

# Яндекс

## Яндекс.Новости как партнерский сервис

Лев Гершензон, Татьяна Исаева

# Модель Яндекс.Новостей

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

Издания



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



Издания

Яндекс

НОВОСТИ



Читатели

- 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

# Релевантный новостной ответ

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

- Первоисточник
- Оперативность
- Цитируемость
- Информативность

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

# Принципы Яндекс.Новостей

- Полностью автоматический сервис
- Единые требования для всех партнеров
- Быть зеркалом русскоязычной медиа-среды, а не одним из ее субъектов

Цели:

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

# 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...)

# Стать партнером

можно только по обоюдному согласию, подтвержденному документально

- Прочитать <http://news.yandex.ru/cooperation.html>
- Подать заявку на [info@news.yandex.ru](mailto:info@news.yandex.ru)
- Выполнить технические требования
- Подписать соглашение с Яндексом



# Что всех волнует

1. Формирование и представление сюжета
2. Ранжирование сюжетов

# Определение дубликатов

Перед формированием сюжета автоматически определяются группы очень похожих документов («дубликатов»).

Из каждой группы остается по одному сообщению («мастеру»), которое может быть представлено на страницах сюжета и в результатах поиска.

«Мастер» определяется по времени публикации сообщения на сайте, по сравнительному анализу текстов, по цитированию источника (учитываются гиперссылки, текстовые упоминания).

«Дубликаты» не представлены в сюжетах и не участвуют в поиске по Яндекс.Новостям.

# 1. Сюжет в Яндекс.Новостях

Фото (всего 145)



[ПитерБургер.ru](#)



[Кремль.Орг](#)

Видео (всего 9)



[Комсомольская правда](#)  
[Первый канал](#)  
[Звезда](#)

Аудио (всего 19)

[Радио Культура](#)  
[Радио Маяк](#)  
[Радио России](#)

Общество

## Первая запись появилась в блоге Дмитрия Медведева в "Живом журнале"

Блог самого высокопоставленного пользователя Рунета, президента России [Дмитрия Медведева](#) - [blog\\_medvedev](#) - появился в среду в сообществе Live Journal, "Живом Журнале", или просто "ЖЖ", выражаясь интернет-новоязом. [Самотлор-экспресс](#) 14:54

Президент России выступил против попыток одностороннего регулирования сети какой-либо из стран, считает необходимым выработать общие международные нормы - Пытаться регулировать сеть, делать это примитивно - очень сложно. [Наш век](#) 12:37

"Ситуация с открытостью, с готовностью государства к различным государственным сервисам у нас, мягко говоря, пока далека от идеальной, а если говорить откровенно, то почти нет этого", - заявил глава государства в новой записи в своем видеоблоге. [РИА Новости](#) 11:34

Еще на эту тему

[Медведев заводит ЖЖ, но еще не определился с никнеймом](#)  
21.04.2009

[314 мнений](#) в блогах

Президент РФ Дмитрий Медведев считает, что главными принципами функционирования интернета должны стать его доступность, удобство для пользователей, а также отражение в нем деятельности органов власти, сообщает "Интерфакс".

Дмитрий Медведев, Россия, президент [все цитаты сюжета](#) (83)

**В сюжете:** [Дмитрий Медведев](#), [Наталья Тимакова](#), [Эндрю Полсон](#), [Александр Мамут](#) • [Живой журнал](#)

**Все новости сюжета** ([всего 402](#)) **Статьи** ([всего 30](#)) **Интервью** ([всего 2](#))

[Президент страны Дмитрий Медведев всего лишь за несколько часов после официального открытия ...](#) [Эхо Москвы](#) 18:29

[Официально открыт блог Дмитрия Медведева в "LiveJournal"](#) [Первый канал](#) 18:07

[Человек дня: Дмитрий Медведев](#) [Актуальные комментарии](#) 17:46

[За первый час в blog\\_medvedev написали 200 человек](#) [Вести.Ru](#) 13:41

[Медведев: Кризис даст старт новым интернет-проектам](#) [BFM.ru](#) 12:18

[Первое видео в ЖЖ Дмитрий Медведев посвятил Интернету](#) [Радио Маяк](#) 12:07

[Президент РФ заявил, что государство должно обеспечить доступность и информационное удобство ...](#) [Интерфакс](#) 11:41

[Президент Медведев посоветовал чиновникам создавать ЖЖ-дневники](#) [KM-News](#) 11:32

[Первая запись появилась в блоге Дмитрия Медведева в "Живом журнале"](#) [Gzt.Ru](#) 11:21

# Выбор заголовка сюжета

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

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.

Цель:

показать предложения из сообщений, содержащие основные факты события

# Аннотация сюжета.

## Информативность

### Первая запись появилась в блоге Дмитрия Медведева в "Живом журнале"

Блог самого высокопоставленного пользователя Рунета, президента России Дмитрия Медведева [blog\\_medvedev](#) - появился в среду в сообществе Live Journal, "Живом Журнале", или просто "ЖЖ", выражаясь интернет-новоязом. [Самотлор-экспресс](#) 14:54

Президент России выступил против попыток одностороннего регулирования сети какой-либо из стран, считает необходимым выработать общие международные нормы - Пытаться регулировать сеть, делать это примитивно - очень сложно. [Наш век](#) 12:37

"Ситуация с открытостью, с готовностью государства к различным государственным сервисам у нас, мягко говоря, пока далека от идеальной, а если говорить откровенно, то почти нет этого", - заявил глава государства в новой записи в своем видеоблоге. [РИА Новости](#) 11:34


Президент РФ Дмитрий Медведев считает, что главными принципами функционирования интернета должны стать его доступность, удобство для пользователей, а также отражение в нем деятельности органов власти, сообщает "Интерфакс".

Дмитрий Медведев, Россия, президент [все цитаты сюжета](#) (83)

#### Еще на эту тему

[Медведев заводит ЖЖ, но еще не определился с никнеймом](#)

21.04.2009

 [314 мнений](#) в блогах

**В сюжете:** [Дмитрий Медведев](#), [Наталья Тимакова](#), [Эндрю Полсон](#), [Александр Мамут](#) • [Живой журнал](#)

**Все новости сюжета** ([всего 402](#))   **Статьи** ([всего 30](#))   **Интервью** ([всего 2](#))

[Президент страны Дмитрий Медведев всего лишь за несколько часов после официального открытия ...](#) [Эхо Москвы](#) 18:22

[Официально открыт блог Дмитрия Медведева в "LiveJournal"](#) [Первый канал](#) 18:07

[Человек дня: Дмитрий Медведев](#) [Актуальные комментарии](#) 17:46

[За первый час в blog\\_medvedev написали 200 человек](#) [Вести.Ru](#) 13:41

[Медведев: Кризис даст старт новым интернет-проектам](#) [BFM.ru](#) 12:18

[Первое видео в ЖЖ Дмитрий Медведев посвятил Интернету](#) [Радио Маяк](#) 12:07

[Президент РФ заявил, что государство должно обеспечить доступность и информационное удобство ...](#) [Интерфакс](#) 11:41

[Президент Медведев посоветовал чиновникам создавать ЖЖ-дневники](#) [KM-News](#) 11:32

[Первая запись появилась в блоге Дмитрия Медведева в "Живом журнале"](#) [Gzt.Ru](#) 11:21

# Заголовок сюжета. «Ядерность» лексики



первая запись

только в этом сюжете

[расшир](#)

[Главные новости](#)

[Политика](#)

[В мире](#)

[Общество](#)

[Экономика](#)

[Спорт](#)

[Происшествия](#)

[Культура](#)

[Наука](#)

## Первая запись появилась в блоге Дмитрия Медведева в "Живом журнале"

Результат поиска — сообщений : 42

1. [Медвед, превед!](#) 19:00 iToday.ru  
**Первая запись** появилась еще раньше - 7 октября 2008 года.
2. [Дмитрий Медведев стал блоггером-тысячником в ЖЖ](#) 17:07 CNews  
**Первую запись** Дмитрия Медведева прокомментировали более пятисот человек.
3. [Первую запись](#) Дмитрия Медведева в живом журнале прокомментировало более 1000 человек 17:03 Перспектива  
**Первую запись** Дмитрия Медведева в живом журнале "О развитии Интернета в России" прокомментировало уже более 1000 человек.
4. [Медведев опубликовал первый пост в ЖЖ](#) 16:23 Голос Самары.ru  
**Первую запись** президента России Дмитрия Медведева в ЖЖ, посвященную перспективам развития интернета в России, откомментировали около тысячи человек.
5. [Блог президента. Первая запись](#) Дмитрия Медведева 16:11 Звезда [посмотреть](#)  
**Первая запись** появилась в блоге Дмитрия Медведева в "Живом журнале".
6. [Первая запись](#) в ЖЖ Медведева. Блоггеры: "Писать можно всем и ничего за это не будет?", "Ты живой? :)" 15:52 РИА Ура.Ru

# 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: [Дмитрий Медведев](#), [Наталья Тимакова](#), [Эндрю Полсон](#), [Александр Мамут](#) • Живой журнал

**Все новости сюжета** ([всего 402](#))   **Статьи** ([всего 30](#))   **Интервью** ([всего 2](#))

[Президент страны Дмитрий Медведев всего лишь за несколько часов после официального открытия ...](#) Эхо Москвы 18:22

[Официально открыт блог Дмитрия Медведева в "LiveJournal"](#) Первый канал 18:07

[Человек дня: Дмитрий Медведев](#) Актуальные комментарии 17:46

[За первый час в blog\\_medvedev написали 200 человек](#) Вести.Ru 13:41

[Медведев: Кризис даст старт новым интернет-проектам](#) BFM.ru 12:18



# Выбор заголовков для первой страницы. Цитируемость в сюжете

Яндекс

НОВОСТИ

интерфакс

только в этом сюжете

расу

[Главные новости](#)

[Политика](#)

[В мире](#)

[Общество](#)

[Экономика](#)

[Спорт](#)

[Происшествия](#)

[Культура](#)

[Наука](#)

Первая запись появилась в блоге Дмитрия Медведева в "Живом журнале"

Результат поиска — сообщений : 24

1. [ЖЖ номер один](#) 14:53 Interfax-russia.ru

Ответ на то, для чего главе государства сразу два блога, дал **"Интерфаксу"** во вторник сотрудник интернет-департамента администрации Кремля.

2. [В видеоблоге Дмитрия Медведева появилась первая запись](#) 14:36 Финам FM

Как отмечает **"Интерфакс"**, отклики посетителей ЖЖ лаконичны, при этом отсутствует жесткость в модерации записей, которые изобилуют интернет-сленгом.

3. [Медведев впервые выступил в "ЖЖ"](#) 13:51 Осинформ

Я думаю, что главной задачей должно стать создание хороших условий для максимальной доступности интернет-услуг", передает **"Интерфакс"**. По мнению президента, это очень непростая задача с учетом масштабов государства.

4. [Медведев высказался об интернете](#) 13:39 JustMedia

Президент РФ Дмитрий Медведев считает, что главными принципами функционирования интернета должны стать его доступность, удобство для пользователей, а также отражение в нем деятельности органов власти, сообщает **"Интерфакс"**.

5. [Зафрендить Медведева уже можно](#) 13:20 Edinros.ru

По материалам **"Интерфакса"** и "Первоисточника"

# 2. Выбор главных новостных сюжетов



**Почта**

логин

пароль

запомнить меня

Войти

[Забыли пароль?](#)

[Завести почтовый ящик](#)

**Фотки**

- Сегодня в новостях** 23:13 все Москва
1. Первая запись появилась в блоге Дмитрия Медведева в [«Живом журнале»](#)
  2. «Убийство» ректора Полярной академии [организовали ее подчиненные](#)
  3. Арсен Венгер: Не думал, что [Аршавин способен на такое чудо](#)
  4. Результат матча «Локомотив» — «Зенит» [оставлен без изменений](#)
  5. Финансовый директор Freddie Mac [покончил жизнь самоубийством](#)

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

Например, [Периодический закон](#)



- Каталог сайтов**
- [Развлечения](#)
  - [Спорт и отдых](#)
  - [Работа и учеба](#)
  - [Компьютеры](#)
  - [Бизнес](#)

- Маркет**  
[купить телевизор](#)
- Авто**  
поиск по объявлениям
- Мой Круг**  
[новые вакансии](#)

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

- Количество новостных запросов к Яндексу
- Записи в блогах

# 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

# Топ-10 самых цитируемых агентств (в алфавитном порядке)

«Ведомости»

«Газета.ru»

«Интерфакс»

«Коммерсант»

РИА «Новости»

«Спорт-Экспресс»

«Эхо Москвы»

BBCRussian

Lenta.ru

ИА REGNUM

По данным отчета «Медиафера Рунета» на конец 2008 года  
<http://company.yandex.ru/researches/>

# 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

# Робот и человек в Яндекс.Новостях

• Человек

• Робот

• Для правки структуры



– Пятый  
уровень



– Пятый  
структу  
уровень  
– Шестой



# Наиболее часто встречающиеся проблемы

«

- Наше сообщение не попало в сюжет
- В Яндекс.Новости попала перепечатка нашей заметки
- Мы считаем, что получаем от вас недостаточно трафика
- Мы изменили текст сообщения на своем сайте. Внесите эти изменения в вашу базу данных

»



**Хотите узнать больше  
о работе Яндекс.Новостей?  
Приходите к нам на семинар!**

**Я**

**Запись по адресу  
[info@news.yandex.ru](mailto:info@news.yandex.ru)**

**Спасибо за внимание! Вопросы?**

**Лев Гершензон, Татьяна Исаева**