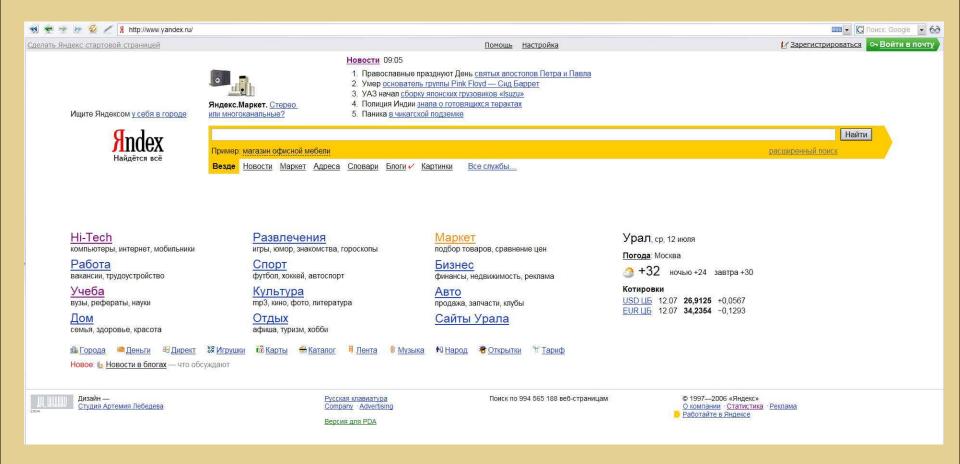
• Рамблер www.rambler.ru

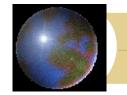
Aпорт www.aport.ru

- Официально поисковая машина Yandex.Ru была анонсирована 23 сентября 1997 года на выставке Softool.
- Слово «Япdex» было придумано за несколько лет до этого и означало «Языковой index», или, если по-английски, «Yandex» — «Yet Another indexer».

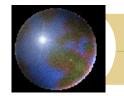
Яндекс

www.yandex.ru

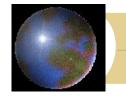




- Основными отличительными чертами ПС Яндекс были и остаются:
 - проверка уникальности документов (исключение копий в разных кодировках),
 - учет морфологии русского языка,
 - поиск с учетом расстояния,
 - оценка релевантности (соответствия ответа запросу).



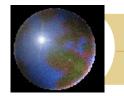
- В ноябре 1997 года был реализован естественно-языковый запрос.
- К Яндексу можно обращаться просто «порусски», задавать длинные запросы и получать точные ответы.
- Средняя длина запроса в ПС сейчас 2,7 слова.
- В 1997 году она составляла 1,2 слова, тогда пользователи поисковых машин были приучены к телеграфному стилю.



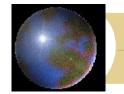
- В 1998 году на Яндексе появилась возможность найти
 - «похожий документ»,
 - список найденных серверов,

а также

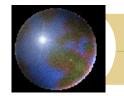
- поиск в заданном диапазоне дат,
- сортировка результатов поиска по времени последнего изменения.



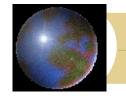
- В марте 2004 г. были реализованы новые возможности:
 - система может отличать мнения людей от технической, вспомогательной и рекламной информации;
 - система автоматически определяет, в каком городе находится пользователь и предлагает уточнение поиска по региону.



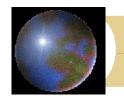
- Яндекс позволяет производить поиск
 - web-документов,
 - изображений,
 - новостей,
 - сообщений блогов,
 - товаров из интернет-магазинов,
 - адресов предприятий и организаций,
 - а также сайтов из собственного каталога



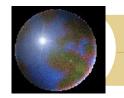
- ПС поддерживает шесть языков: русский, английский, украинский, белорусский, французский и немецкий.
- Язык документа определяется автоматически. Нужным языком можно ограничить область поиска.



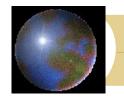
- С 2001 года, по опросам исследовательских компаний Gallup Media и Комкон, Яндекс является самым популярным поисковым ресурсом в рунете.
- Сегодня ежедневная аудитория Яндекса (включая зарубежных пользователей) – 12 млн. человек



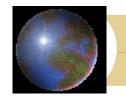
• Яндекс – не просто поисковая система или каталог, а весьма продвинутый интернет-портал, обладающий широким набором вспомогательных сервисов:



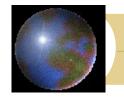
- Hapoд.Py www.narod.ru narod.yandex.ru
- Служба бесплатного хостинга и набор средств, позволяющий каждому желающему создать свой сайт с форумами, чатами, гостевыми книгами и бесплатной почтой.



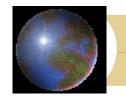
- Яндекс. Деньги money.yandex.ru
- Система оплаты цифровой наличностью.
- На виртуальный счет можно переводить деньги, переводить на другие счета, оплачивать услуги, снимать и обналичивать



- Яндекс. Новости news.yandex.ru
- Сервис предлагает автоматическое выделение главных тем дня в различных информационных агентствах и объединение их в информационные блоки.



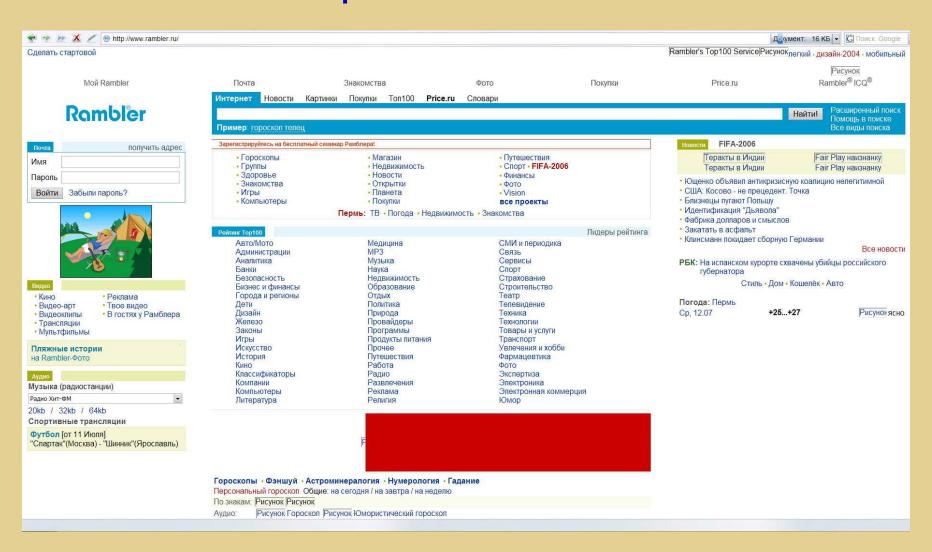
- Яндекс.Энциклопедии encycl.yandex.ru
- Поиск в энциклопедиях, справочниках и толковых словарях.

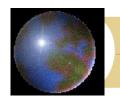


Рамблер www.rambler.ru

- В 1996 году программист Дмитрий Крюков написал поисковую программу для ресурсов интернета.
- Она сразу же была введена в эксплуатацию на www.rambler.ru
- Слово «rambler» означает «скиталец, странник, бродяга»

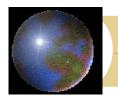
Рамблер www.rambler.ru





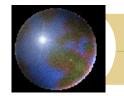
Rambler's Top100

• Весной 1997 года появляется Rambler's Top100 - уникальный рейтинг-классификатор, который не только оценивает на основе объективных данных популярность российских ресурсов, но и позволяет одним «кликом» попасть на них.



Rambler's Top100

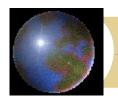
- Возникновение универсального счетчика, способного направлять аудиторию на разные ресурсы, создало в рунете динамичную конкурентную среду.
- Авторы сайтов стали более тщательно работать над своими сайтами, стремясь занять в Топ100 более высокие строчки.
- Хороший рейтинг давал дополнительные преимущества быстрый рост аудитории.



Рамблер

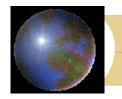
www.rambler.ru

- В июне 2003 года Рамблер запустил новую версию поисковой машины, которая отличается от предыдущей по двум основным параметрам:
 - во-первых, значительно увеличилась скорость поиска;
 - во-вторых, благодаря новой архитектуре системы обновление поискового индекса происходит несколько раз в день.



Rambler Mass Media

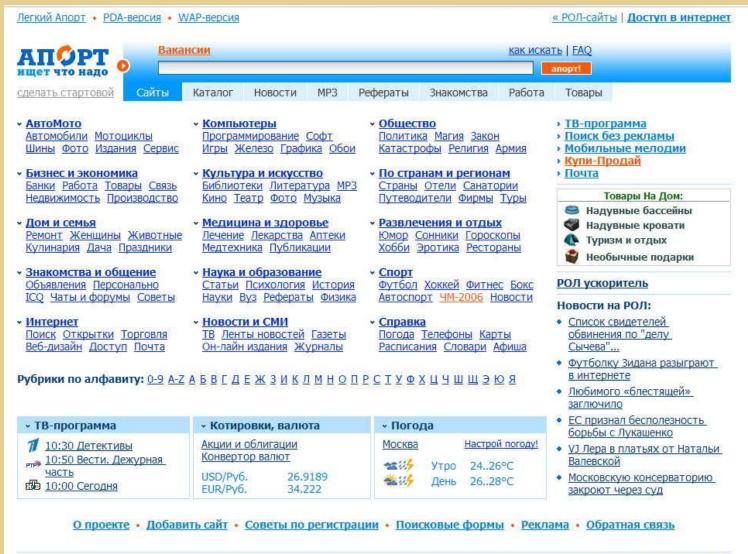
- Rambler Mass Media одна из крупнейших информационных площадок рунета.
- Проект круглосуточно взаимодействует с более чем 50 ведущими российскими и зарубежными СМИ.
- Использование большого количества информационных источников позволяет Rambler Mass Media предоставлять аудитории полную и достоверную информационную картину дня.



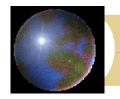
Anopm www.aport.ru

- Поисковая система Апорт появилась в интернете в 1996 году.
- В 2000 году была официально представлена новая версия «Апорт 2000».
- Поиск в ПС производится в таких сегментах интернете, как web-сайты, рефераты, товары, работа, знакомства, музыка, новости, энциклопедия «Кругосвет», а также в собственном каталоге.

Апорт www.aport.ru







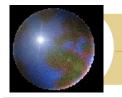
Anopm www.aport.ru

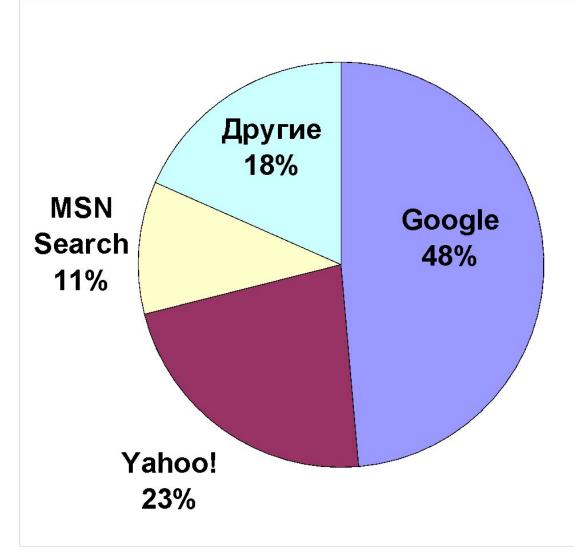
- Разработчики Апорта удачно использовали некоторые идеи, впервые реализованные в системе Google: в конечном итоге поиска одними из первых выдаются сайты, чей «ранг страницы» выше, т.е. чем больше сайтов ссылаются на данный сайт, тем он выше в списке.
- Система ранжирования (похожая на PageRank y Google) была реализована и в ПС Апорт.





Опрос зарубежных пользователей интернета





Наиболее популярные ПС в мире

(по данным Nielsen//NetRatings за февраль 2006 г.)

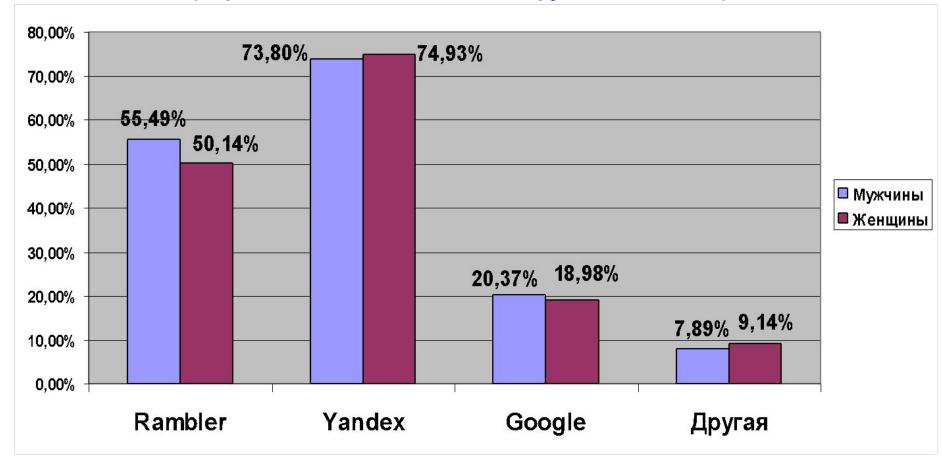


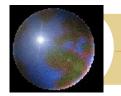


Опрос российских пользователей интернета

Опрос: какими поисковыми системами Вы пользуетесь?

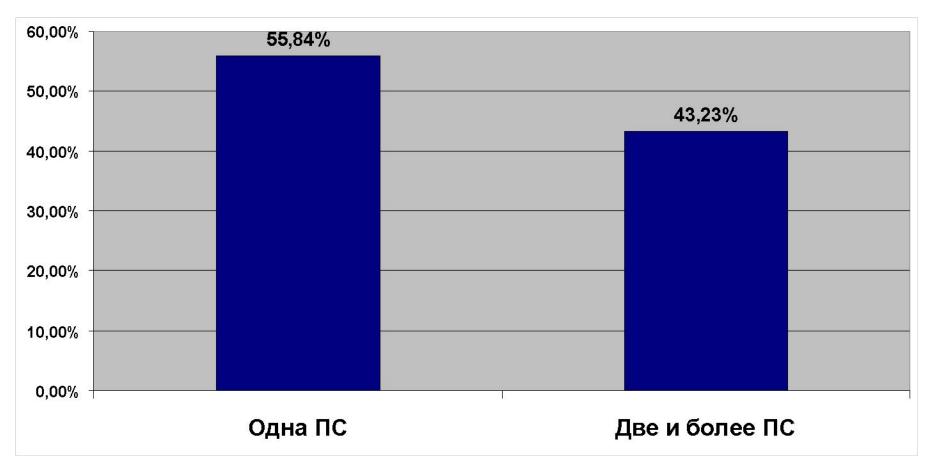
(март 2006 г., 1800 человек в 7 округах РФ, ЦМиСИ)

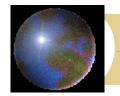




Каким количеством ПС Вы пользуетесь?

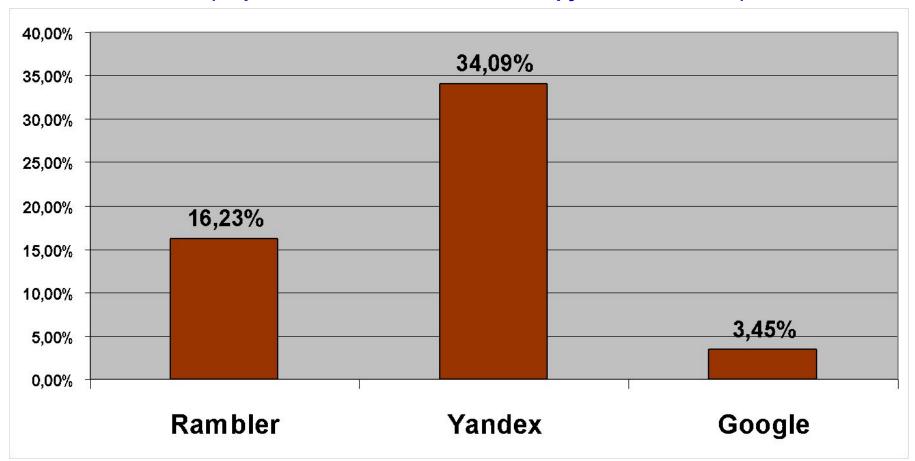
(март 2006 г., 1800 человек в 7 округах РФ, ЦМиСИ)





Среди тех, кто использует одну ПС: Какой именно ПС Вы пользуетесь?

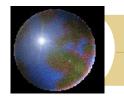
(март 2006 г., 1800 человек в 7 округах РФ, ЦМиСИ)





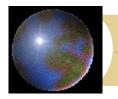


Задание по поиску в каталогах



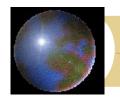
Задание по поиску в каталогах

- Вы планируете выпуск информационноразвлекательной газеты.
- Большей частью Ваше издание будет представлять собой интернет-дайджест.
- Вам необходимо найти источники (интернет-сайты) для наполнения тематических разделов газеты.



Тематические разделы газеты

- Новости страны и мира
- Курьезы и забавные случаи
- Необъяснимые явления и происшествия
- Новости из мира звезд
- Советы домохозяйкам
- Советы автовладельцам
- Гороскоп
- Кроссворды
- Анекдоты



Источники информации

- Дудихин В.В., Дудихина О.В. Конкурентная разведка в Интернет. — М.: ООО «Изд-во АСТ»: изд-во «НТ Пресс», 2004.
- Ландэ Д.В. Поиск знаний в INTERNET. Профессиональная работа.: Пер. с англ. М.: «Вильямс», 2005.
- Синебрюхов Л. Американский гражданин Сергей Михайлович Брин // Известия, №79. 05.05.2006 г.