

## SiteSputnik News. Текстовая или рукописная форма новостного Проекта

### Командный способ формирования Проектов

#### Часть 1

- В данной публикации показано, как без особого труда можно создавать Проекты для мониторинга открытых Источников интернета.
- Проекты своими руками можно создать, ознакомившись с содержанием **Части 1** данной публикации, то есть, буквально, через 10 минут после установки программы.
- Для этого достаточно освоить всего два типа команд, которые по формату аналогичны командам, используемым в ОС Windows: **ИмяКоманды [Операнды]**.
- Рассматриваемая форма Проекта называется **текстовой** или **рукописной**. Все действия по проектированию выполняются на **одном листе**.
- Пояснения сделаны на примере **готового Проекта**, взятого из сообщения на форуме СПКР о [Динамических рубриках СайтСпутника](#).
- Этот Проект выглядит следующим образом:

```
http://www.volgograd-trv.ru/rss/  
http://xml.v102.ru/razdel.php?183  
http://vd-tv.ru/rss.xml  
http://krestyane34.ru/ +.html
```

```
rubric://Россия/ЮФО/Волгоград/Областная администрация  
губернатор /2 (Волгоградской /1 области | Бочаров)
```

```
rubric://Россия/ЮФО/Волгоград/Сельское хозяйство/Весенние работы  
Волгоград /7 (озимые | яровые | посевная |сельхоз*)  
КП: C:\Волгоград\Апрель
```

- В Проекте задействованы два типа команд: «**Определение Источника**» и «**Определение Рубрики**».
  - Первые четыре строки – это четыре команды типа «**Определение Источника**».
    - 1-3 строки – это живые RSS-потоки. Команды имеют только имя. Операндов в данном случае нет.
    - 4 строка – это эмуляция RSS-потока, а именно: со страницы <http://krestyane34.ru/> выбираются и отправляются на мониторинг (скачивание, рубрикации и некоторые другие операции) только те страницы, в адресе которых есть лексема **.html**. Эта лексема является операндом. Для того, чтобы понять какой операнд следует добавить в команду, в большинстве случаев достаточно открыть в браузере или СайтСпутнике ссылку-источник и посмотреть какие лексемы характерны для размещенных на ней **адресов ссылок на полное содержание статей или сообщений**. Если это не понятно, то пришлите Вашу ссылку-источник на E-mail [1a@sitesputnik.ru](mailto:1a@sitesputnik.ru), и Вам подскажут полный формат команды. Одной-двух подсказок будет достаточно для того, чтобы далее самостоятельно формулировать команды данного типа.
  - 5-6 и 7-9 строки – это две команды типа «**Определение Рубрики**».
    - Имя первой команды - `rubric://Россия/ЮФО/Волгоград/Областная администрация`. Оно определяет название Рубрики. Операнд всего один - `губернатор /2 (Волгоградской /1 области | Бочаров)`. Он определяет Правило

попадания информации в Рубрику. В данном случае это сведения о губернаторе области. Примеры основных Правил и пояснения к ним приведены в Приложении 1.

- Имя второй команды - <rubric://Россия/ЮФО/Волгоград/Сельское хозяйство/Весенние работы>. Операндов два. Второй из них - КП: C:\Волгоград\Апрель. Он определяет дополнительную папку локального компьютера (клиентскую папку), в которую будет отправлено содержание данной Рубрики.
- Собственно, все «начальные трудности» позади, можно перейти к созданию своих Проектов. Последовательность действий такова:
  - Нажмите на «Золотой ключик», затем на кнопку «Пакет».
  - Если Вы увидите табличную форму формирования Проекта, то лучше её перевести в текстовую форму: нажмите на кнопку-иконку с подсказкой «К текстовой форме Проекта».
  - Введите свои команды. Можно ввести или живьём скопировать приведенные выше команды, чтобы протестировать программу.
  - По кнопке «Сохранение» сохраните сделанный Проект в библиотеке.
  - По кнопке «Совместно» отправьте его на выполнение.
  - Далее совершенствуйте созданный готовый Проект:
    - добавляйте в него новые Источники и Рубрики,
    - оттачивайте Правила попадания в Рубрики,
    - изучайте [другие возможности СайтСпутника](#) и добавляйте их в Проект.

\* \* \*

## Часть 2

### О саморазвивающемся Проекте

- Можно применить [модуль +NewStreams](#), который в автоматическом режиме будет находить новые Источники и проверять: содержат ли они интересующую Вас информацию. Эта проверка осуществляется следующим образом. Все попадающие в поле зрения Проекта ссылки анализируются СайтСпутником на предмет нахождения в их контенте [адресов ссылок](#), являющихся потоками, приносящими информацию в Рубрики Проекта. Проект будет сам наблюдать за изменяющимся информационным пространством и предлагать Вам новые потоки для включения их в Проект. Вам достаточно будет контролировать действия автоматике. Освоение этого модуля также не требует больших усилий.

### Подробнее о двух основных командах

#### **Первая команда: «Определение Источников»**

- Имя команды - это URL ссылка, далее могут следовать операнды, например: [http://krestyane34.ru/ +.html /fast](http://krestyane34.ru/+ .html /fast)
- Имя команды начинается с префикса <http://> или <https://>.
- В Проекте таких команд может быть несколько. Находиться они могут в любом месте Проекта.
- Ещё подробнее о команде - в приведенной ниже Таблице 1.

#### **Вторая команда: «Определение Рубрики»**

- Это имя рубрики и от одного до трех операндов, например:  
[rubric://Россия/ЮФО/Волгоград/Сельское хозяйство/Весенние работы Волгоград /7 \(озимые | яровые | посевная |сельхоз\\*\)](rubric://Россия/ЮФО/Волгоград/Сельское хозяйство/Весенние работы Волгоград /7 (озимые | яровые | посевная |сельхоз*))  
КП: C:\Волгоград\Апрель  
[Искать только в родительской](#)

Здесь к двум операндам, пояснения к которым даны выше, добавлен третий операнд «Искать только в родительской»

- В данном случае третий операнд имеет смысл, если в Проекте задействована родительская Рубрика:  
[rubric://Россия/ЮФО/Волгоград/Сельское хозяйство](#)  
В нашем Проекте она не задействована.
- Имя команды начинается с префикса [rubric://](#) или [рубрика://](#). Префиксы эквивалентны.
- В Проекте таких команд может быть несколько. Находиться они могут в любом месте Проекта.

\* \* \*

### **Таблица 1. Основные команды SiteSputnik News**

Красным цветом выделены две основные команды: «Определение источников» и «Определение Рубрик».

№	Префикс	Корень	Операнды	Пояснения	Способ *
1	<a href="#">http://</a> <a href="#">https://</a>	Ссылка конкретная	+L1;L2;LN –m1;m2;mN	L1,L2,LN – лексемы, хотя бы одна из которых содержится в адресе ссылки, попадающей на мониторинг. Лексема может содержать знаки <b>&amp;&amp;</b> , например, <a href="#">news&amp;&amp;tv/</a> - это означает, что в ссылке должны быть обе лексемы: и <a href="#">news</a> , и <a href="#">tv/</a> . m1,m2,mN – запрещённые лексемы.	/fast /slow
2	<a href="#">file:///</a>	Имя файла локального		Аналогично пункту № 1, только рассматривается не Ссылка конкретная, а конкретный файл на Вашем компьютере. Это <a href="#">локальные источники информации</a> . Ограничение временное: файл может быть или html, или текстового типа. В файле должны быть ссылки.	
3	<a href="#">file://///</a>	Имя файла локальной сети		Аналогично пункту № 2, только рассматривается конкретный файл из локальной сети. Это <a href="#">источники информации, размещённые в Вашей локальной сети</a> .	
4	<a href="#">folder:///</a>	Имя папки локального компьютера		Аналогично пункту № 1, только рассматривается не Ссылка конкретная, а конкретная папка на Вашем компьютере. И лексемы ищутся не в адресах ссылок, а в именах файлов, находящихся в папке и всех вложенных в неё папок. Это мониторинг <a href="#">файлов на Вашем компьютере</a> .	
5	<a href="#">folder://///</a>	Имя папки локальной сети		Аналогично пункту № 4, только рассматривается папка из локальной сети. Это мониторинг <a href="#">файлов из Вашей локальной сети</a> .	
6	<a href="#">rubric://</a> <a href="#">рубрика://</a>	Имя рубрики	Правило попадания в Рубрику	Смотрите ниже Приложение 1.	-
			КП: f1;f2;f3	f1,f2,f3 – папка, папка локальной сети, E-mail или папка сайта	
			Искать только в родительской	<a href="#">Правила попадания в Рубрику</a> проверяются только для ссылок, попавших в родительскую Рубрику	
7		Поиск по запросу	Имя Источника из «Настройка поиска»	Поиск по запросу в поисковики и встроенные в сайты поиски, например: <a href="#">президент России    Yandex=7    news=3</a>	/fast /slow
8				В качестве команды может использоваться другой готовый Проект. Подробнее на <a href="#">ссылке</a> по коричневой стрелке.	

\* Параметры /fast и /slow – явная прописка быстрого или обычного способа открытия Источника. Подробнее на [ссылке](#).

### **О макрокомандах**

- Макрокоманды действуют на все команды, заданные в Проекте.

- Они имеют более высокий приоритет, чем установленные в Настройках параметры, определяющие аналогичные функции.
- Запоминать как пишутся макрокоманды не обязательно, все они доступны в диалоге из контекстного меню.

**Таблица 2. Основные макрокоманды**

№	Название	Пояснения
1	^^Новости;	Проект типа «Новости». Определяются новые ссылки, они скачиваются и отправляются на Рубрикацию. Если в Проекте прописаны Рубрики, то эта макрокоманда может быть опущена.
2	^^WebSpider;	Проект типа «Обновление». Все ссылки скачиваются, сравниваются с их предыдущим состоянием, выделяются новые фрагменты, которые и отправляются на Рубрикацию.
3	^^Рубрики	Указывается имя статической (созданной заранее) Рубрики. Подробнее на <a href="#">ссылке</a> .
4	^^Комментарии;	Выделяются комментарии к найденным ссылкам. Они отправляются на Рубрикацию на те же Рубрики, что и сами ссылки. Замечание: выделить комментарии удастся не всегда.
5	^^Комментарии+Рубрики	Аналогично пункту 4, но Рубрикация на «своих» Рубриках.
6	^^Сигналы	Подключение Сигнализации о значимых событиях. Подробнее на <a href="#">ссылке</a> .
7	^^RSS-потоки новые=3 ^^URL-потоки новые=4	Указывает максимальную глубину <a href="#">поиска RSS-потоков новых</a> , несущих искомую информацию. Указывает максимальную глубину <a href="#">поиска URL-потоков новых</a> , несущих искомую информацию.
8	^^Одинаковый контент+ ^^Одинаковый контент-	Отсеивать новости и сообщения с одинаковым контентом. Анализ на одинаковый контент не выполняется.
9	^^Экспорт; ^^Импорт;	<a href="#">Экспорт</a> в Клиентские папки Чистого потока. <a href="#">Импорт</a> из Клиентских папок Чистых потоков.
10	^^Пауза=4	Задаёт 4-х секундную паузу после окончания обработки очередного элементарного Источника.
	^^Пауза.Yandex=3 ^^Пауза.Google=4 ^^Пауза.Yahoo=2	«Персональные паузы» после открытия страниц Яндекса, Google и Yahoo.

**Приложение 1. Некоторые Правила попадания в Рубрики на примерах**

- Со всеми Правилами можно ознакомиться на [ссылке](#).
- Для начала работы рекомендуется ознакомиться со следующими правилами:
  - **Сергей Иванов** - где-то в тексте обязательно есть словоформы от Сергей и Иванов  
**Сергей /2 Иванов** - между этими словоформами может быть не более 2-х пробелов (не более 1-го слова).  
**Сергей & Иванов** - эти словоформы должны быть в одном предложении.  
**Сергей && Иванов** - эти словоформы должны быть в одном абзаце.  
**'Сергей Иванов'** - строго в тексте должно быть написано то, что Вы заключили в одинарные кавычки.  
**(Сергей | Иванов)** - хотя бы одно из двух слов в тексте.  
**Сергей ~Иванов** - Сергей есть, а Иванов нет.  
**!стройка** - отсечение слева, слово перестройка - не учитывается.  
**Алексе\*** - справа вместо \* - что угодно, Александр - не учитывается.  
**!Путин\*** - не учитываются: путь, пути, Распутин.  
**(!Путин\* | Медведев | Нарышкин) /4 (приехал | прилетел | прибыл) /3 Волгоград** - хотя бы один из трёх посетил Волгоград.
  - Допускается многоуровневая скобочная структура.
  - Слова в одной строке Правила (если нет специальных операторов) - логическое "И", то есть, должны быть все в тексте.
  - Строки соединяются логическим "ИЛИ", то есть, текст учитывается, если удовлетворяет хотя бы одной строке Правила.
  - **Регулярные выражения** могут быть прописаны в любом месте Правила.

**Все подробности:**

- на [главной странице сайта о программе](#) и
- в [перечне публикаций о программе](#).

### **Обратная связь**

- По всем вопросам можно обращаться на форум СПКР в раздел [о программе SiteSputnik \(СайтСпутник\)](#);
- к автору программы по E-mail или телефонам, указанным на <http://sitesputnik.ru/>

[Copyright © 2003-2019 Индивидуальный предприниматель Мыльников Алексей Борисович. Все права защищены.](#)