

Воронежский государственный университет
Факультет компьютерных наук
Кафедра информационных систем

Становление ИПС Web: краткий исторический обзор

Краткая история ИПС Web

- **1990** – *Archie* (Alan Emtage – McGill University, Montreal):
 - программа-обработчик списков *ftp*-ресурсов, поиск по именам файлов с использованием рег. Выражений
- **1993** – *Veronica* (Nevada University)
 - серверы Gopher
- *WWW Wanderer* (Matthew Gray)
 - база данных URL
 - высокий трафик
- Октябрь **1993** – *ALIWEB* (Martijn Koster)
 - http-серверы
 - владельцы сайтов сами представляют информацию о сайте
 - малый объем базы данных

Краткая история ИПС Web

- Декабрь 1993:
 - *Jump Station* – индексирование заглавий и заголовков – линейный поиск в базе данных
 - *WWW Worm* – индексирование заглавий и URL, использование рег. выражений для поиска
 - *RSBE* (Repository-Based Software Engineering) – ранжирование по ключевым словам
- Февраль 1993 - *Excite*
 - использование статистического анализа связей между словами для эффективного поиска

Краткая история ИПС Web

- Январь 1994 - *Tradeware Galaxy*
 - группировка гиперссылок по иерархическим категориям
 - ручная группировка ссылок – высокое качество ссылок – незначительное количество найденных страниц
- Апрель 1994 – *Yahoo* (David Filo, Jerry Yang)
 - каталог с возможностями поиска
 - автоматизация процесса сбора и классификации

Краткая история ИПС Web

- 1994 – WebCrawler (Brian Pinkerton – Wash. Univ.)
 - индексирование полного текста страницы
- Июль 1994 - Lycos (Michael Mauldin – Carnegie Mellon Univ.)
 - ранжирование по релевантности
 - частичное соответствие
 - учет пространственной близости слов
- Январь 1994 – Infoseek
 - дружественный интерфейс
 - каталог

Краткая история ИПС Web

- Декабрь 1995 – *AltaVista* (DEC)
 - Первая ИПС промышленного масштаба (миллионы запросов в день)
 - Использование естественного языка в запросах
 - Поддержка логических операторов
 - Поиск по названиям изображений, Java апплетов и ActiveX объектов
 - Добавление и удаление пользователем своего URL в индексе
 - Поиск сайтов, ссылающихся на конкретный URL
 - Контекстная помощь

Краткая история ИПС Web

- 1995 – *MetaCrawler* (Eric Selburg – Wash. Univ.)
 - Выполнение запроса одновременно несколькими поисковыми системами
 - Сбор и переформатирование результатов на одной странице
 - Отображение рекламы

Краткая история ИПС Web

- Май 1996 - *Inktomi*
первая система, которая включала сайты в выдачу за плату.
- Апрель 1997 - *Ask Jeeves* (затем *Ask.com*)
попытка ранжировать сайты на основе их популярности, однако алгоритм имел слабую защиту перед спамом.
в 2000 появилась поисковая система *Teoma*, которая использовала кластеризацию для группировки сайтов по *Тематической Популярности* (*Subject Specific Popularity*).
На разработку данной системы большое влияние оказала статья Джона Кляйнберга “*Authoritative sources in a hyperlinked environment*”.

Краткая история ИПС Web

- 1998 - *Overture*

Впервые была реализована модель “*платы за клик*”.

Краткая история ИПС Web

- 1999 – Google (Sergey Brin, Larry Page – Stanford Univ.)
 - Алгоритм ранжирования PageRank
 - Оптимизация архитектуры системы
 - Масштабирование и распараллеливание (Linux-кластеры)
 - Восстановление текста документа без обращения к его оригиналу в Интернет

История ИПС Web в России

- Яндекс
- Rambler
- Апорт

Яндекс

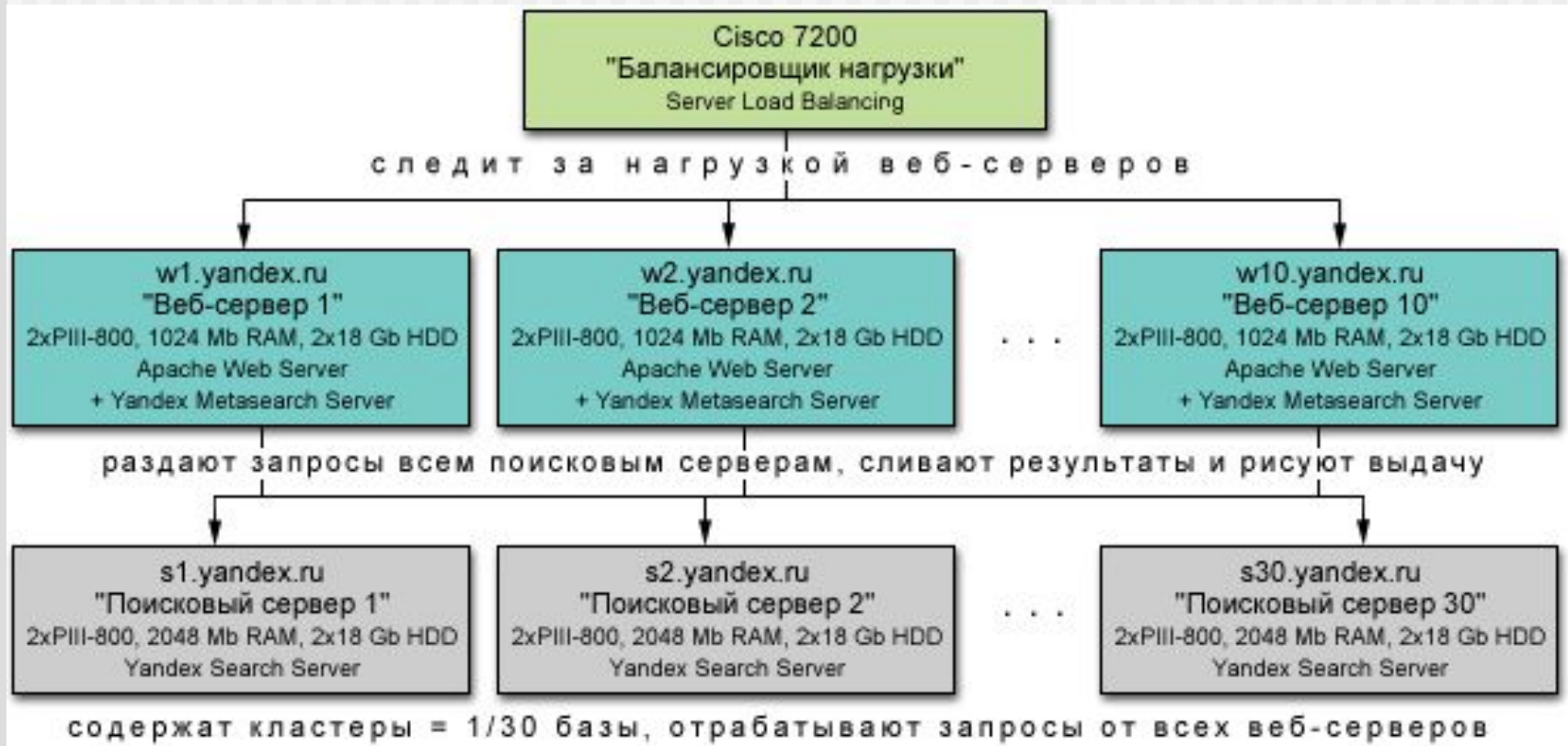
- **1990** – компания *Аркадия* (Аркадий Борковский и Аркадий Волож):
 - Начало разработки поискового программного обеспечения. За два года работ созданы *Международная классификация Изобретений* и *Классификатор Товаров и Услуг*
- **1993** – *Аркадия* становится подразделением *ComrTek*. Сотрудничество с лабораторией Ю.Д.Апресяна (ИППИ РАН). Создание словаря для поиска с учетом морфологии русского языка (300 КБ).
- **Начало 1996 г.** Разработан алгоритм построения гипотез (снятие привязки к словарю) для незнакомых слов.
- **Лето 1996.** Руководство *ComrTek* и разработчики пришли к выводу о самоценности развития самой технологии. Появление “*Yandex*” в Интернете.
- “*Yandex*” = “языковой индекс”, “Yet Another Index”. “Y” = “Я”.

Яндекс

- **Осень 1997 г.:**
 - открытие *Yandex.ru*
 - реализован естественно-языковой интерфейс
- **Май 2001 г. Новая версия ИПС “Яндекс”.**

Яндекс

Как устроен Яндекс



Яндекс

Яндекс в железе



Rambler

- **1991** – команда единомышленников.
- **1996** – запуск в эксплуатацию поискового механизма Rambler.ru
- **Февраль 1997** – рейтинговая система Rambler's Top100.

Апорт

- **1996** – демонстрация поисковой машины “Апорт”.
- **Октябрь 1999** – “Апорт 2000”, интегрированная с каталогом “Каталог-Апорт”.
- Первая российская поисковая машина, реализовавшая учет “ранга страницы” и обработку запроса с ориентацией на HTML-код страницы.

Источники

- <http://www.search-marketing.info/search-engine-history/>
- <http://company.yandex.ru/history/pred.xml>
- http://company.yandex.ru/programs/web_200203.html
- <http://www.rambler.ru/doc/about.shtml>