

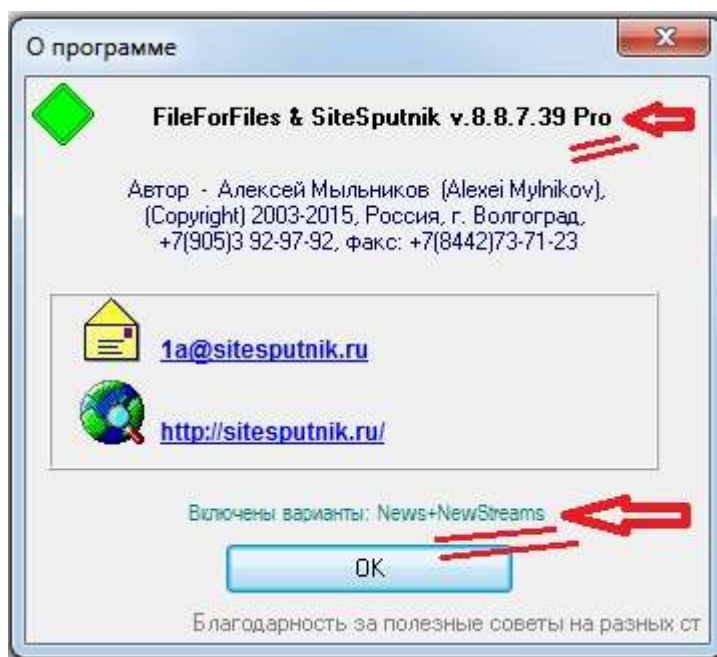
Инструкция по применению программы SiteSputnik Pro+News+NewStreams

Назначение программы

- Организация и автоматизация поиска Источников информации: именно **Источников, а не самой информации**.
- Мониторинг интернета на предмет появления в нём Источников, которые могут быть интересны.
- Модуль **+NewStreams** можно назвать поставщиком Источников для [SiteSputnik News](#) или других программ.
- Его также можно использовать для более глубокого мониторинга открытых источников интернета.

Установка программы

- Удалите демоверсию программы [FileForFiles & SiteSputnik](#) (сокращенно [SiteSputnik](#) или [СайтСпутник](#)), установленную на Вашем компьютере. Для этого достаточно прямо из Проводника удалить папку FileForFiles.
- Папку FileForFiles из персонального дистрибутива программы разместите, например, прямо в корне диска D:\.
- Откройте файл FileForFiles.exe и ответьте на два вопроса. Программа установится и откроется.
- Для того, чтобы убедиться, что Вы имеете дело с вариантом комплектации **SiteSputnik Pro+News+NewStreams** (далее **SiteSputnik NewStreams**), выберите из главного меню пункт: **? – О программе**. Отрывшееся окно должно иметь вид:



Обновление программы и переход на новые версии

- Рекомендуется систематически обновляться: **? – Проверить наличие обновлений**.
- Обновления и переходы на новые версии программы бесплатные.

Рекомендации

- Ознакомление с инструкцией по [SiteSputnik News](#) желательно.

Терминология и пояснения

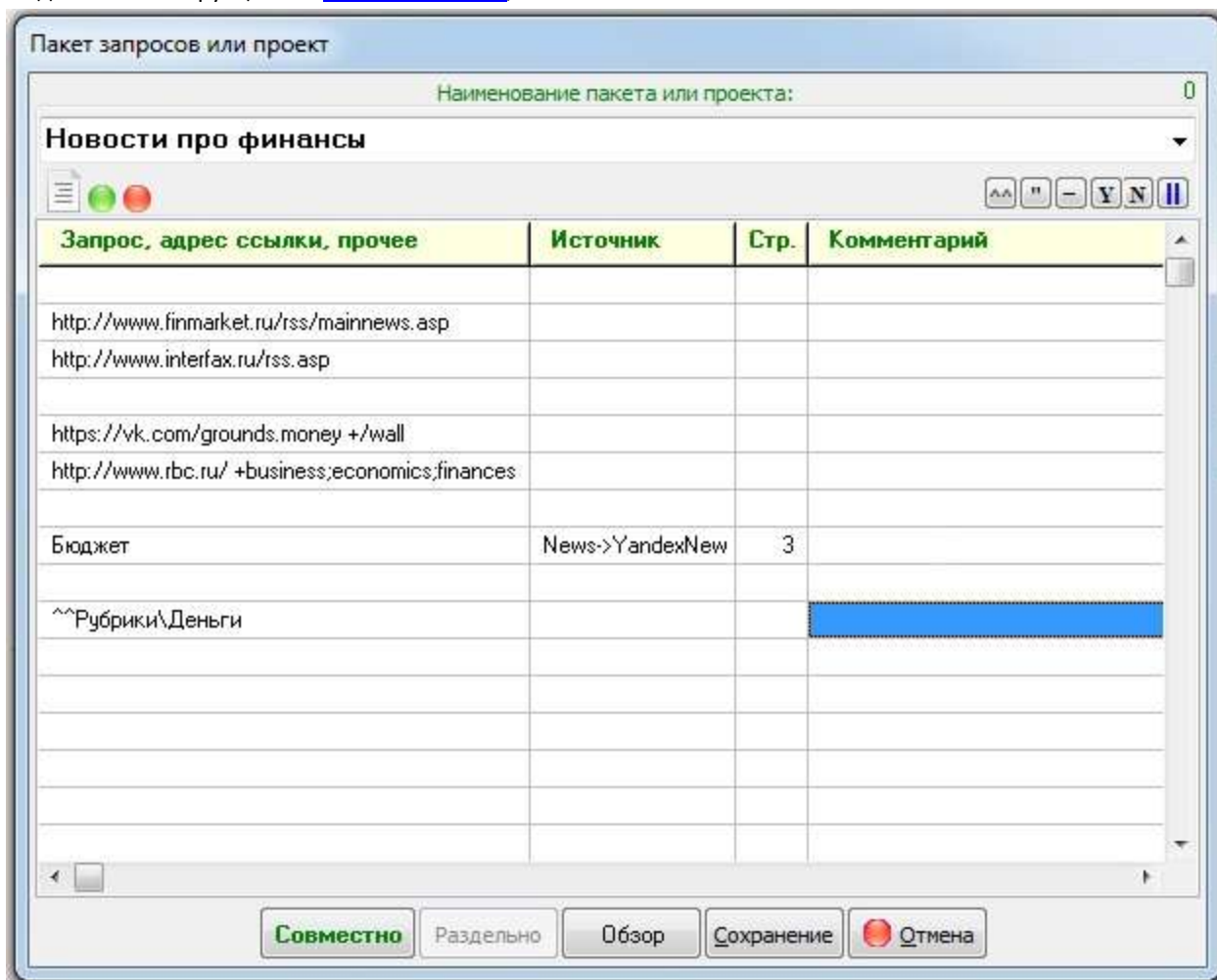
- **Источники**, которые СайтСпутник будет находить и автоматически включать в **Проект**, могут быть двух видов:
 - **RSS-потоки** - существующие на сайтах потоки информации.
 - **URL-потоки** – программная эмуляция потоков информации с произвольных страниц сайтов.
- **Проекты** могут быть следующих четырёх видов:
 - **Проект главный** – проект, который мы сделали сами и сохранили в Библиотеке готовых проектов.
 - **Проект расширенный** содержит обнаруженные и включенные СайтСпутником Источники. Критерий включения в Проект расширенный – Источник должен приносить информацию в Рубрики.
 - **Проект проверочный** формируется СайтСпутником из Источников-кандидатов на включение в Проект расширенный.
 - **Проект исполняемый** – проект, который выполняется. Выполнить можно Проект главный, проверочный, расширенный или исполняемый.

Другие существующие в СайтСпутнике виды Источников и Проектов в данной инструкции не затрагиваются.

- В настоящее время функционирует **бета-версия SiteSputnik NewStreams**.
- Пример применения **SiteSputnik NewStreams**, рассматриваемый ниже, носит учебный характер.

Простой проект

- Приступим к созданию Проекта для поиска новых Источников. В качестве Проекта главного возьмём Проект, созданный в инструкции по [SiteSputnik News](#), а именно:



и настроим его на поиск новых нужных Источников.

Напомню, что этот Проект:

- следит за появлением новой информации в пяти Источниках, а именно:
 1. RSS-поток с Финмаркета - <http://www.finmarket.ru/rss/mainnews.asp>,
 2. RSS-поток с Интерфакса - <http://www.interfax.ru/rss.asp>,
 3. URL-поток с ВКонтакте «Территория денег» из записей на стенах – <https://vk.com/grounds.money+/wall>,
 4. URL-поток с РБК о бизнесе, экономике и финансах – <http://www.rbc.ru/+business;economics;finances>,
 5. Поиск по запросу: бюджет, - в Яндекс.Новости на глубину 3 страницы,
- и создает на основе этой информации три тематические ленты в Рубриках:
 1. Деньги
 2. Сбербанк (вложена в Рубрику Деньги)
 3. Дефицит бюджета (вложена в Деньги).

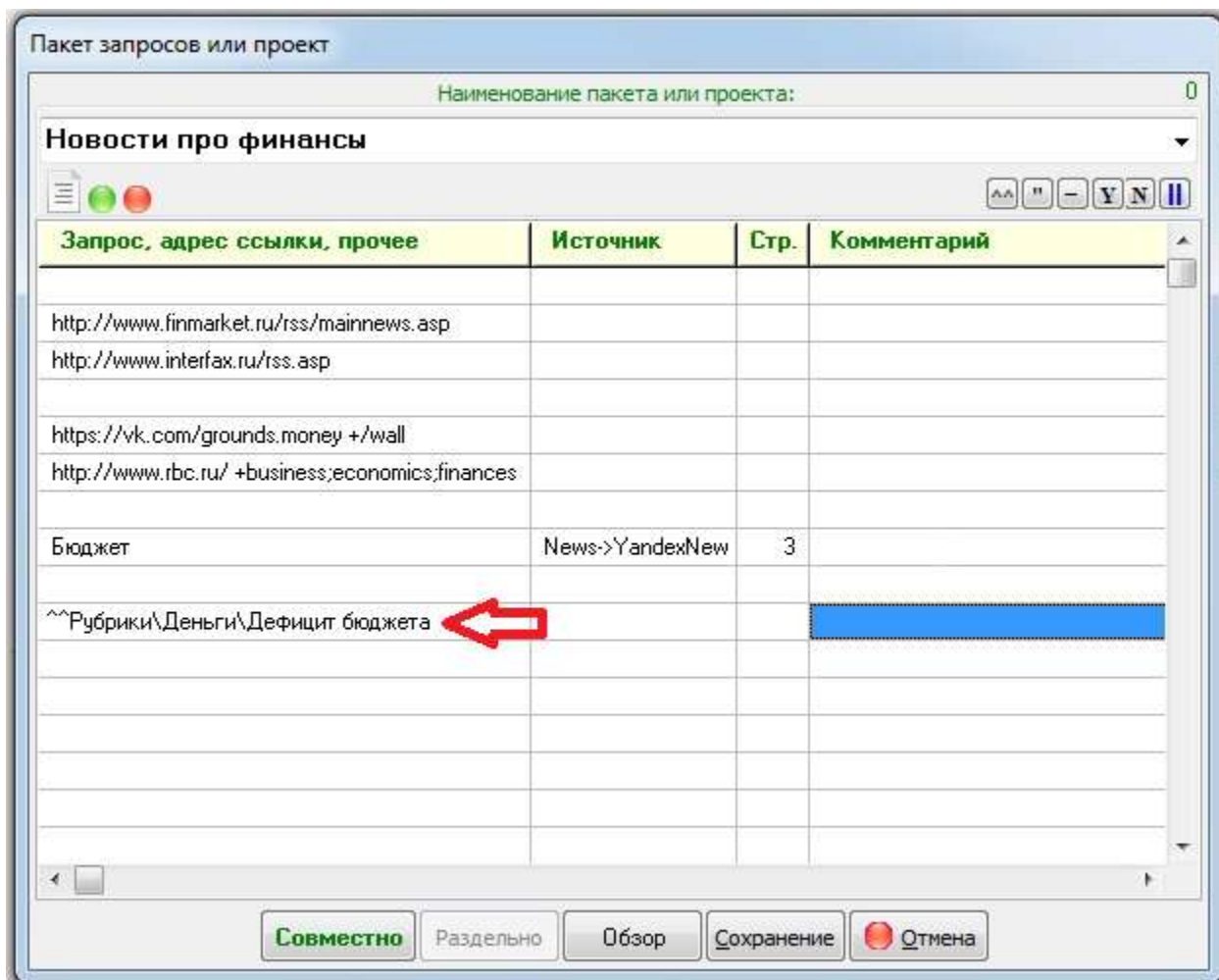
Допустим, что нас интересуют Источники, где речь может идти о дефиците бюджета. Мы хотим находить и добавлять в Проект такие и только такие Источники, в которых может обсуждаться дефицит бюджета. Правило попадания в Рубрики, которым мы в инструкции по [SiteSputnik News](#) описали наш интерес, было таким:

секвестр

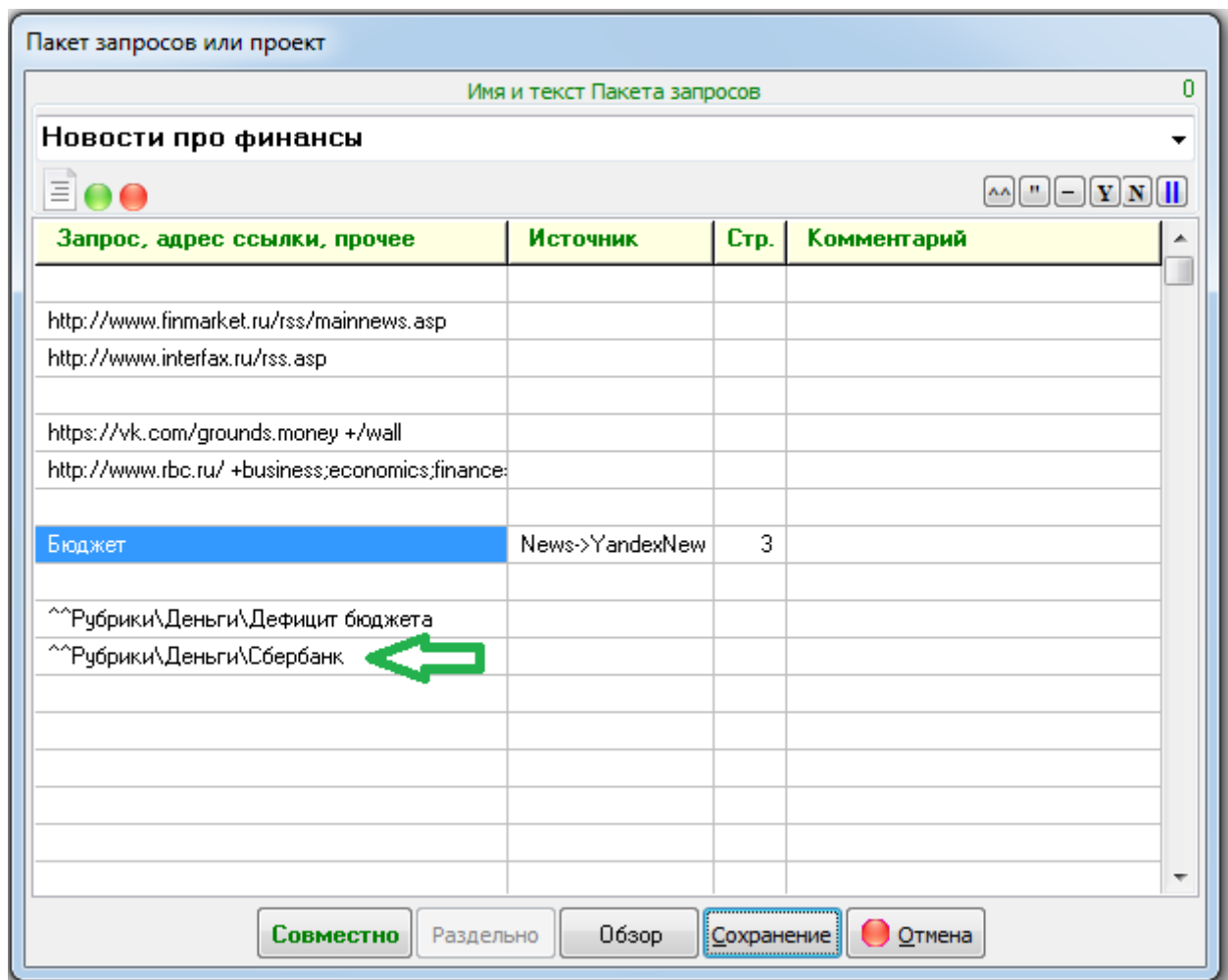
(дефицит | Кризис) /2 бюджета

(Сокращение | уменьшение | снижение) /2 расходов /5 бюджета

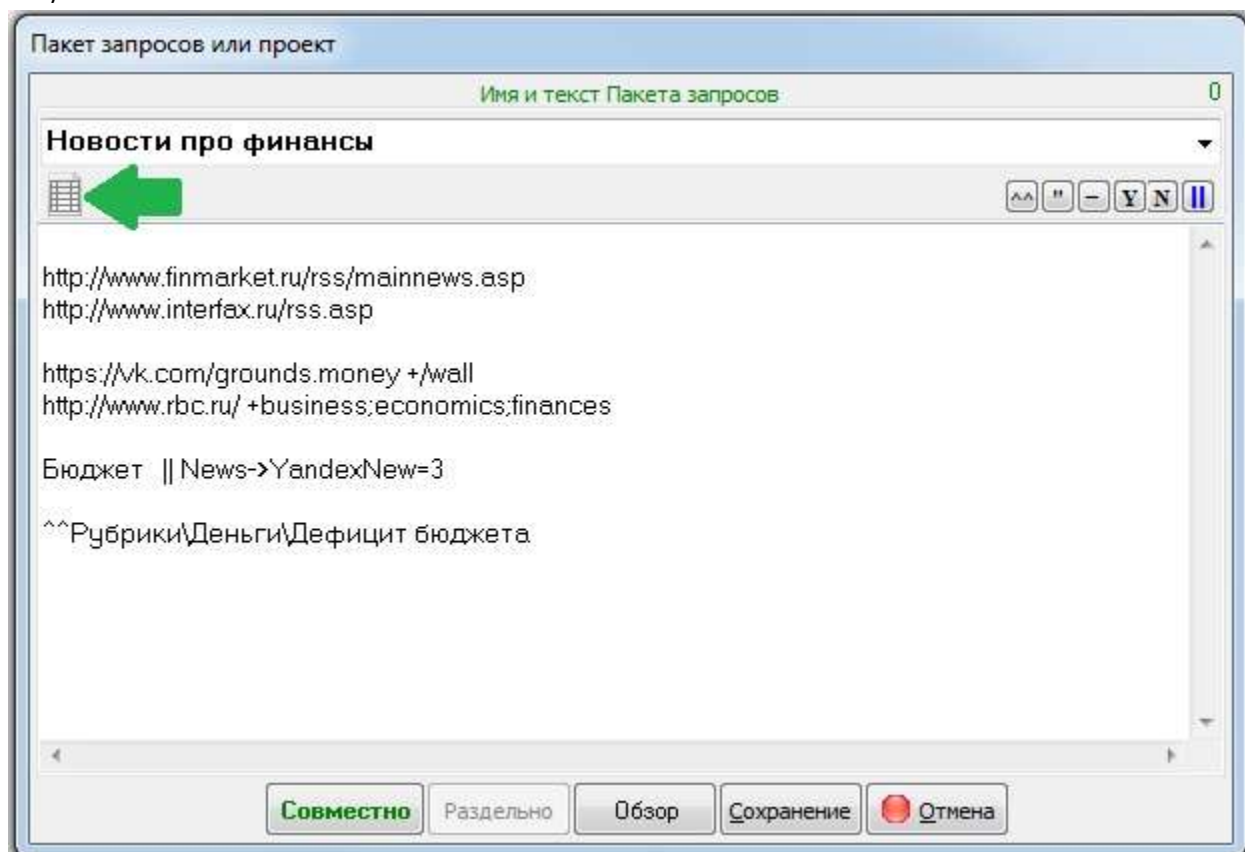
Редактируем Проект следующим образом:



Как видно по картинке, изменения произошли только в последней строке, а именно, мы оставили только Рубрику «Дефицит бюджета». Если бы нам ещё нужны были Источники, где упоминается Сбербанк, то Проект имел бы следующий вид:

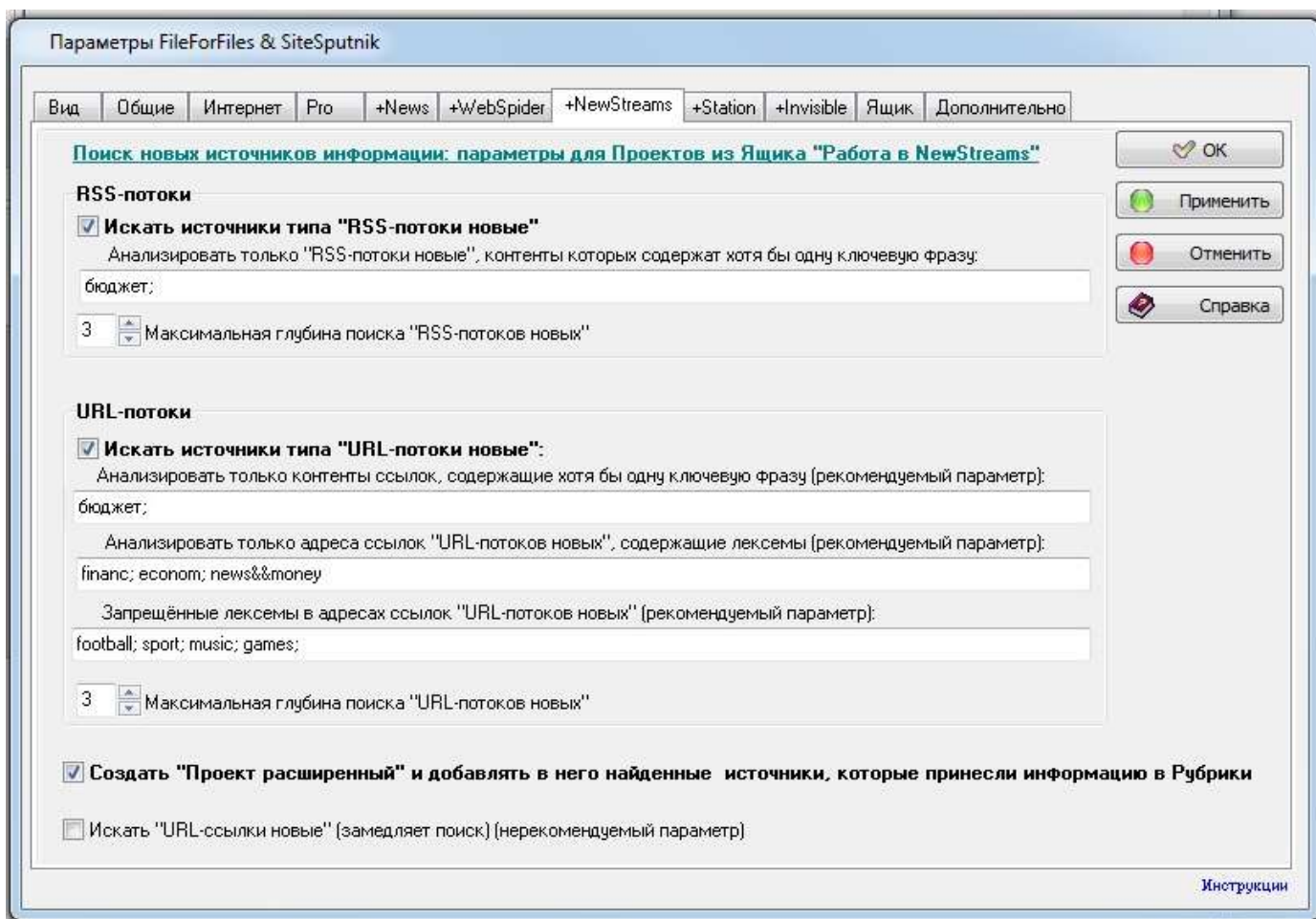


Перейдем к текстовому представлению Проекта. Это совсем не обязательно, но, как выяснится ниже, нагляднее. Для этого, нажмем на иконку, на которую показывает зелёная стрелочка (см. картинку ниже). Получим:



Настройки

- Установим основные Настройки для поиска нужных Источников. Они доступны из главного меню:
 - **Настройки – Параметры – вкладка +NewStreams.**
 - Нужно **внимательно** прочитать о всех параметрах, прописываемых на этой вкладке, если Вы хотите понять принцип поиска Источников информации.



- **Искать источники типа «RSS-потоки новые»:**
 - Все контенты ссылок, которые откроет или скачает СайтСпутник на любом этапе своей работы, будут проанализированы на предмет содержания в них адресов RSS-потоков. Все проверенные на новизну адреса RSS-потоков будут помещены в список «RSS-потоки новые».
- Анализировать только «RSS-потоки новые», контент которых содержит хотя бы одну ключевую фразу:
 - Ключевые фразы или слова записываются через знак «точка с запятой». Хотя бы одна фраза или одно слово должно присутствовать в контенте ссылки из «RSS-потока нового», чтобы она попала в список «RSS-потоки новые+». Это вспомогательный фильтр «грубой очистки» при поиске нужных RSS-потоков. Прошедшие вспомогательный фильтр ссылки считаются кандидатами на принадлежность к искомым нами Источникам и попадают в «Проект проверочный».
 - В нашем конкретном случае (смотрите что прописано в этом параметре на картинке), если на самой ссылке RSS-потока нет слова **бюджет**, то, вряд ли, на ссылках, размещённых в этом потоке, будет идти речь о дефиците бюджета. Если Вы считаете, что это не так, то пропишите свои фильтры или оставьте этот параметр пустым, чтобы проанализировать все новые RSS-потоки.
- Максимальная глубина поиска «RSS-потоков новых»:
 - В нашем случае прописано 3. Это означает, что СайтСпутник сделает три следующие итерации:
 1. Выполнит Проект главный и создаст Проект проверочный.
 2. Выполнит Проект проверочный, полученный на предыдущем шаге итерации, что приведет к двум основным результатам:
 - к добавлению в Проект расширенный из Проекта проверочного, тех Источников, которые принесли информацию в Рубрики.

- к получению очередного Проекта проверочного, содержащего новые Источники-кандидаты.
3. Выполнить то, что написано в предыдущем шаге итерации, только с полученным на нём же новым Проектом проверочным.
- Фильтром «тонкой очистки» являются Правила попадания в Рубрики. В нашем случае это только одна Рубрика «Дефицит бюджета», а её Правило приведено выше.
 - Если очередной Источник-кандидат содержит хотя бы одну ссылку, контент которой удовлетворяет Правилам, то кандидат становится членом, а именно, попадает в Проект расширенный.
- **Искать источники типа «URL-потоки новые»:**
 - Все контентные ссылки, которые откроет или скачает СайтСпутник на любом этапе своей работы, будут проанализированы на предмет содержания в них новых адресов URL-потоков. Новыми считаются те адреса URL-потоков, которые не встречались ранее при выполнении Проекта.
 - Анализировать только контентные ссылки, содержащие хотя бы одну ключевую фразу (рекомендуемый параметр):
 - Ключевые фразы или слова записываются через знак «точка с запятой». Хотя бы одна фраза или одно слово должно присутствовать в контенте новой ссылки, чтобы она была исследована на предмет размещения на ней адресов URL-потоков. Это вспомогательный фильтр «грубой очистки» при поиске нужных URL-потоков. В нашем конкретном случае, если на самой ссылке нет слова **бюджет**, то вряд ли на ссылках, размещённых на ней, будет идти речь о дефиците бюджета. Вы можете прописать другой фильтр, если у Вас на этот счет другое мнение.
 - Анализировать только адреса ссылок «URL-потоки новые», содержащие лексемы (рекомендуемый параметр):
 - Лексемы записываются через знак «точка с запятой». Хотя бы одна из них должна присутствовать в адресе ссылки, чтобы она была исследована на предмет принадлежности к URL-потокам. Это второй вспомогательный фильтр при поиске нужных URL-потоков.
 - В нашем конкретном случае, если в адресе ссылки нет лексем **financ**, **econom** или в одном адресе и **news** и **money**, то, вряд ли, на ссылках, размещённых на этой ссылке, будет идти речь о дефиците бюджета. Только новые ссылки, удовлетворяющие обоим вспомогательным критериям и не содержащие «Запрещённые лексемы» (см. параметр ниже), попадают в список «URL-потоки новые+» и на фильтр «тонкой очистки». Они считаются кандидатами на принадлежность к искомым Источникам и попадают в «Проект проверочный».
 - Запрещённые лексемы в адресах ссылок «URL-потоков новых» (рекомендуемый параметр):
 - Запрещённые лексемы записываются через знак «точка с запятой». Ни одна из них не должна присутствовать в адресе ссылки.
 - В нашем конкретном случае, если в адресе ссылки есть хоть одна лексема из **football**, **sport**, **music** или **games**, то ссылка не анализируется.
 - Максимальная глубина поиска «URL-потоков новых»:
 - Этот параметр Аналогичен по функциям параметру «Максимальная глубина поиска «RSS-потоков новых»». Значения обоих параметров могут отличаться друг от друга.
 - **Создавать «Проект расширенный» и добавлять в него найденные источники, которые принесли информацию в Рубрики:**
 - Установите галочку, чтобы создавать Проект расширенный.
 - Искать «URL-ссылки новые» (замедляет поиск) (нерекомендуемый параметр):
 - В этот список попадают обычные новые ссылки, не являющиеся потоками. В настоящей версии они никак не используются, но на их поиск и проверку на существование затрачиваются вычислительные ресурсы, поэтому не рекомендуется устанавливать галочку у этого параметра. Параметр будет задействован в будущем.

Последовательность действий Пользователя и получаемые результаты

- Заполняем вкладку +NewStreams. Это действие уже сделано выше.
- Подготавливаем Проект главный. Действие сделано в инструкции по [SiteSputnik News](#) и отредактировано выше.
- Выполняем Проект главный. Для этого нажимаем на кнопки: «**L1 – Совместно**».

- Согласно параметрам «Максимальная глубина поиска ...» выполняются три Проекта: 1) Проект главный, 2) Проект проверочный, созданный главным и 3) Проект проверочный, созданный на 2-ом шаге. На картинке ниже на три сделанных шага указывает красная стрелочка, а в серой рамочкой обведен конечный результат – содержание Проекта расширенного на последнем 3-м шаге в текстовом виде.

№	Дата	Время	Вкл.	Запрос, пакет, проект	КЧС	КНС
1	17.09.2015	19:30:	Нет	Новости про финансы	177	
2	17.09.2015	19:34:	Нет	Новости про финансы	962	924
3	17.09.2015	19:36:	Нет	Новости про финансы	298	195

№	Дата	Источник	Файл с результатом поиска
* 21	17.09.	Повторные	Internet Files\Работа в Net
* 22	17.09.	Чистый поток	Internet Files\Работа в Net
* 23	17.09.	Дефицит бюджета	Internet Files\Работа в Net
* 24	17.09.	Статистика по Рубрикам	Internet Files\Работа в Net
* 25	17.09.	График по статистике	Internet Files\Работа в Net
* 26	17.09.	Проект проверочный	Internet Files\Работа в Net
* 27	17.09.	Проект расширенный	Internet Files\Работа в Net


```

^^Проект расширенный: все проверенные источники (RSS и URL-потоки)
^^№ 3 - исполнительный номер проекта
^^Рубрики\Деньги\Дефицит бюджета

^^№ 1:
http://www.rbc.ru/ +business;economics;finances
бюджет || News->YandexNew=3
^^№ 2:
http://vesti.ru/vesti.rss
http://regnum.ru/news/economy/ +financ;econom;news&&money
^^№ 3:
http://vestifinance.ru/yandex.xml
http://argumentiru.com/economics/ +financ;econom;news&&money
http://argumenti.ru/economics/ +financ;econom;news&&money
http://top.rbc.ru/desktop/economics/ +financ;econom;news&&money
http://regnum.ru/analytics/economy.html +financ;econom;news&&money

```

- Проанализируем содержание созданного Проекта расширенного:
 - Символы ^^ и стоящие от них правее символы – это комментарий или специальные операторы.
 - ^^Рубрики\Деньги\Дефицит бюджета – это Рубрики, взятые из главного Проекта.
 - Далее идут Источники, которые принесли информацию на каждом из шагов итерации:
 1. <http://www.rbc.ru/ +business;economics;finances>
бюджет || News->YandexNew=3
 2. <http://vesti.ru/vesti.rss>
<http://regnum.ru/news/economy/ +financ;econom;news&&money>
 3. <http://vestifinance.ru/yandex.xml>
<http://argumentiru.com/economics/ +financ;econom;news&&money>
<http://argumenti.ru/economics/ +financ;econom;news&&money>
<http://top.rbc.ru/desktop/economics/ +financ;econom;news&&money>
<http://regnum.ru/analytics/economy.html +financ;econom;news&&money>

Этот Проект расширенный при желании может быть выполнен. О том, как инициировать выполнение разных видов Проектов, написано ниже.

- Просмотрим содержание Проекта проверочного, полученного на 3 шаге. В него попал один RSS-поток и пять URL-потоков:

```

^^Проект проверочный: очередные Источники - кандидаты на включение в Проект расширенный
^^Рубрики\Деньги\Дефицит бюджета

http://newsler.ru/rss?r=2&k=6

http://vestifinance.ru/ +financ;econom;news&&money
http://vestifinance.ru/articles/ +financ;econom;news&&money
http://vestifinance.ru/videos/ +financ;econom;news&&money
http://finance.comon.ru/ +financ;econom;news&&money
http://top.rbc.ru/mobile/economics/ +financ;econom;news&&money

```

Этот Проект проверочный тоже может быть выполнен. Мы это сделаем ниже.

- Остальную выдачу и выдачу на предыдущих шагах итерации, Вы можете посмотреть, выполнив данный или свой Проект.

Как можно продолжить поиски

Совсем не обязательно изначально задавать большую глубину поиска. Поиск всегда можно продолжить.

- Продолжение нашего поиска, остановившегося на глубине 3. Для этого:

Время	Вкл.	Запрос, пакет, проект	КУС	КНС
2015 19:30	Нет	Новости про финансы	177	
2015 19:34	Нет	Новости про финансы	962	924
2015 19:36	Нет	Новости про финансы	298	195

Источник	Файл с результатом пои
18 17.09. Объединение	Internet Files\Работа в Nei
19 17.09. Ссылка-Рубрики	Internet Files\Работа в Nei
20 17.09. Рубрика-Ссылки	Internet Files\Работа в Nei
21 17.09. Повторные	Internet Files\Работа в Nei
22 17.09. Чистый поток	Internet Files\Работа в Nei
23 17.09. Дефицит бюджета	Internet Files\Работа в Nei
24 17.09. Статистика по Рубрикам	Internet Files\Работа в Nei
25 17.09. График по статистике	Internet Files\Работа в Nei
26 17.09. Проект проверочный	Internet Files\Работа в Nei
27 17.09. Проект расширенный	Internet Files\Работа в Nei


```


^^Проект проверочный: очередные Источники (RSS и URL-потоки новы
^^Рубрики\Деньги\Дефицит бюджета

http://newsler.ru/rss?r=2&k=6

http://vestifinance.ru/ +financ;econom;news&&money
http://vestifinance.ru/articles/ +financ;econom;news&&money
http://vestifinance.ru/videos/ +financ;econom;news&&money
http://finance.comon.ru/ +financ;econom;news&&money
http://top.rbc.ru/mobile/economics/ +financ;econom;news&&money

```

мы сначала становимся на Проект проверочный, затем в подменю иконки  выбираем пункт «Выполнить Проект проверочный». Это обеспечит поиск на глубину 6.

- Для сведения: можно выполнять Проекты и из меню на правой кнопке мышки.
- Аналогично можно выполнить и Проект расширенный. В Источниках, которые уже приносили информацию в Рубрики, появятся новые ссылки, которые могут содержать новые Источники.
- Всегда можно в очередной раз выполнить Проект главный. Он найдет новые ссылки и, соответственно, могут быть найдены новые Источники. Выполнить Проект главный можно двумя способами:
 - или «L1 – Совместно».
 - или  - иконка «Глобус с линзой».
- Назначение Проекта расширенного Проектом главным: становимся на Проект расширенный, на правой кнопке мышки выбираем пункт «Выполнить, открыть, обновить – Сохранить как Проект главный».
- Очередное выполнение Проекта проверочного, полученного на 3 шаге итерации, на глубину 4 принесло результаты:

№	Дата	Время	Вкл.	Запрос, пакет, проект	КУС	КНС
1	17.09.2015	19:30	Нет	Новости про финансы	177	
2	17.09.2015	19:34	Нет	Новости про финансы	962	924
3	17.09.2015	19:36	Нет	Новости про финансы	298	195
4	18.09.2015	10:05	Нет	Новости про финансы	180	141
5	18.09.2015	10:08	Нет	Новости про финансы	241	147
6	18.09.2015	10:12	Нет	Новости про финансы	404	364
7	18.09.2015	10:17	Нет	Новости про финансы	393	310

№№	Дата	Источник	Файл с результатом пои
5	18.09.	RSS-потоки все	Internet Files\Работа в Nei
6	18.09.	RSS-потоки новые	Internet Files\Работа в Nei
7	18.09.	RSS-потоки новые+	Internet Files\Работа в Nei
8	18.09.	URL-неоткрывшиеся	Internet Files\Работа в Nei
9	18.09.	URL-ссылки все	Internet Files\Работа в Nei
10	18.09.	URL-потоки новые+	Internet Files\Работа в Nei
11	18.09.	Открытые	Internet Files\Работа в Nei
12	18.09.	Новые	Internet Files\Работа в Nei
13	18.09.	Заголовки	Internet Files\Работа в Nei
14	18.09.	Статистика текущая	Internet Files\Работа в Nei
15	18.09.	Статистика обобщенная	Internet Files\Работа в Nei
16	18.09.	Статистика полная	Internet Files\Работа в Nei
17	18.09.	Все сниппеты	Internet Files\Работа в Nei
18	18.09.	Объединение	Internet Files\Работа в Nei
19	18.09.	Ссылка-Рубрики	Internet Files\Работа в Nei
20	18.09.	Рубрика-Ссылки	Internet Files\Работа в Nei
21	18.09.	Повторные	Internet Files\Работа в Nei
22	18.09.	Чистый поток	Internet Files\Работа в Nei
23	18.09.	Дефицит бюджета	Internet Files\Работа в Nei
24	18.09.	Статистика по Рубрикам	Internet Files\Работа в Nei
25	18.09.	График по статистике	Internet Files\Работа в Nei
26	18.09.	Проект проверочный	Internet Files\Работа в Nei
27	18.09.	Проект расширенный	Internet Files\Работа в Nei

```

^^Проект расширенный: все проверенные источники (RSS и URL-потоки
^^№ 7 - исполнительный номер проекта
^^Рубрики\Деньги\Дефицит бюджета

^^№ 1:
http://www.rbc.ru/ +business;economics;finances
бюджет || News->YandexNew=3
^^№ 2:
http://vesti.ru/vesti.rss
http://regnum.ru/news/economy/ +financ;econom;news&&money
^^№ 3:
http://vestifinance.ru/yandex.xml
http://argumentiru.com/economics/ +financ;econom;news&&money
http://argumenti.ru/economics/ +financ;econom;news&&money
http://top.rbc.ru/desktop/economics/ +financ;econom;news&&money
http://regnum.ru/analytics/economy.html +financ;econom;news&&money
^^№ 4:
http://vestifinance.ru/ +financ;econom;news&&money
^^№ 6:
http://newsler.ru/economics?p=6 +financ;econom;news&&money
http://ria.ru/economy/ +financ;econom;news&&money
http://ng.ru/economics/ +financ;econom;news&&money
http://vz.ru/economy/ +financ;econom;news&&money
http://mk.ru/economics/ +financ;econom;news&&money
^^№ 7:
http://feeds.newsru.com/com/www/section/finance
https://vedomosti.ru/rss/rubric/economics
https://vedomosti.ru/rss/news/economics
http://vedomosti.ru/rubrics/economics +financ;econom;news&&money

```


На картинке показан Проект расширенный в текстовом виде. А Проект проверочный, сформированный на шаге 7, имеет следующий текстовый вид:

^^Проект проверочный: очередные Источники - кандидаты на включение в Проект расширенный
^^Рубрики\Деньги\Дефицит бюджета

<http://feeds.newsru.com/com/www/dossier/659>
<http://feeds.newsru.com/com/www/dossier/423>
<https://vedomosti.ru/rss/rubric/economics/regulations>
<https://vedomosti.ru/rss/rubric/economics/macro>

<http://m.vedomosti.ru/rubrics/economics+financ;econom;news&&money>
<http://newizv.ru/economics/+financ;econom;news&&money>

В него попало четыре RSS-потока и два URL-потока. Поиск может быть продолжен на любую глубину.

- К сведению. Если на вкладке +Pro включен параметр «Создавать список «Новые» ссылки относительно всех поисков в Ящике», то Проект расширенный будет создаваться не только для одноименных, но и разноименных Проектов – один для всех.

Естественный интеллект

- Автоматика автоматикой, но естественный интеллект никто не отменял.
- Просматривайте вручную списки «RSS-потоки новые», «RSS-потоки новые+», «URL-потоки новые+» и другую выдачу. Открывайте, например, в браузере содержание новых потоков, попавших в эти списки. Оценивайте их своим умом.
- Такая «ручная» работа не порок! СайтСпутник, чтобы подготовить информацию для размышления, размещённую в своей выдаче, уже целенаправленно «перерыл» **html-контенты** тысяч ссылок на разной глубине веб-пространства, проанализировал сотни потоков информации, и нашёл **новые** Источники, которые Вы ранее не встречали. Возможно, что Вы включите в Проект «забракованные» СайтСпутником Источники, или, наоборот, удалите Источники, «понравившиеся» ему. На практике вручную не реально сделать такую работу.
- Интересные Потоки информации могут встретиться на любом шаге итерации.

Особое предложение

- Если у Вас есть предложения по «оснастке», которую можно «повесить» на рассмотренный в данной публикации вычислительный процесс, то сделайте это.

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/>