

Приложение № 1
к Соглашению
об информационном сотрудничестве,
опубликованному 27.04.2004 г.
(<http://partner.news.yandex.ru/agreement.pdf>)
с изменениями от 17.08.2011г.

Дата последнего изменения – 27 августа 2015 г.

ТЕХНИЧЕСКИЕ ТРЕБОВАНИЯ

Экспорт Данных для размещения заголовков и аннотаций новостей на Яндекс.ру (в том числе на сайте Яндекс.Новости) осуществляется в XML-based (<http://www.w3.org/TR/REC-xml>) формате RSS 2.0 (<http://blogs.law.harvard.edu/tech/rss>). Ниже содержится описание используемых для экспорта Данных элементов RSS 2.0, необходимые комментарии и пример экспортного файла.

1. Описание элементов RSS 2.0, используемых для экспорта Данных

Корневым элементом RSS-файла является **<rss>**, атрибут `version` которого должен иметь значение 2.0:
`<rss xmlns:yandex="http://news.yandex.ru" xmlns:media="http://search.yahoo.com/mrss/" version="2.0">`

Внутри элемента **<rss>** содержится элемент **<channel>**, который включает информацию об источнике и его содержание. Элементами **<channel>** считаются следующие элементы:

<title> – название RSS-потока.

В случае, если экспортируется содержание целого сайта, то здесь должно быть его название, например: `<title>Российские новости</title>`; если же части сайта, то в названии должно быть отражено, какая именно часть, например: `<title>Российские новости: технологии</title>`. На Яндекс.ру название RSS-потока не показывается, экспортируемые Данные маркируются тем названием источника, которое было указано в анкете.

<link> – URL сайта, данные которого транслируются в потоке.

Пример: `<link>http://www.rossiyskie-novosti.ru</link>`

<description> – описание потока. Одно предложение.

Пример: `<description> Ежедневная московская общественно-политическая газета.</description>`

<yandex:logo> – стандартный логотип. Обязательный элемент!

<yandex:logo type="square"> – квадратный логотип. Обязательный элемент!

В `<channel>` должны быть добавлены два элемента: `<yandex:logo>`, со ссылкой на стандартный логотип, и `<yandex:logo type="square">`, со ссылкой на крупный квадратный логотип.

Пример:

`<yandex:logo>http://example.net/normal_logo.png</yandex:logo>` -> стандартный

`<yandex:logo type="square">http://example.net/square_logo.png</yandex:logo>` -> квадратный

В `<yandex:logo>` нужно указать ссылку на стандартный логотип издания в формате PNG, GIF или JPG/JPEG. Рекомендуемая длина максимальной стороны – 100 пикселей. Этот логотип отображается на странице издания в Базе данных СМИ <http://news.yandex.ru/smi/>.

Атрибут `type` у квадратного логотипа может быть только `square`! Тип и размер квадратного логотипа, передаваемого посредством `<yandex:logo type="square">`, – PNG, 180*180 пикселей, при этом его ширина и высота должны быть одинаковыми. Крупный квадратный логотип отображается в Яндекс.Новостях рядом с заголовками новостей.

Если вы решите использовать один и тот же логотип в качестве квадратного и стандартного, ссылку на него всё равно необходимо указать в обоих элементах – и в `<yandex:logo>`, и в `<yandex:logo type="square">`.

Старая схема с передачей логотипа в `<image>` пока поддерживается (его допустимо использовать вместо `<yandex:logo>`). Однако после добавления новых тегов `<yandex:logo>` и `<yandex:logo type="square">`, тег `<image>` и вложенные в него теги `<url>`, `<title>`, `<link>` нужно убрать.

<item> – Обязательный элемент!

Каждый `<item>` описывает только одно(!) сообщение и должен содержать необходимый элемент `<title>` – заголовок сообщения. В `<channel>` может содержаться любое количество элементов `<item>`.

<title> – заголовок сообщения. Обязательный элемент!

Пример: `<title>Яндекс ищет на президентском сайте</title>`

Внимание: Написание заголовка `<title>` целиком ПРОПИСНЫМИ буквами не допускается. Не допускается также наличие точки в конце заголовка. В заголовке не должны содержаться название источника и дата/время сообщения, а также служебные примечания ("обновлено", "дополнено", "фоторепортаж", "видео" и др.) и неинформативные обороты, не представляющие собой неотъемлемой части заголовка ("Срочно!", "Сенсация:" и т.п.).

<link> – URL сообщения. Обязательный элемент!

Пример: `<link>http://www.rossiyskie-novosti.ru/2003/03/25/yandex.html</link>`

Внимание: каждое сообщение должно располагаться на отдельной странице, открывающейся по указанному адресу, при этом заголовок и начало текста сообщения должны быть доступны в первом экране при разрешении 1024x768. При переходе с заголовка, размещенного на Яндекс.Новостях, должна открываться только одна страница, содержащая сообщение, соответствующее заголовку.

Наличие по URL, указанному в <link>, более одной новости (ленты новостей) не допускается. URL, различающиеся только в части после '#' (только «якорями»), т.е. вида <http://www.some-host.ru/news.html#2545> и <http://www.some-host.ru/news.html#5794>, считаются идентичными и НЕ допускаются.

<pubDate> – время публикации сообщения на сайте источника. Обязательный элемент!

(в формате RFC-822 – см. <http://www.w3.org/Protocols/rfc822/#z28>)

Регистр в названии <pubDate> имеет значение: буква D должна быть прописной.

Пример: <pubDate>Tue, 21 Apr 2015 14:15:00 +0300</pubDate>

Эта запись означает, что новость датирована 21 апреля 2015 г., 14:15 московского времени.

Внимание: +0300 является указанием на часовой пояс (в приведенном примере это московское время). Указанное в экспортном файле время должно обязательно совпадать с фактическим временем публикации на сайте! Актуальными считаются сообщения за 8 дней; остальные проиндексированы не будут.

<yandex:full-text> – для экспорта полного текста сообщений. Обязательный элемент!

Кроме стандартных элементов RSS 2.0, для экспорта Данных используется специальный элемент <yandex:full-text>, который должен содержать полный текст сообщения. Полный текст сообщения необходим для индексирования поисковым роботом и на Яндекс.ру размещаться не будет.

В полном тексте не должны содержаться:

1. название источника
2. дата/время сообщения
3. контактная информация
4. ссылки на изображения, аудио- и видеофайлы; для них нужно формировать отдельные теги <enclosure>, <media:group>

Пример:

<yandex:full-text>ПОЛНЫЙ ТЕКСТ НОВОСТИ</yandex:full-text>

<pdalink> – ссылка на адаптированную для мобильных устройств версию сообщения.

Чтобы эта версия материалов источника была доступна на <http://pda.news.yandex.ru/> и <http://m.news.yandex.ru/>, необходимо формировать в экспортном файле тег <pdalink>, в который включается ссылка на соответствующее сообщение на мобильной версии сайта источника.

Пример: <pdalink><http://www.m.rossiyskie-novosti.ru/2003/03/25/yandex.html></pdalink>

<description> – аннотация сообщения.

Пример: <description>Продукт Яндекс.Site установлен на сайте президента России</description>

<yandex:genre> – жанр сообщения.

Здесь нужно указать латиницей жанр сообщения:

lenta (короткое новостное сообщение, 50-80 символов)

message (более развернутое новостное сообщение)

article (статья)

interview (интервью)

Пример: <yandex:genre>article</yandex:genre>

<author> – автор сообщения.

Пример: <author>Иван Петров</author>

<category> – рубрика (раздел, категория) сообщения.

Здесь должно быть размещено название рубрики (оригинальное, как в издании), в которой опубликовано сообщение. одному сообщению может соответствовать только одна рубрика!

Пример: `<category>Технологии</category>`

Внимание: Об изменении рубрикации издания или о добавлении в экспортный файл материалов новых рубрик необходимо сообщать по адресу info@news.yandex.ru. Без такого уведомления сообщения, принадлежащие ранее не существовавшим или переименованным рубрикам издания, не индексируются.

<enclosure> – элемент для иллюстраций, аудио- и видеофайлов.

Если в сообщении содержится несколько иллюстраций, или иллюстрация и видеофайл, элемент `<enclosure>` повторяется. Принимаются иллюстрации с шириной не менее **240** пикселей и длиной не менее **300** пикселей. Если есть несколько вариантов одной иллюстрации, отличающихся размером, то в `<enclosure>` должен быть указан URL изображения наибольшего размера.

Внимание: иллюстрации размером менее **300*240** пикселей не индексируются.

Иллюстрации должны быть разрешены к индексированию в файле robots.txt. Дополнительную информацию о robots.txt можно посмотреть здесь: <http://help.yandex.ru/webmaster/controlling-robot/robots-txt.xml>

Для изображений параметр `type` должен совпадать с тем, что отдается по указанному URL.

Значение `url` обязательно, значение `type` крайне рекомендуется и обязательно в случае, если невозможно определить тип контента по расширению.

Пример: `<enclosure url="http://www.rossiyskie-novosti.ru/2003/03/25/yandex.jpg" type="image/jpeg"/>`

<media:group> – позволяет объединить связанные медиа-объекты. Например, можно объединить два видео, отличающиеся лишь форматами, плеер и тумбнейл. Различные видео объединять не рекомендуется.

Предлагаем использовать `<media:group>` для более качественного индексирования видеофайлов, сопровождающих сообщение, вместе с тегом `<enclosure>` или вместо него.

В теге могут быть несколько вложенных тэгов `<media:content>`, в одном из которых может быть проставлен атрибут `isDefault`.

Допускается лишь один тэг `<media:player>`.

Допускаются несколько тэгов `<media:thumbnail>`, при этом указывать их следует в порядке убывания приоритета.

Внимание: ширина тумбнейла должна быть не меньше 240 пикселей, длина — не меньше 300 пикселей. Видео, сопровождающееся тумбнейлом размером менее 300*240 пикселей, не будет проиндексировано. Также допускается группировать вместе видео и аудио, если аудио является аудиодорожкой к видео. Если в группе содержится тэг `<media:player>`, то вместо прямой ссылки на файл укажите ссылку на плеер. Можно указать любое количество тэгов `<media:group>`. Вложенные группы не допускаются. Атрибутов нет.

Пример:

`<media:group>`

`<media:content url="ссылка на видео-файл на вашем сайте"/>`

`<media:player url="ссылка на плеер на вашем сайте"/>`

`<media:thumbnail url="ссылка на иллюстрацию, которая должна быть использована в качестве preview"/>`

`</media:group>`

Внимание: Preview может быть несколько. Ссылка на видео или на плеер обязательна.

Видео- и фоторепортажи можно транслировать в Яндекс.Новости даже в том случае, если они не сопровождаются опубликованным на вашем сайте текстом или расшифровкой.

Для трансляции такого видео необходимо в соответствующем <item>:

- сформировать тег <description>, который должен содержать лид (краткое описание) видеорепортажа, если оно есть и опубликовано на сайте;
- сформировать тег <yandex:tags>, в который поместить ключевые слова для этого видео, повторяя этот элемент для каждого слова (то есть каждое ключевое слово должно быть в отдельном элементе <yandex:tags>). Вы можете не публиковать эти ключевые слова у себя на сайте, они нужны для успешной работы алгоритма кластеризации, минимальное их количество - 5.

Пример:

```
<yandex:tags>путин</yandex:tags>  
<yandex:tags>порошенко</yandex:tags>  
<yandex:tags>меркель</yandex:tags>  
<yandex:tags>лукашенко</yandex:tags>  
<yandex:tags>минск</yandex:tags>
```

- обязательно в соответствующем теге указать ссылку на само видео.

Для трансляции фоторепортажей необходимо в соответствующем <item>:

- указать жанр сообщения <yandex:genre>photogallery</yandex:genre>;
- сформировать тег <description>, который должен содержать лид (краткое описание) фоторепортажа, если оно есть и опубликовано на сайте;
- сформировать тег <yandex:tags>, в который поместить ключевые слова для этого фоторепортажа, повторяя этот элемент для каждого слова (то есть каждое ключевое слово должно быть в отдельном элементе <yandex:tags>). Вы можете не публиковать эти ключевые слова у себя на сайте, они нужны для успешной работы алгоритма кластеризации, минимальное их количество -5.

Пример:

```
<yandex:tags>путин</yandex:tags>  
<yandex:tags>порошенко</yandex:tags>  
<yandex:tags>меркель</yandex:tags>  
<yandex:tags>лукашенко</yandex:tags>  
<yandex:tags>минск</yandex:tags>
```

- обязательно в соответствующих тегах указать ссылки на все иллюстрации, входящие в фоторепортаж.

<yandex:related> – если на странице источника для сообщения указаны ссылки на другие, в том числе не новостные, источники (сайты по теме), необходимо добавить в <item> этого сообщения ссылки на них. Для этого формируется специальный блок <yandex:related>. Число элементов <link> внутри этого блока может быть любым.

Пример:

```
<yandex:related>  
<link url="http://www.kremlin.ru/">Президент России</link>  
</yandex:related>
```

<yandex:online> – если в вашем экспорте для Яндекс.Новостей есть ссылки на онлайн-репортажи различных событий и/или онлайн-пресс-конференции, в **<item>** сообщения-трансляции в вашем обычном фиде для Яндекс.Новостей нужно включить тег **<yandex:online>** и поместить в него ссылку на ещё один фид, в который нужно положить непосредственно в текст трансляции:

Пример:

```
<item>
<title>заголовок новости</title>
<pubDate>дата сообщения</pubDate>
<yandex:full-text>полный текст сообщения</yandex:full-text>
<yandex:online> http://babruysk.com/translation.xml</yandex:online>
<yandex:genre>online</yandex:genre>
</item>
```

<yandex:online> <http://babruysk.com/translation.xml>**</yandex:online>** – структура экспортного файла онлайн-трансляции должна быть такой:

```
<?xml version="1.0" encoding="UTF-8" ?>
<rss version="2.0">
<channel>
<заголовок трансляции (не обязательно)>
<title>Революция в Бобруйске</title>
<url трансляции на сайте (не обязательно)>
<link>http://www.babruysk.com/revolution.html</link>
<дата обновления (не обязательно)>
<pubDate>Mon, 27 Dec 2011 16:45:00 +0000</pubDate>
<item>
<само сообщение>
<description>Бобруйские рабочие предприняли попытку атаковать здание администрации</description>
<url сообщения, не обязательно>
<link>http://www.babruysk.com/revolution.html#1488</link>
<время сообщения>
<pubDate>Mon, 27 Dec 2011 16:45:00 +0000</pubDate>
</item>
</channel>
</rss>
```

Обязательные элементы в этой структуре – pubDate внутри item и description внутри item.

item = одна «реплика» внутри трансляции. Чтобы ссылка на сообщение-трансляцию попала в список «Все онлайн-трансляции», в теге для определения жанра сообщения (новость/статья/интервью) указать жанр сообщения online. Пример: **<yandex:genre>online</yandex:genre>**

Внимание: Тег **<yandex:online>** нужно формировать только для трансляций онлайн-конференций, онлайн-репортажей и т.п. Онлайн-трансляции спортивных событий мы по-прежнему индексируем с помощью формируемых вами специальных экспортных файлов, технические требования к которым опубликованы в разделе "Технические документы" партнёрского интерфейса.

2. Символы и кодировки

По умолчанию (если это не указано явно в заголовке) кодировкой файла считается utf-8. В противном случае выставление кодировки xml-файла обязательно. Наиболее часто употребляемые кодировки: windows-1251, utf-8, koï8-r.

Внимание: фактическая кодировка, отдаваемая веб-сервером, ВСЕГДА должна совпадать с кодировкой, указанной в заголовке XML!

Встречающиеся в тексте символы < > & ' " необходимо заменять на соответствующие элементы:

& на &

< на <

> на >

' на '

" на "

(здесь точка с запятой – это не разделитель данного списка, а обязательная часть элемента!)

Замены должны производиться во всех элементах <item> и <channel> – в <yandex:full-text>, <link>, <title>, <enclosure> и др.

Пример, ссылка "http://some.host.ru/?id=1&page=10" приводится к виду
"http://some.host.ru/?id=1&page=10"

В случае, если RSS-файл передается в koï8-r, необходимо также заменить встречающиеся в тексте символы кодировки windows-1251 на аналоги из koï8-r:

многоточие – код символа 133

en-dash (короткое тире) – код символа 150

em-dash (длинное тире) – код символа 151

"Русский" номер – код символа 185

Кавычки-"ёлочки" – коды символов 171 и 187

"Сглаженные" кавычки-"лапки" – коды символов 147 и 148

"Сглаженные" апострофы – коды символов 145 и 146

3. Механизм экспорта Данных

Для экспорта Данных необходимо положить RSS-файл на сервер издания и обновлять его с определенной периодичностью (например, файл может пополняться в течение дня и перезаписываться утром). Файл должен быть доступен по http, его индексирование (скачивание) происходит каждые 5 минут. Экспортный файл, который не удалось полностью загрузить за 10 секунд, считается недоступным.

Для корректной индексации необходимо, чтобы у робота Яндекса было разрешение на скачивание RSS-фида в файле **robots.txt**.

Делается это следующим образом:

В начале файла robots.txt надо добавить следующие 3(!) строки:

User-agent: Yandex

Allow: путь до фида без имени хоста, например /file.rss

Третья строка должна быть пустая.

Добавить их нужно обязательно в начале, а не в конце!

(Проверить корректность добавления можно тут <http://webmaster.yandex.ru/robots.xml>)

Если Вы используете в robots.txt директиву Crawl-delay, то необходимо убедиться, что ее значение достаточно мало, чтобы робот мог оперативно перекачивать RSS-фид.

4. Пример экспортного файла (Обязательные элементы выделены цветом)

```
<?xml version="1.0" encoding="windows-1251"?>
<rss xmlns:yandex="http://news.yandex.ru" xmlns:media="http://search.yahoo.com/mrss/" version="2.0">

<channel>

<title>Российские новости</title>
<link>http://www.rossiyskie-novosti.ru/</link>
<description> Ежедневная иллюстрированная московская общественно-политическая газета.</description>

<yandex:logo>http://www.rossiyskie-novosti.ru/normal_logo.png</yandex:logo>
<yandex:logo type="square">http://www.rossiyskie-novosti.ru/square_logo180x180.png</yandex:logo>

<item>
<title>Яндекс ищет на президентском сайте</title>
<link>http://www.rossiyskie-novosti.ru/2003/03/25/yandex.html</link>
<pdalink>http://www.m.rossiyskie-novosti.ru/2003/03/25/yandex.html</pdalink>
<description>Программный продукт Яндекс.Site установлен на официальном сайте президента
России</description>

<author>Иван Петров</author>
<category>Технологии</category>

<enclosure url="http://www.rossiyskie-novosti.ru/2003/03/25/yandex.jpg" type="image/jpeg"/>
<enclosure url="http://www.rossiyskie-novosti.ru/2003/03/25/yandex1.jpg" type="image/jpeg"/>
<enclosure url="http://www.rossiyskie-novosti.ru/video/100237" type="video/x-ms-asf"/>

<pubDate>Sun, 29 Sep 2002 19:59:01 +0300</pubDate>

<yandex:genre>message</yandex:genre>

<yandex:full-text>Для поиска по сайту www.kremlin.ru выбрана программа Яндекс.Site. Этот программный
продукт был исследован провайдером президентского сайта – Федеральным агентством правительственной
связи и информации. ФАПСИ сочло возможным использование поисковой системы &lt;Яндекса&gt; на
www.kremlin.ru. По результатам исследования программа была скомпилирована, протестирована и
установлена на сайт </yandex:full-text>
<yandex:online> http:// www.rossiyskie-novosti.ru /translation.xml</yandex:online>

<yandex:related>
<link url="http://www.kremlin.ru/">Президент России</link>
</yandex:related>

</item>

</channel>
</rss>
```