



Подготовка данных для поиска Картинок и Видео

Дмитрий Беляев

Руководитель группы

мультимедийных поисковых сервисов

Я расскажу о том:

- Зачем нужен мультимедийный импорт
кому он пригодится и какая от него польза
- Какие требования предъявляются к сайтам и данным
кого мы берем, а кого – нет
- Как это работает и с чего начать
XML-импорт и Media RSS
- На что обратить внимание

Я.Картинки и Я.Видео

Яндекс.Картинки (images.yandex.ru):

- 1,1 млрд. изображений
на сайтах в Рунете и за его пределами
- 1,5 млн. посетителей, 8 млн. запросов в сутки

The screenshot shows the Yandex Images search interface. At the top, there is a search bar with the text 'озеро Байкал' and a search button. Below the search bar, there are navigation links for 'Любые', 'Обои', 'Большие', 'Средние', 'Маленькие', and 'Портреты'. The search results are displayed in a grid format, showing various images of Lake Baikal. Each image has a caption and a link to the source website. The search results include:

- Изображение глобуса с надписью 'Найдите город вашей мечты'.
- Изображение озера Байкал с надписью 'Озеро Байкал находится на юге Восточной Сибири. 585x453 ... 600x465 sohes.net'.
- Изображение озера Байкал с надписью 'куда вложить деньги, озеро байкал, инвестиции недвижимости... 250x154 ... 649x400 www.baikal.nbbr.ru'.
- Изображение озера Байкал с надписью 'Озеро Байкал 450x289 ... 622x400 copypast.ru'.
- Изображение озера Байкал с надписью 'Озеро Байкал. 400x616 www.magicbaikal.ru'.
- Изображение озера Байкал с надписью 'куда вложить деньги, озеро байкал, инвестиции недвижимости... 250x157 ... 635x400 baikal.nbbr.ru'.
- Изображение озера Байкал с надписью 'Озеро Байкал находится на юге Восточной Сибири. 321x200 ... 642x400 my-fishing.org.ru'.
- Изображение озера Байкал с надписью 'Западное побережье озера Байкал. 604x400 kiprida14narod.4bb.ru'.
- Изображение озера Байкал с надписью 'Озеро Байкал 450x239 ... 602x400 copypast.ru'.
- Изображение озера Байкал с надписью 'озеро Байкал покрытое льдом 640x480 ... 1024x768 cheboksar.net'.
- Изображение озера Байкал с надписью 'Озеро Байкал (Lake Baikal) 540x405 ... 800x600 www.travelling.ru'.
- Изображение озера Байкал с надписью 'Озеро Байкал. 140x93 ... 593x400 www.magicbaikal.ru'.

Я.Картинки и Я.Видео

Яндекс.Видео (video.yandex.ru):

- 13 млн. видео-роликов
на видео-хостингах, новостных и других тематических сайтах
- 450 тыс. посетителей, 1,4 млн. запросов в сутки

The screenshot shows the Yandex.Video search results page for the query "Евровидение 2009". The search bar at the top contains the text "Евровидение 2009" and a "Найти" button. Below the search bar, it indicates "Найдено видео: 4 337" and a filter for "умеренный". The main content area displays a grid of video thumbnails with their titles and source URLs. The thumbnails include:

- A.Рыбак - Сказка (Евровидение 2009) from www.liveinternet.ru - 21.07.2009
- С.Лобода Украина (Евровидение 2009) from www.liveinternet.ru - 21.07.2009
- Patricia Kaas - Евровидение 2009 from www.liveinternet.ru - 12.06.2009
- ... объявления результатов полуфинала Евровидения... from www.liveinternet.ru - 26.07.2009
- ЕВРОВИДЕНИЕ 2009= ИЗРАИЛЬ from www.liveinternet.ru - 7.08.2009
- ЕВРОВИДЕНИЕ 2009= АНАСТАСИЯ ПРИХОДЬКО from www.liveinternet.ru - 7.08.2009
- ЕВРОВИДЕНИЕ 2009 ...туре "Евровидения" стала для 4,5-миллионной... from www.liveinternet.ru - 7.08.2009
- Евровидение-2009. Открытие from www.1tv.ru - 10.05.2009

The left sidebar contains navigation options such as "Источники", "Ролики", and "Сортировать".

Мультимедийный импорт

Кому это пригодится:

- Фото- и видео-хостингам
- Сайтам с динамически обновляющимся мультимедийным контентом (СМИ, фото- и видео-архивам и т.п.)

Для чего это нужно:

- Скорость индексирования
- Качество индексирования
- Полнота и актуальность данных
- Рост трафика

Требования к сайтам и данным

Берем:

- Сайты с уникальным и оригинальным контентом
- Фотографии и видео-ролики:
 - имеющие внятные и качественные описания
 - размещенные в свободном и бесплатном доступе
 - прошедшие модерацию на сайте-источнике

Не берем:

- Материалы оскорбительной или порно-тематики
- Сайты:
 - нарушающие требования законов РФ
 - не отвечающие требованиям «Лицензии...»

<http://rules.yandex.ru/termsfuse.xml>

Как это работает

XML-импорт:

- написать письмо с описанием вашего ресурса:
 - Картинки: xml-import@images.yandex.ru
 - Видео: xml@video.yandex.ru
- сделать скрипт, готовящий данные в XML-формате:
 - Картинки: <http://images.yandex.ru/xml-import.html>
 - Видео: <http://video.yandex.ru/partners.xml>

Media RSS:

- <http://search.yahoo.com/mrss>
- далее – аналогично XML-импорту

На что обратить внимание

- **Корректность структуры и данных**
соответствие XML- и RSS-стандартам, отслеживание поломок
формат дат: год-месяц-день[Тчас:минута:секунда[Z | +час[:]минута]]
возвращать стандартные коды (500, 501, 503 и т.п.)
- **Полнота данных**
длительность, рейтинги и т.п.
- **Оперативность обновления данных**
исправления, удаления, паролирование, постмодерация и т.п.
- **Недопустимость спама и накруток**
в описаниях, тэгах и рейтингах



Дмитрий Беляев

Руководитель группы мультимедийных
поисковых сервисов

119021, Россия, Москва,
ул. Льва Толстого, д.16

+7 (495) 739-00-00

+7 (495) 739-70-70 — факс

dvbelyaev@yandex-team.ru