

Тенденции развития поисковых систем

Александр Садовский

Яндекс: что нового в 2005

Робот

- быстрый (8-10 раз в сутки, 20 млн. URL)
- кластеризация основного (скорость в 1,5 раза больше)

Поиск

- ранжирование коротких запросов
- колдовство по опечаткам
- yandex.ua (морфология, регион сайта)

Антиспам

- внедрен автоматический отлов



Яндекс: что нового в 2005



Яндекс: что нового в 2005



Поиск революции в поиске

Принципы Web 2.0 и веб-поиск

- пользователи «улучшают» сервис:
 исправление опечаток на основе запросов;
 ИЦ; ссылки из блогов
- ценность уникальных данных: архивы блогов, словари, архивы запросов
- «длинный хвост»: поиск, а не только навигация; реклама по ключевым словам
- открытые API: Яндекс.XML, RSS
 поисковых результатов в Новостях, Блогах
- есть развитие, но нет версий: обратная связь от веб-мастеров и пользователей

Поиск революции в поиске

Борьба с мусором

- дубликаты и клоны: шинглы, лексические сигнатуры, сниппеты
- технический мусор: высокая изменчивость текстов и URL

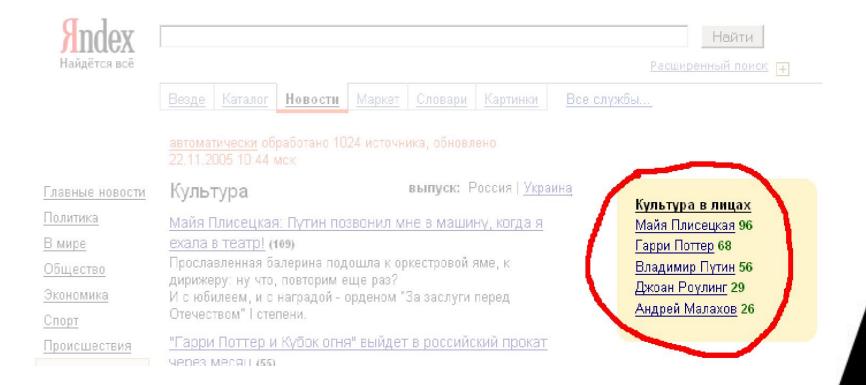
Антиспам

- статистический анализ текста
- анализ распределения ссылок
- обнаружение колец
- TrustRank

Поиск революции в поиске

Извлечение фактов: первый шаг

– «Новости в лицах»





Яндекс: поиск революции в поиске

Стимуляция исследований: научные стипендии

- 252 заявки, 34 стипендии
- сборник «Интернетматематика-2005»

Tadex Recorded to the second second

Найдется все.