

Яндекс

Яндекс.Новости: как и зачем

Татьяна Исаева

Задачи Яндекс.Новостей

Миссия Яндекса – отвечать на заданные и незаданные вопросы пользователей

Яндекс.Новости:

- **На главной странице** Яндекса удовлетворяют потребность пользователей в актуальной информации

Незаданный вопрос: *«Что сейчас происходит?»*

- **В Поиске** отвечают на явно сформулированные вопросы пользователей

Заданные вопросы: *«Кастер Семеня», «отставка Воронина», «11 сентября башни близнецы»*

Принципы Яндекс.Новостей

- Полностью автоматический сервис
- Партнерский сервис
- Единые требования для всех партнеров
- Зеркало русскоязычной медиа-среды

Цели:

отразить новостную картину дня и дать ссылки на источники, содержащие наиболее полную информацию о событиях

Модель сервиса

Яндекс
НОВОСТИ

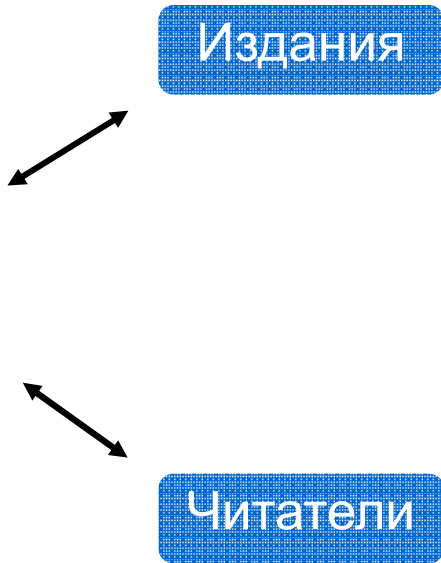
Издания

- Обеспечивают быструю трансляцию контента (30 тыс. сообщений в сутки)
- Получают читателей (1 млн переходов в сутки)

Читатели

- Узнают новости на Яндексе (1-5 млн человек в сутки)
- Читают подробности у источников (500 тыс. человек в сутки)
- Задают вопросы о новостях (150 тыс. запросов в сутки)

Интересы участников



- Все хотят трафика
- Первоисточники хотят приоритетного размещения и подавления рерайтеров
- Оптимизаторы хотят прозрачных правил

- Хотят быстро узнавать, не случилось ли чего
- Если вдруг случилось, получить ссылку на статью от знакомого источника
- Получить ссылку на первоисточник

Интересы Яндекс.Новостей

- Расширение аудитории, повышение ее лояльности
 - качественный сервис
- Привлечение новых поставщиков интересной информации и сохранение существующих
 - выгодная для партнеров модель сотрудничества

Основные продукты Яндекс.Новостей

1. Страница сюжета
2. Новостной блок на главной странице Яндекса
3. Главная страница Новостей, страницы рубрик
4. Поиск по новостям (новостные результаты в поиске Яндекса)

Релевантный новостной ответ

Алгоритм, формирующий ответ на новостные запросы пользователей, ориентируется на следующие свойства сообщений:

- принадлежность первоисточнику
- оперативность
- цитируемость
- информативность

Источники, у которых доминируют такие сообщения, предлагаются пользователям в первую очередь.

Новостной контент

- **Соответствие формату** (не подходят для трансляции реклама, пресс-релизы компаний, блоги, тв-программа, гороскопы...)
- **Технические требования** (доступность сайта, отдельные адреса для каждого сообщения...)
- **Требования законодательства** (защита прав на интеллектуальную собственность; противодействие терроризму; защита чести, достоинства, репутации...)

Для чего нам нужны полные тексты сообщений

Анализ полных текстов сообщений позволяет алгоритму:

- полнее и точнее сформировать «сюжет» - множество новостных сообщений, посвященных одному событию.
- отделять текст сообщения от «обвязки»
- находить документы в базе по любому фрагменту их текста

**Нигде на сервисе полный текст сообщений
партнеров не показывается**

Я

Технология экспорта

- Ссылку на экспортный файл, размещённый на сервере партнёра, нужно прислать на info@news.yandex.ru
- Робот автоматически считывает новости со всех экспортных файлов наших партнёров раз в 5-10 минут
- Время от включения сообщения в экспортный файл до публикации ссылки на него в Яндекс.Новостях может составлять до 40 минут

Сообщение в экспортном файле

```
<item>
<title>Число жертв теракта в Киркуке достигло 100 человек</title>
  <link>http://www.vesti.ru/doc.html?id=232133</link>
  <pdalink>wap.vesti.ru/doc.wml?id=232133</pdalink>
  <description>До 100 человек возросло число жертв теракта,
совершенного в иракском городе Киркук. Смертник привел в действие
закрепленную на теле взрывчатку в курдском ресторане.</description>
  <pubDate>Thu, 11 Dec 2008 15:57:00 +0300</pubDate>
  <category>В мире</category>
  <enclosure url="http://www.vesti.ru/p/b_293790.jpg"
type="image/jpeg" />
  <enclosure url="http://www.vesti.ru/videos?vid=169629"
type="video/x-flv" />
  <yandex:full-text>До 100 человек возросло число жертв теракта,
совершенного в четверг на севере Ирака. Об этом сообщает телеканал &quot;Аль-
Арабия&quot;.
Смертник привел в действие закрепленную на теле взрывчатку в курдском
ресторане, расположенном в нескольких километрах от города Киркук.
Во время теракта заведение было переполнено - иракцы пришли семьями в
последний день праздника жертвоприношения Ид аль-Адха. После взрыва место
инцидента было оцеплено иракскими и американскими военнослужащими, передает
ИТАР-ТАСС.
Этот теракт стал крупнейшим за последние год-полтора, когда насилие в стране
резко пошло на спад после серии антитеррористических операций в Багдаде и
мятежных провинциях страны, где были сильны позиции &quot;Аль-
Каиды&quot;.</yandex:full-text>
</item>
```

Что мы делаем с тегами -1

<**title**> - заголовок сообщения. Отображается в Яндекс.Новостях, участвует в кластеризации

<**link**> - URL сообщения, считается уникальным идентификатором сообщения. Кликая на заголовок в Яндекс.Новостях, пользователь идёт по этой ссылке

<**author**> - имя и фамилия автора сообщения. Даёт возможность найти все материалы автора

<**category**> - рубрика. Помогает определить тематическую принадлежность сюжета, в который попадает сообщение

Что мы делаем с тегами -2

<enclosure> - элемент для иллюстраций, аудио и видеофайлов. Содержит ссылку на страницу с мультимедиа

<pubDate> - время публикации сообщения. Отображается в Яндекс.Новостях рядом с заголовком сообщения

<yandex:full-text> - полный текст сообщения. Не отображается полностью в Яндекс.Новостях, но участвует в кластеризации и служит материалом для создания аннотаций

Что мы делаем с разметкой

Теги форматирования позволяют роботу не принимать подзаголовков за часть предложения

Параметр **type** позволяет роботу отличить фото, видео и аудио друг от друга

Корректное включение ссылок на цитируемые материалы помогает правильно определять первоисточник

Часто встречающиеся ошибки

кодировка – не совпадает с фактической

xml-формат – некоторые символы необходимо заменять

время публикации – неверное смещение часового пояса, сообщения датированы будущим временем

<yandex:full-text> – включение лишней информации, уже содержащейся в других тегах; помещение только части текста сообщения

заголовки – некорректное повышение регистра, служебные вставки («фото»; «срочно»; «эксклюзив»)

доступность экспортного файла – не скачивается полностью за 10-15 секунд

**Хотите сотрудничать с сервисом
Яндекс.Новости?**



**Пришлите нам заявку по адресу
info@news.yandex.ru**

Татьяна Исаева, руководитель службы по
работе с партнерами Яндекс.Новостей

1. Сюжет в Яндекс.Новостях

Фото (всего 19)



ПитерБурлер.ru



OpenSpace.ru

Аудио (всего 1)

Эхо Москвы

Культура // Санкт-Петербург

Эрмитаж потребовал от Мадонны отказаться в Петербурге от богохульства

Один из фирменных номеров певицы - это появление распятой на большом зеркальном кресте во время исполнения баллады Live To Tell. [JustMedia](#) 09:59

Русская православная церковь назвала богохульством и оскорблением христианства этот номер во время выступления Мадонны в Москве два года назад. [Новые известия](#) 06:00

ВОСТОК-МЕДИА - В выступлении певицы Мадонны, которое запланировано на Дворцовой площади Санкт-Петербурга 2 августа, не должно быть богохульства, считает Михаил Пиотровский - директор государственного Эрмитажа в Петербурге, расположенного рядом с местом проведения концерта, передает РИА "Новости". [Восток-Медиа](#) 05:05

Директор Государственного Эрмитажа Михаил Пиотровский заявил, что в выступлении Мадонны, которое пройдет 2 августа на Дворцовой площади Санкт-Петербурга, не должно быть богохульства, пишет РИА Новости.

Михаил Пиотровский, Эрмитаж, директор [все цитаты сюжета](#) (17)

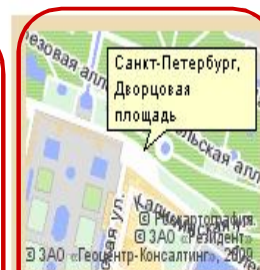
В сюжете: [Михаил Пиотровский](#) • Эрмитаж

Все новости сюжета (всего 33) Статьи (всего 1)

[Эрмитаж разрешил Мадонне петь, но не богохульствовать](#) УРА-информ 11:43

[Мадонну попросили не порочить Дворцовую площадь](#) Агентство национальных новостей 11:09

[Эрмитаж потребовал от Мадонны отказаться в Петербурге от богохульства](#) РИА Новости 14:31



На выступлении певицы Мадонны, которое должно состояться на Дворцовой площади Санкт-Петербурга 2 августа, не место богохульству, считает директор расположенного рядом с местом проведения концерта Государственного Эрмитажа Михаил Пиотровский. [Новые известия](#) 06:00

[25 мнений](#) в блогах

Схема создания сюжета

Создание поискового запроса из ключевых слов документа, построение матрицы близости документов



Аннотирование сюжета: выбор заголовка и фрагментов сообщений, содержащих основные факты сюжета



Выбор основных заголовков сюжета



Выбор статей и интервью

Определение дубликатов

Эксклюзивный – материал, созданный силами редакции источника

Первоисточник определяется по времени публикации сообщения на сайте, по сравнительному анализу текстов, по цитированию агентств (учитываются гиперссылки, текстовые упоминания первоисточника)

Дубликаты отображаются в сюжетах как «похожие сообщения» и не участвуют в поиске по Яндекс.Новостям.

Выбор заголовка сюжета

- Соответствие региону пользователя
- Лексическая и фактологическая «ядерность»
- Актуальность фактов
- Информативность и читаемость

Заголовки сюжетов меняются в каждой итерации.

Цель:

первый по времени заголовок, наиболее полно отражающий актуальную фактическую сторону сюжета, не содержащий нехарактерных для сюжета слов и фактов.

Создание аннотации сюжета (дайджеста)

Из всех сообщений сюжета автоматически выделяются наиболее значимые объекты, имена людей, названия организаций, географические объекты, даты и числа. Они, наряду с ключевыми словами сюжета и новостными запросами, определяют выбор текстов для аннотации.

Предложения, составляющие аннотацию меняются в каждой итерации.

Цель:

показать предложения из сообщений,
содержащие основные факты события

Выбор основных заголовков сюжета

Цитирование
источника
в сюжете

Дата публикации

Вес источника

Цель:

Должны быть представлены первоисточник(и), наиболее цитируемые источники и заголовки, отражающие актуальное состояние сюжета

В сюжете: [Михаил Пиотровский](#) • [Эрмитаж](#)

Все новости сюжета ([всего 33](#)) Статьи ([всего 1](#))

[Эрмитаж разрешил Мадонне петь, но не богохульствовать](#) УРА-информ 11:09

[Мадонну попросили не порочить Дворцовую площадь](#) Агентство национальных новостей 11:43

[Эрмитаж потребовал от Мадонны отказаться в Петербурге от богохульства](#) РИА Новости 14:31

Определение рубрики и географии сюжета

- Тематическая рубрика
 - Специализация изданий
 - Рубрикация от источника
 - Лексические запросы
 - География
 - Выделение обозначений географических объектов из сообщений сюжета
 - Автоматическое определение релевантных для сюжета географических объектов
 - Карта населенного пункта, адреса
- Цель: получить все новости рубрики и региона

2. Выбор главных новостных сюжетов

Сегодня в новостях 15:06 **все** Москва

1. СКП выяснил причину смерти китайки в поезде Благовещенск-Москва
2. В Чечне сегодня отменен режим контртеррористической операции
3. Решение об отставке Лаудрупа принял владелец «Спартака»
- новость часа** 4. СКП РФ предъявил в окончательной редакции обвинение Сторчаку
5. Патриарх Кирилл омыл ноги двенадцати священнослужителям



Почта

 запомнить меня

[Забыли пароль?](#)

Поиск [Карты](#) [Маркет](#) [Новости](#) [Словари](#) [Блоги](#) [Видео](#) [Картинки](#) [ещё](#) ▼

Например, улица Дерибасовская



[Каталог сайтов](#)
[Развлечения](#)

[Маркет](#)
[купить телевизор](#)

Цель:
отобрать самые освещаемые в СМИ,
общезначимые, актуальные и вызывающие
интерес пользователей события.

Определение веса сюжета

Вес отдельного сообщения

- Время создания
- Вес источника
- Соответствие тематики сюжета специализации агентства

Динамика сюжета

- Плотность потока сообщений
- Динамика кликабельности новости на главной странице
- Время жизни в топе

Интерес пользователей и медийность

- Количество новостных запросов к Яндексу
- Записи в блогах
- Место новости на сайтах партнёров

Вес источника

Цитируемость учитывает, насколько часто ссылаются на источник другие новостные ресурсы

Оперативность учитывает, насколько часто источник быстро реагирует на события

Вес источников пересчитывается каждые два месяца

Робот и человек в Яндекс.Новостях

Человек

1. Принимает решение о сотрудничестве
2. Удаляет ссылку на сообщение из базы данных Яндекс.Новостей в случае, если текст сообщения изменён или отсутствует на сайте источника

Робот

1. Скачивает и индексирует новостные сообщения
2. Определяет и исключает из поиска по Яндекс.Новостям дубли
3. Определяет жанр сообщения
4. Выделяет ключевые слова и факты
5. Объединяет сообщения в сюжет
6. Определяет рубрику
7. Ранжирует сюжеты