

ООО «ЭЛТЕК»

Руководство по установке и эксплуатации
программного обеспечения
«ЛАН.Студия управления источниками»

2024

1 Общие сведения

Программное обеспечение «ЛАН.Студия управления источниками» предназначено для настройки информационных источников в сети Интернет. Каждый источник представляет собой группы настроек, которые регламентируют выполнение различных процессов.

Набор модулей формируется индивидуально в зависимости от назначения программного компонента и может отличаться от примеров, указанных в данном руководстве. Далее приведено описание отдельных модулей, которые могут быть включены в программное обеспечение «ЛАН.Студия управления источниками».

Графический интерфейс программного обеспечения «ЛАН.Студия управления источниками» представлен на рисунке 1 и состоит из вкладок:

- «Выделение текста» – вкладка, предназначенная для просмотра и добавления главных страниц, выбора режима и перехода к правилам сбора источников;

- «Общие параметры» – вкладка, предназначенная для просмотра и изменения настроек процессов;

- «Описание» – в данной вкладке расположены прописываемые вручную параметры, описывающие выбранный источник;

- «XML-представление» – отображение источника в виде XML конфигурации;

- «Результаты тестирования» – результаты работы правил сбора после тестирования источника;

- «Рубрики» – вкладка, предназначенная для классификации источников по тематическим классификаторам;

- «Привязки» – связывание правил и главных страниц, для оптимизации процессов сбора с источника.

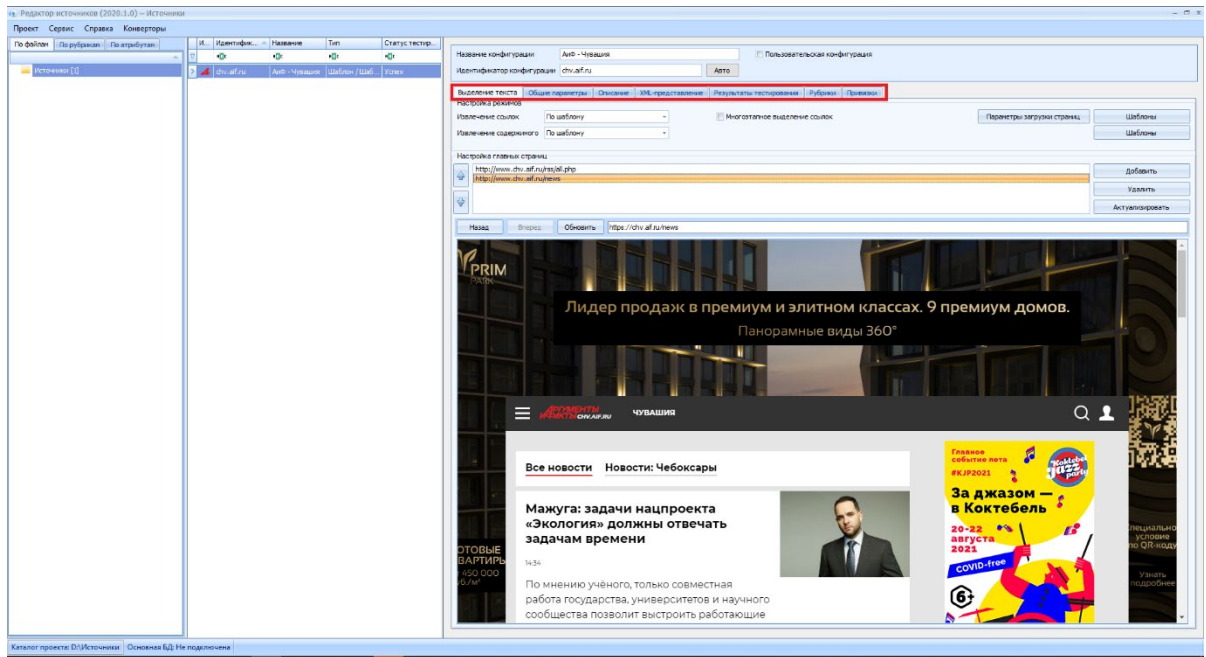


Рисунок 1 – Графический интерфейс программного обеспечения
«ЛАН.Студия управления источниками»

2 Установка

Для установки программного обеспечения «ЛАН.Студия управления источниками» требуется запустить установочный файл.

В результате запуска установочного файла откроется мастер установки программного обеспечения, представленный на рисунке 2.

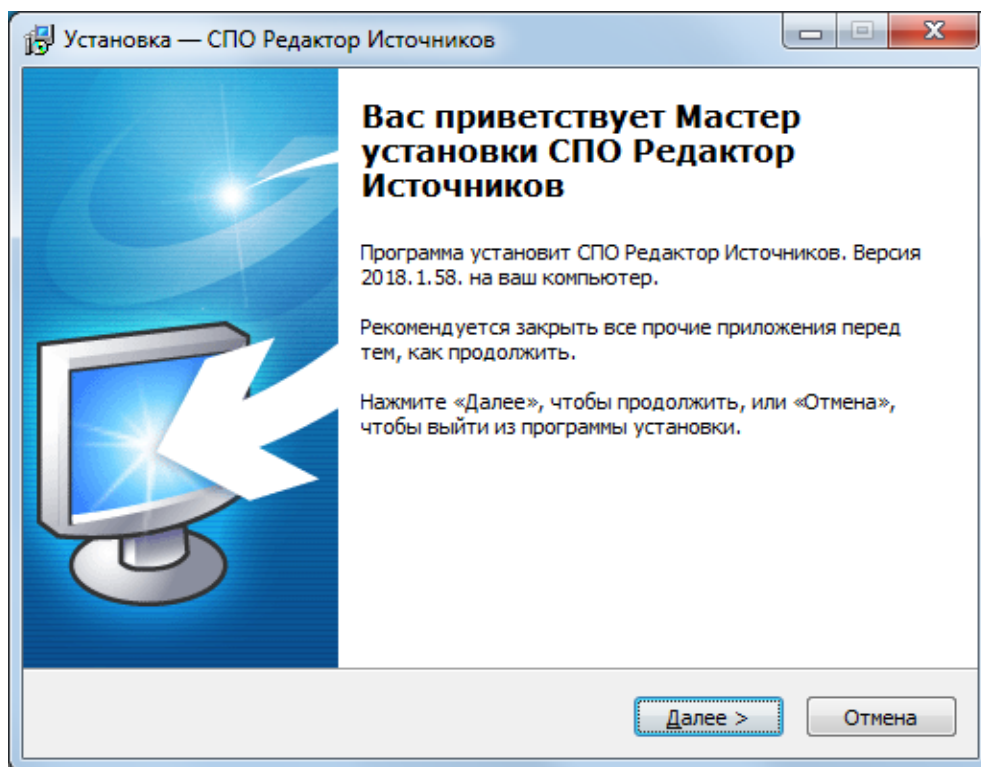


Рисунок 2 – Мастер установки «ЛАН.Студия управления источниками»

Следуя инструкциям данного мастера, необходимо выбрать директорию установки программного обеспечения (по умолчанию «C:\Program Files\Lan\WebCrawlerConfig») и нажать на расположенную в нижней части окна мастера кнопку «Далее», как показано на рисунке 3.

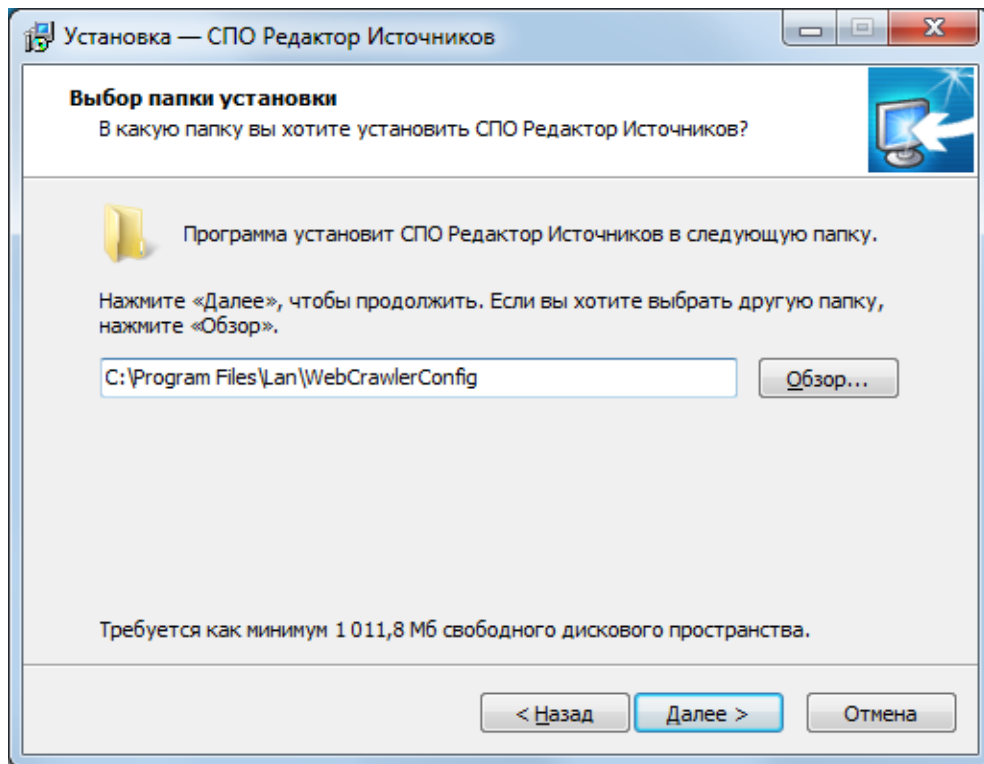


Рисунок 3 – Выбор директории, в которую будет установлена «ЛАН.Студия управления источниками»

Далее требуется указать необходимость установки программного обеспечения «Служба интернет аккаунтов». Если служба не была установлена ранее, необходимо установить метку, как указано на рисунке 4.

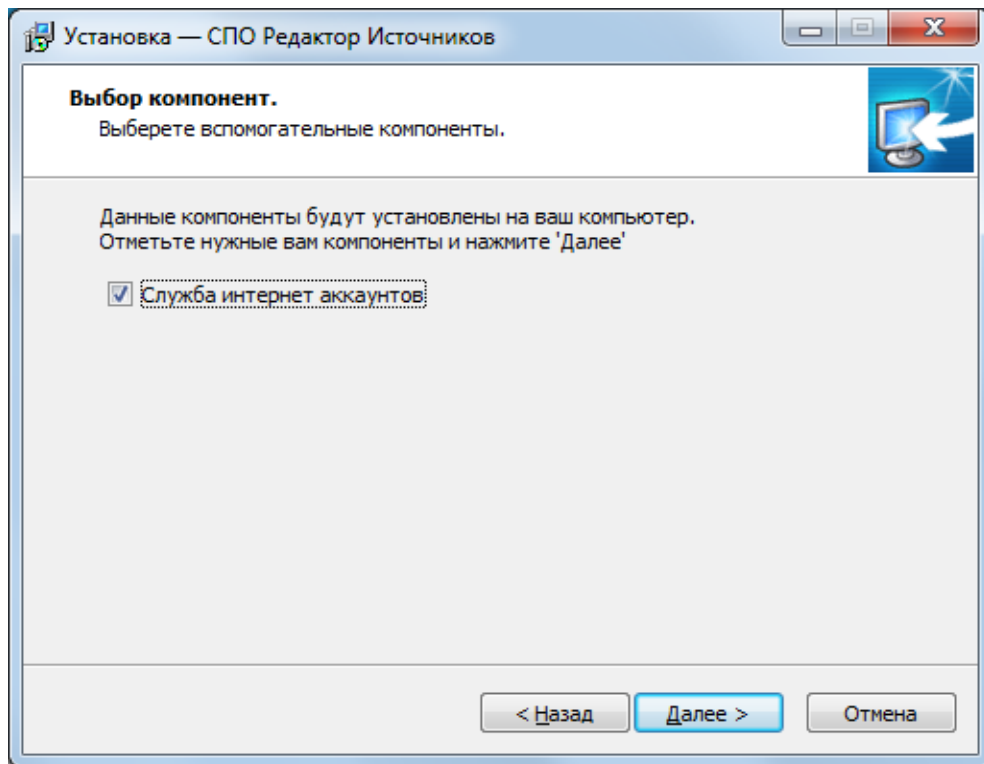


Рисунок 4 – Выбор дополнительных компонентов

Следующее окно мастера установки, представленное на рисунке 5, позволяет выбрать тип установки. В том случае, если программное обеспечение никогда не было установлено ранее, необходимо выбрать пункт «Полная установка». Соответственно, если мастер установки был запущен с целью обновления версии программного обеспечения до более новой, необходимо осуществить выбор пункта «Обновление установленной программы».

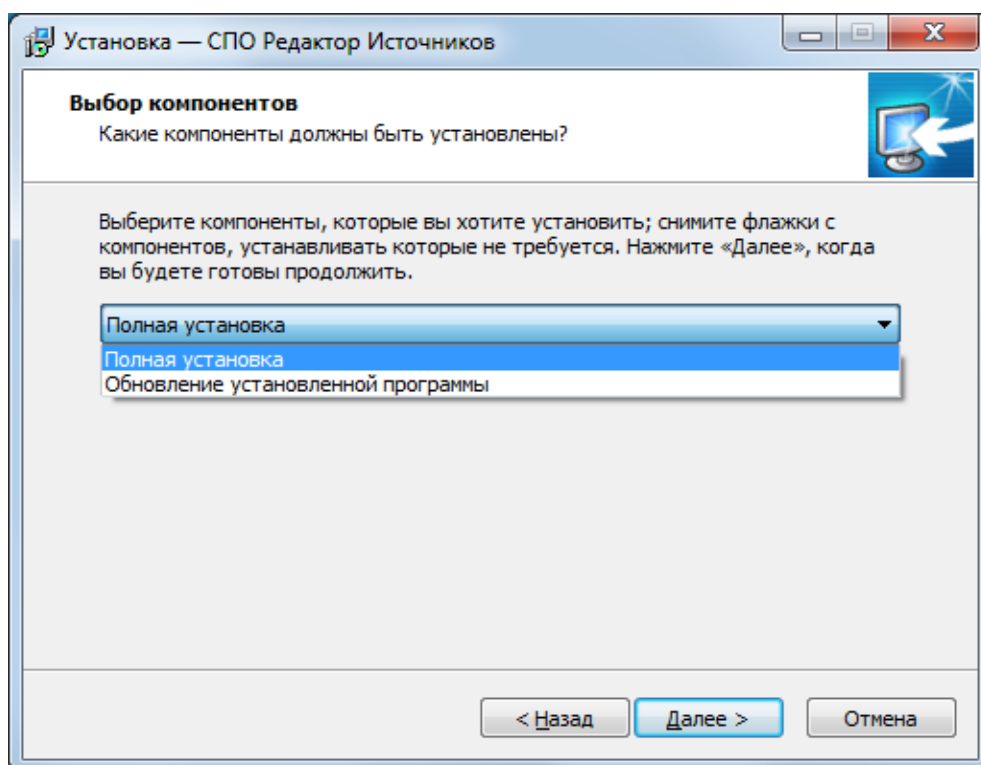


Рисунок 5 – Выбор типа установки

Далее необходимо указать название раздела меню «Пуск», в котором мастер создаст ярлык для запуска программы (по умолчанию – «СПО Редактор Источников»), и нажать на кнопку «Далее», как показано на рисунке 6.

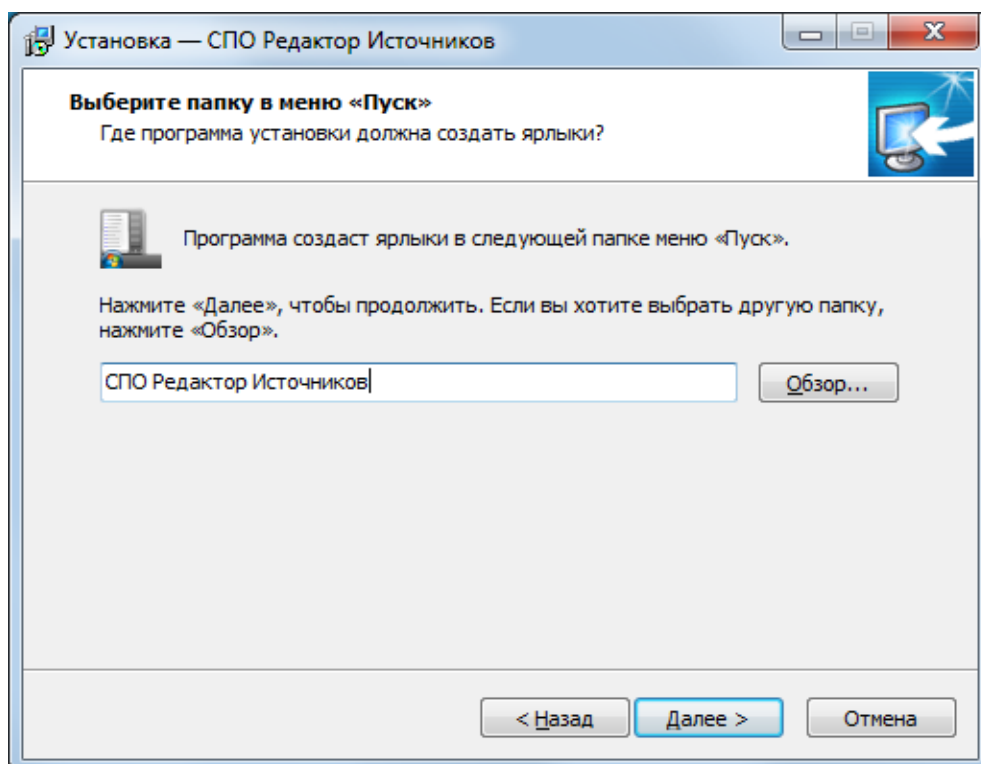


Рисунок 6 – Выбор названия раздела меню «Пуск»

В следующем окне, представленном на рисунке 7, необходимо установить метку, если требуется создавать ярлык для программного обеспечения на рабочем столе, и нажать на кнопку «Далее».

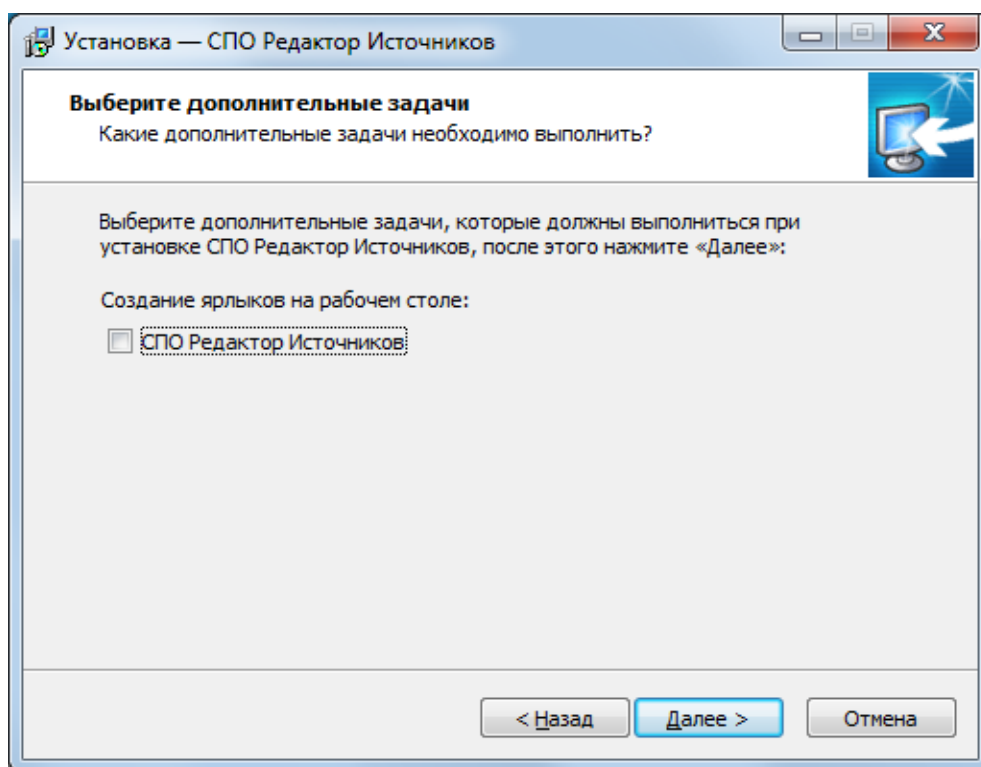


Рисунок 7 – Создание ярлыка на рабочем стол

После указания всех параметров, которые необходимы мастеру, следует запустить процесс установки с помощью кнопки «Установить», представленной на рисунке 8.

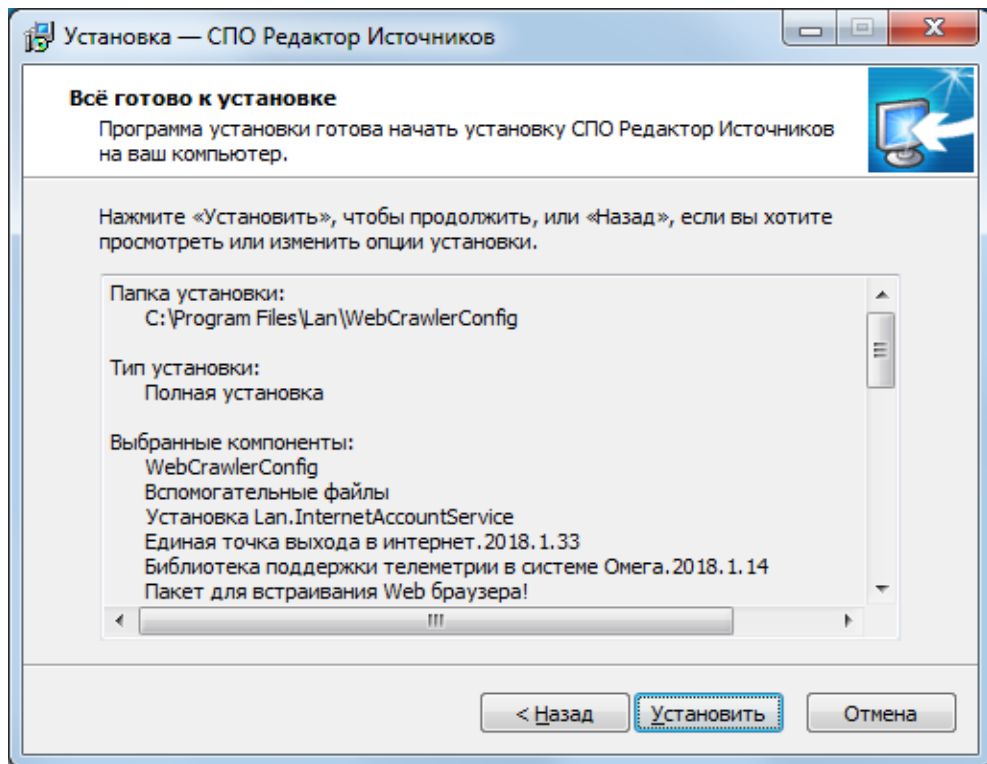


Рисунок 8 – Запуск процесса установки

В результате выполнения данного действия появится окно, в котором отображается процесс установки, представленное на рисунке 9.

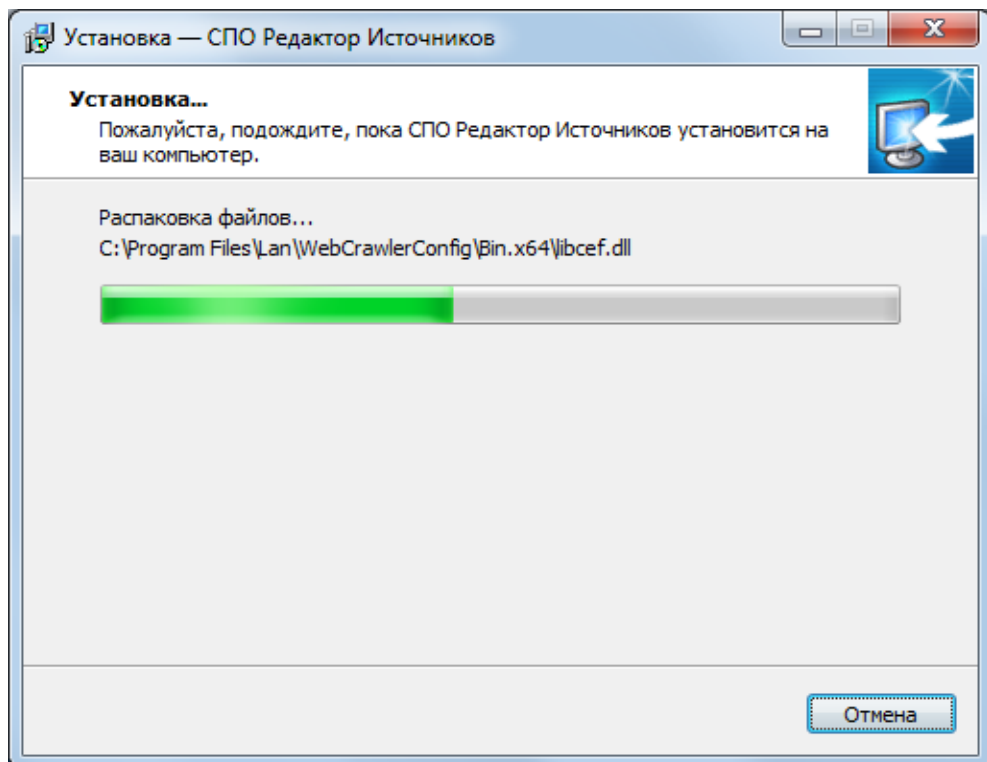


Рисунок 9 – Прогресс процесса установки

По завершении процесса установки ПО, если требовалось установить компонент «Служба интернет аккаунтов» - запустится новый мастер установки.

В случае успешной инсталляции программного обеспечения появится окно, представленное на рисунке 10, в котором необходимо нажать на кнопку «Завершить» для закрытия мастера установки.

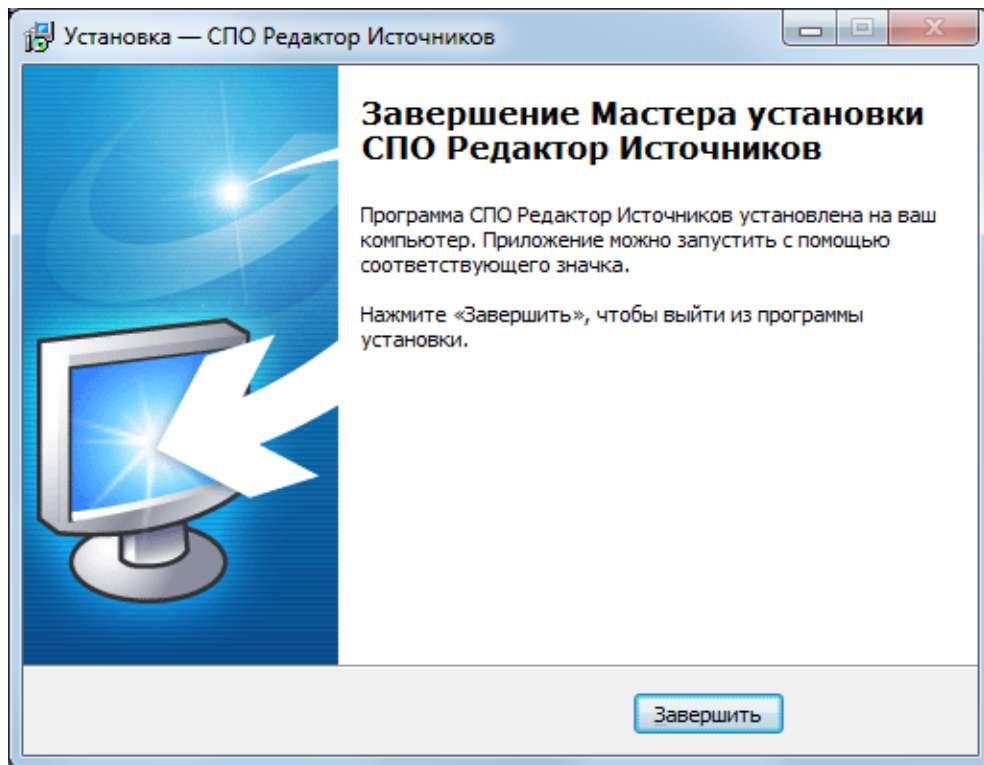


Рисунок 10 – Завершение установки

Результатом установки программного обеспечения будет его отображение в панели меню «Пуск», представленное на рисунке 11.

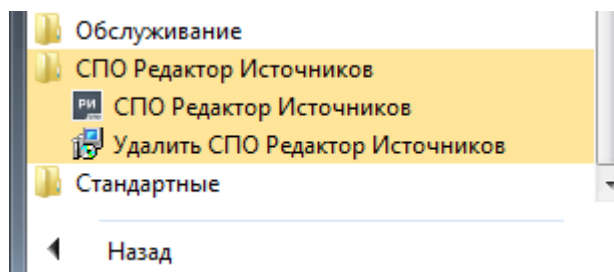


Рисунок 11 – Результат установки ПО

3 Эксплуатация

3.1 Запуск

«ЛАН.Студия управления источниками» запускается с помощью ярлыка на рабочем столе «СПО Редактор Источников». После запуска откроется окно, представленное на рисунке 12.

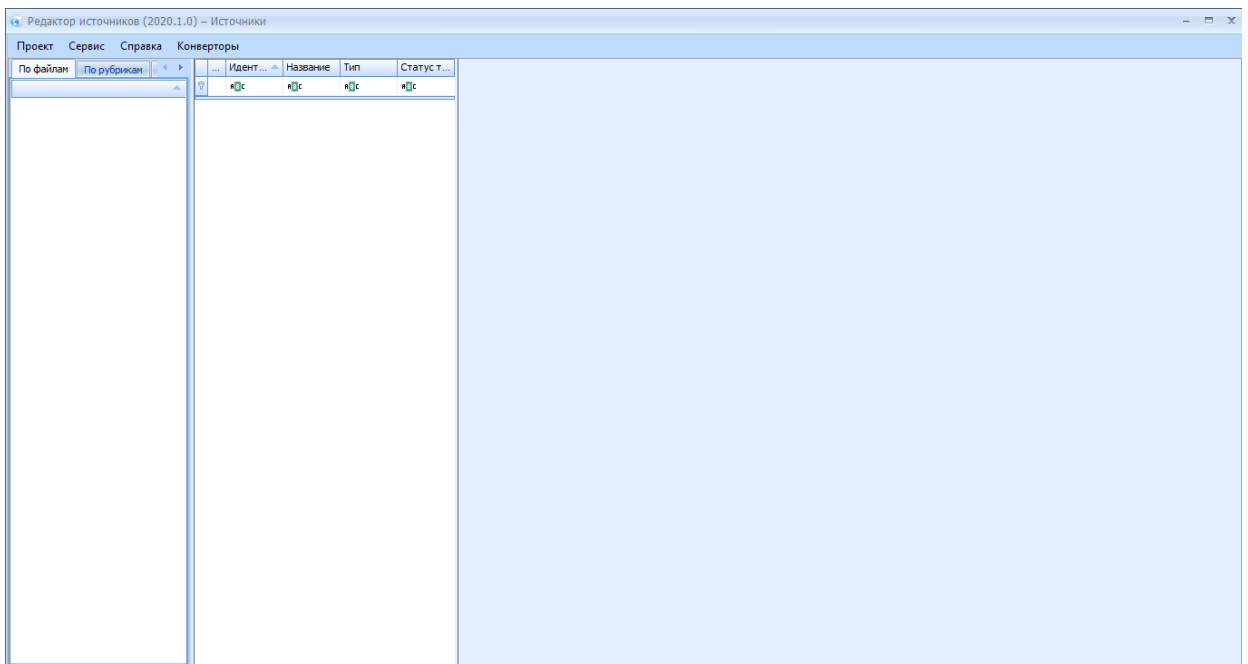


Рисунок 12 – Главная страница приложения

Для начала работы в приложении необходимо создать проект как показано на рисунке 13.

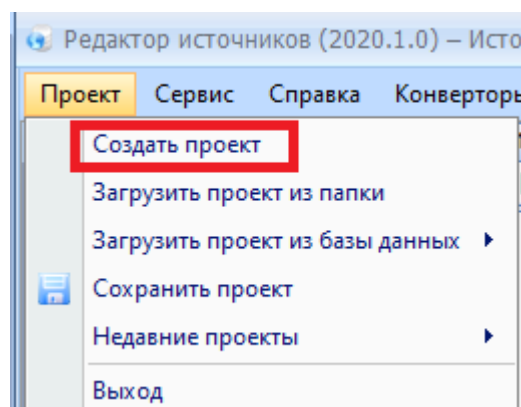


Рисунок 13 – Создание нового проекта

Далее нужно создать источник для настройки правил сбора контента.

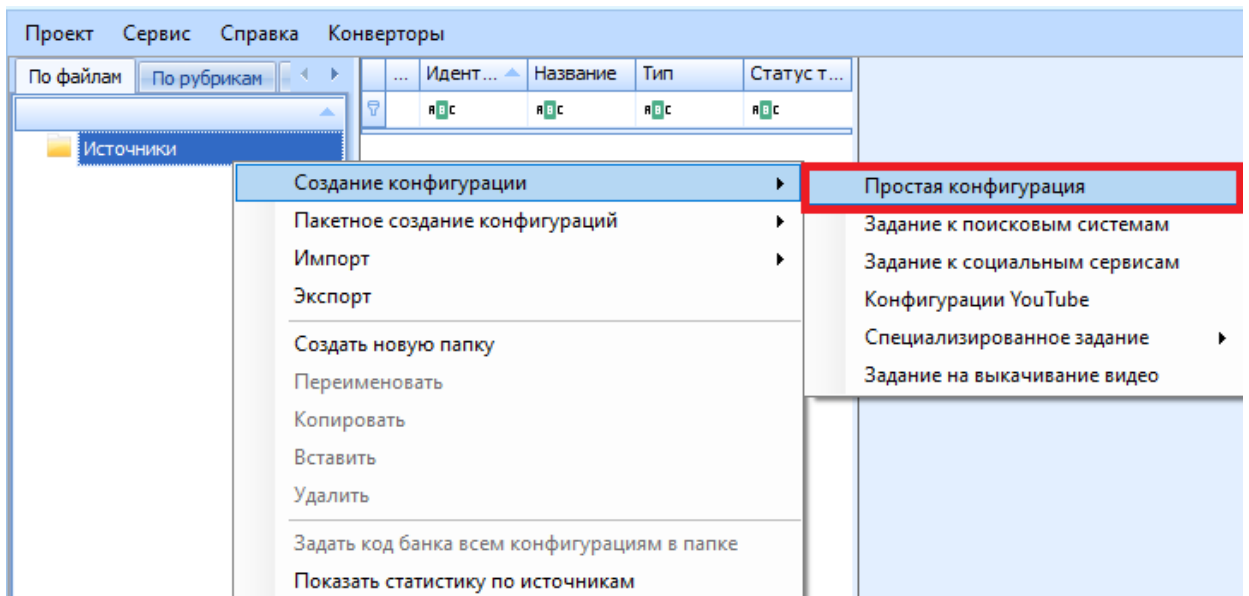


Рисунок 14 – Создание источника

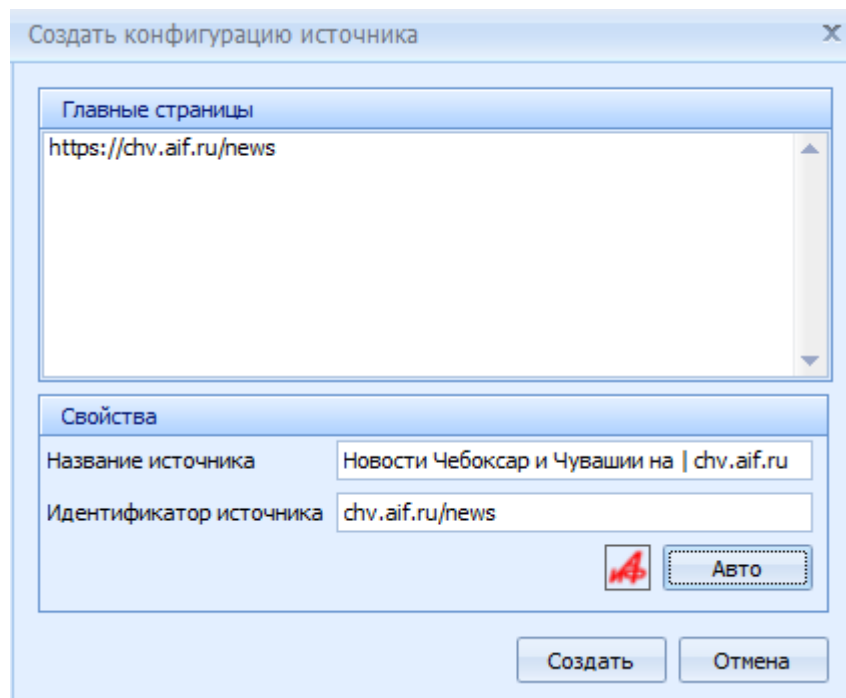


Рисунок 15 – Создание источника

Для проверки результатов работы необходимо дождаться обработки и открыть вкладку «Результаты тестирования». В данной вкладки должны содержаться результат работы правил сбора информации с информационного источника.

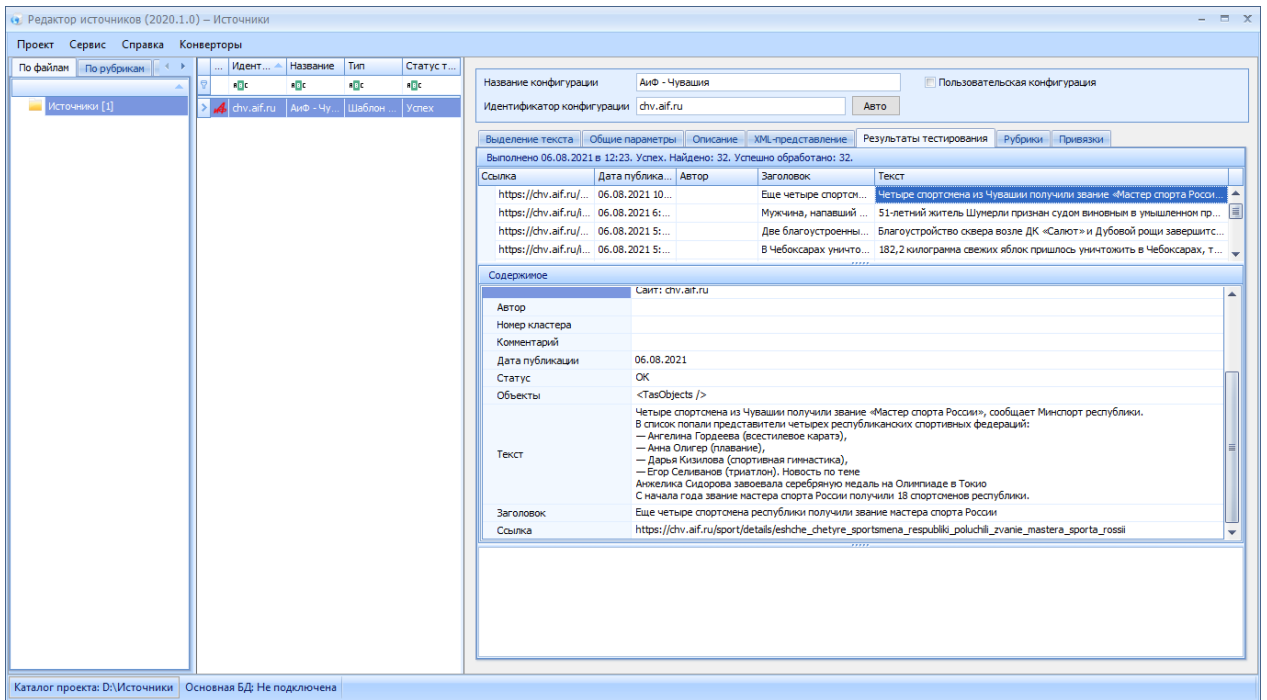


Рисунок 16 – Результат тестирования источника

3.2 Вкладка «Выделение текста»

«Выделение текста» представлена на рисунке 17.

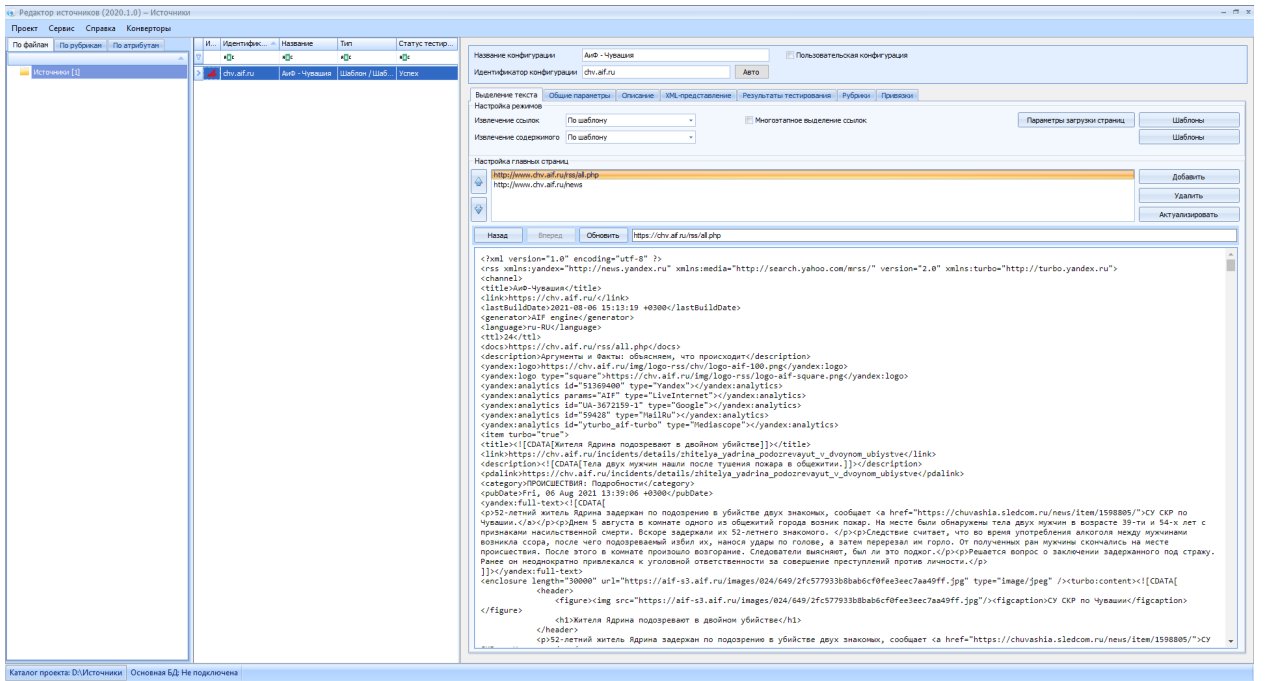


Рисунок 17 – Вкладка «Выделение текста»

По умолчанию, для источников задается режим обхода «Авто». В этом случае программный компонент загрузки документов будет пытаться самостоятельно выделить ссылки на новости, расположенные на странице, указанной при создании, затем переходить по ним и пытаться выделить там текст новости, дату публикации и заголовок. Данный способ не всегда дает удовлетворительные результаты. Это обусловлено фактическим отсутствием каких-либо единых стандартов верстки, применяемых при создании страниц различных сайтов. Для более точной обработки источников рекомендуется использовать режим «По шаблону», где необходимо самостоятельно создать правила обхода для выделения ссылок, заголовков, дат, текстов и авторов сообщений. Переключение режимов производится отдельно для страницы со ссылками на новости и для страниц с самими новостями.

Для первоначального создания нового правила обхода источника необходимо переключить режим обхода с «Авто» на «По шаблону» и нажать на кнопку «Шаблоны», после чего откроется окно «Редактора шаблонов», а главное окно программы заблокируется до закрытия данного окна.

Если требуется получить информацию с сайта, структура страниц которого представляет «дерево», необходимо установить метку «Многоэтапное выделение ссылок». Данный параметр изменяет режим обработки шаблонов извлечения ссылок таким образом, чтобы каждый шаблон выделял ссылки для своего уровня вложенности: первый шаблон – для корня сайта, второй шаблон – для второго уровня сайта, третий – для третьего и так далее до конца списка шаблонов.

Как правило, многоэтапный режим выделения ссылок необходим для обработки сайтов-форумов.

На вкладке «Выделение текста» также имеется возможность редактировать список главных страниц источника, подлежащих обходу. Для добавления новой страницы необходимо зайти на нее во встроенном браузере и нажать на кнопку «Добавить» напротив списка главных страниц. Следует отметить, что если просто вставить желаемую ссылку в адресную строку

браузера и сразу нажать на данную кнопку, то добавится не желаемая ссылка, а ссылка на ту страницу, которая в данный момент открыта в браузере. Поэтому после копирования ссылки необходимо сначала нажать на клавишу «Enter» и только затем добавлять желаемую ссылку в список.

Кнопка «Удалить» удаляет ту страницу, которая выделена курсором в списке уже добавленных главных страниц.

Кнопка «Актуализировать» выполняет поочередный переход на каждую из главных страниц и, в случае, если страница недоступна или перемещена, отображает соответствующее информационное окно с возможностью удалить или изменить главную страницу.

Кнопка «Параметры загрузки страниц» выполняет функцию перехода к окну настроек загрузки страницы таких как: правило ожидания элемента, настройки скриптов, настройки скриптов авторизации, настройки обхода капчи.

3.3 Вкладка «Общие параметры»

На данной вкладке, представленной на рисунке 18, расположены настраиваемые параметры, используемые при обходе сайта программой «ЛАН.Интернет-мониторинг».

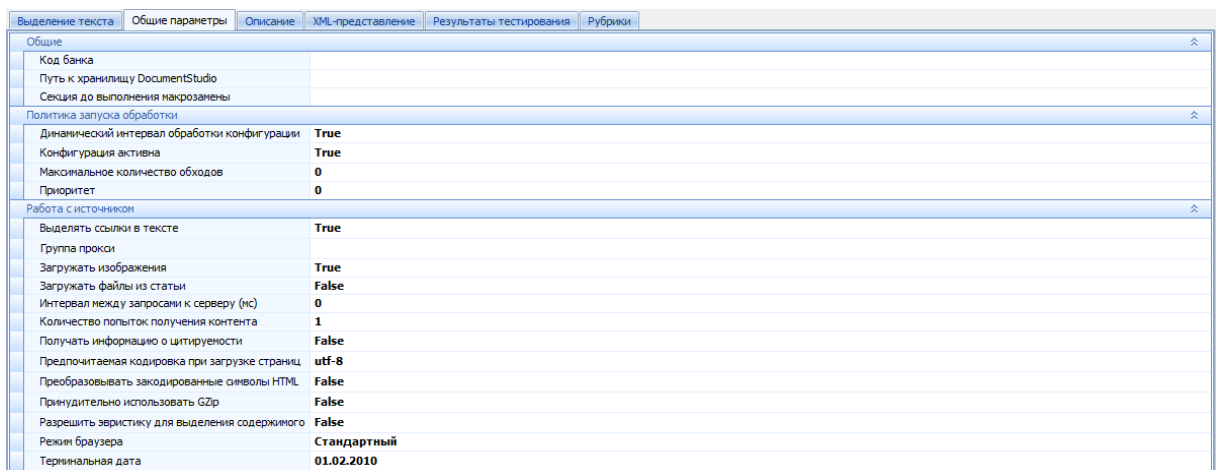


Рисунок 18 – Вкладка «Общие параметры»

3.4 Вкладка «Описание»

На данной вкладке, представленной на рисунке 19, расположены прописываемые вручную параметры, описывающие выбранный источник.

При нажатии на кнопку «Получить изображение» будет выполнена попытка получения текущего логотипа сайта.

При нажатии на кнопку «Получить описание источника» будет произведена «выкачка» информации о нем со специализированного сайта «Whois».


Выделение текста	Общие параметры	Описание	XML-представление	Результаты тестирования	Рубрики
Основные данные источника					
Полное название источника	3DNews.ru				
Название источника	3dnews.ru				
Тип источника	Новостные сайты				
Название организации	3DNews Daily Digital Digest				
Контактные данные					
Статус	3				
Данные о домене источника					
IP-адрес	195.90.131.231				
Дата регистрации домена					
Регистратор					
Адрес расположения домена					
ФИО администратора					
Контактные данные администратора					
Домашний адрес администратора					
Остальная информация об источнике					
Адрес редакции	101000, Moscow, Russia, 108, Profsoyuznaya Street, Rosnet (Russian Data Communication Company);				
Дата регистрации редакции					
Надежность					
Номер свидетельства					
Периодичность	Ежедневно				
Территория распространения					
Учредители					
Главный редактор					
Языки	Русский				
Описание	3DNews Daily Digital Digest – первое независимое российское онлайн-издание, посвященное цифровым технологиям. Основано в 1997 г. Цель проекта – своевременная публикация максимально объективной информации обо всем, происходящем на рынке IT, а также помощь пользователям цифровых устройств в выборе, приобретении и наиболее эффективном использовании оборудования и программного обеспечения. Ежедневно 3DNews.ru посещают более 190 тыс. человек из всех стран СНГ и русскоязычного зарубежья. Более 3 млн посетителей ежемесячно просматривают около 18 млн страниц. Число подписчиков, ежедневно читающих наши рассылки на Subscribe.ru и на Mailist.ru , составляет более 90 000, и этот показатель постоянно растет. Новостные ленты 3DNews транслируются на многие сайты Всемирной сети, среди которых Яндекс.Новости, FeedBurner, Inforotor, микроблоги Twitter, LiveJournal, Friendfeed и др.				
Картинка					
<input type="button" value="Получить изображение"/>					
Данные Whois:					
Сайт: www.3dnews.ru ;					
Регистрация: ;					
IP-адреса:					
195.90.131.231;					
Найденные домены по данному IP-адресу:					
Объекты по домену www.3dnews.ru :					
InetNim: 195.90.131.0 - 195.90.131.255;					
<input type="button" value="Получить описание источника"/>					

Рисунок 19 – Вкладка «Описание»

3.5 Вкладка «Рубрики»

На вкладке «Рубрики», представленной на рисунке 20, расположены все классификаторы, загруженные в Систему.

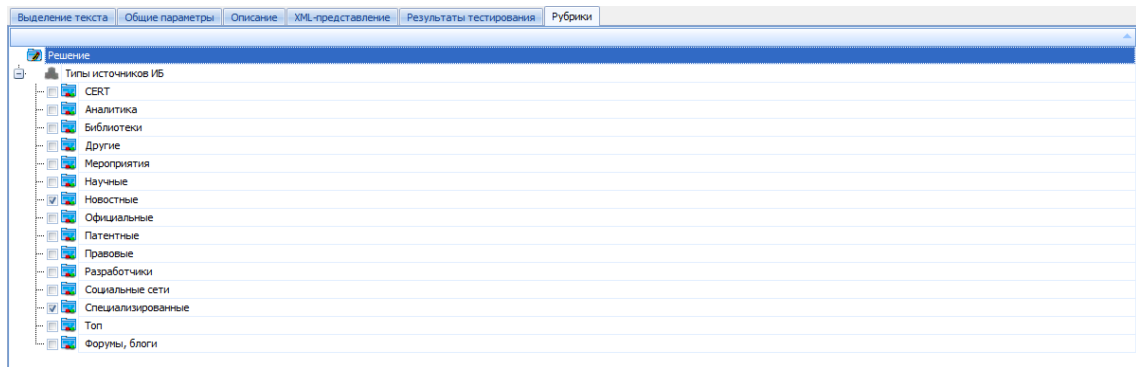


Рисунок 20 – Вкладка «Рубрики»

В данной вкладке можно выбрать желаемые рубрики классификаторов, которые будут автоматически соотнесены со всеми документами, скачанными из выбранного источника, независимо от фактического содержания текста.

3.6 Вкладка «XML-представление»

На данной вкладке, представленной на рисунке 21, расположен исходный код конфигурационного файла (источника). В программе показывается версия, находящаяся в оперативной памяти, которая будет сохранена при нажатии на кнопку «Сохранить проект», и ее вид может отличаться от версии, которая будет видна при одновременном открытии этого конфигурационного файла в текстовом редакторе до выполнения сохранения проекта.

Предусмотрена возможность выделить несколько файлов и запустить тестирование сразу всего списка. Для этого необходимо воспользоваться стандартными командами среды Windows («Ctrl+A» - чтобы выделить все, «Ctrl+клик» - чтобы выделить желаемые позиции и «Shift+клик» - чтобы выделить промежутки).

После успешного завершения тестирования на вкладке «Результаты тестирования», представленной на рисунке 23, станет доступна подробная информация по проведенному тестированию. Стоит заметить, что после закрытия программы вместе с проектом сохраняется только статус тестирования, но не результаты тестирования.

The screenshot shows a software window with several tabs: 'Выделение текста', 'Общие параметры', 'Описание', 'XML-представление', 'Результаты тестирования', and 'Рубрики'. The 'Результаты тестирования' tab is active, displaying a table of search results.

Ссылка	Дата публикации	Автор	Заголовок	Текст
https://3dnews.ru/972660	16.07.2018 12:51:04		Sony выпустила улучшенную в...	Компания Sony, по сообщениям сетевых источников, выпустила улучшенную версию компактной фото...
https://3dnews.ru/972666	16.07.2018 13:32:59		«ВымпелКом» стал экологичен...	«ВымпелКом», оказывающий услуги связи под маркой «Билайн», и финская компания HMD Global, выпу...
https://3dnews.ru/972639	16.07.2018 14:04:24		Складной планшет Microsoft An...	Пару лет назад впервые прозвучало кодовое имя Andromeda, за которым якобы скрывается проект M...
https://3dnews.ru/972654	16.07.2018 11:41:24		Материнские платы на Intel Z37...	Большинство крупных производителей материнских плат начали выпускать новые версии BIOS для св...
https://3dnews.ru/972656	16.07.2018 11:10:57		AMD Ryzen 5 2500X: новые под...	Уже довольно давно стало известно, что компания AMD готовит пополнение семейства процессоров R...

Below the table, the 'Содержимое' (Content) section is expanded for the first result. It shows the title 'Sony выпустила улучшенную версию фотоаппарата RX100 V', the URL 'https://3dnews.ru/972660', the date '16.07.2018 12:51', and the status 'OK'. The 'Текст' (Text) section contains a detailed description of the camera's features, including the 20.1-MP CMOS sensor, 1" Exmor RS sensor, 0.39" OLED display, and Wi-Fi/NFC connectivity. The 'Объекты' (Objects) section shows XML metadata for the article.

At the bottom of the screenshot, there is a photograph of the Sony RX100 V camera, a compact digital camera with a black body and a lens.

Рисунок 23 – Вкладка «Результаты тестирования»

Данная вкладка используется при поиске ошибок выделения даты новости, сравнения количества найденных новостей с их фактическим количеством на главной странице, для выяснения правильности нахождения

конца текста новости и выявления возможных проблем с кодировкой текста новости.

На вкладке «Результаты тестирования» только после успешного выполнения тестирования выводятся результаты выделения, как если бы программный компонент «ЛАН.Интернет-мониторинг» скачивал документы с сайта с использованием текущего конфигурационного файла. Причем тестирование показывает актуальные результаты как в режиме обхода «Авто», так и в режиме «По шаблону».

3.8 Вкладка «Привязки»

Вкладка «Привязки» (рисунок 24) предназначена для связывания правил обработки источников с главными страницами.

У каждой главной страницы может быть своя верстка и исходя из этого нужен свой набор правил сбора информации.

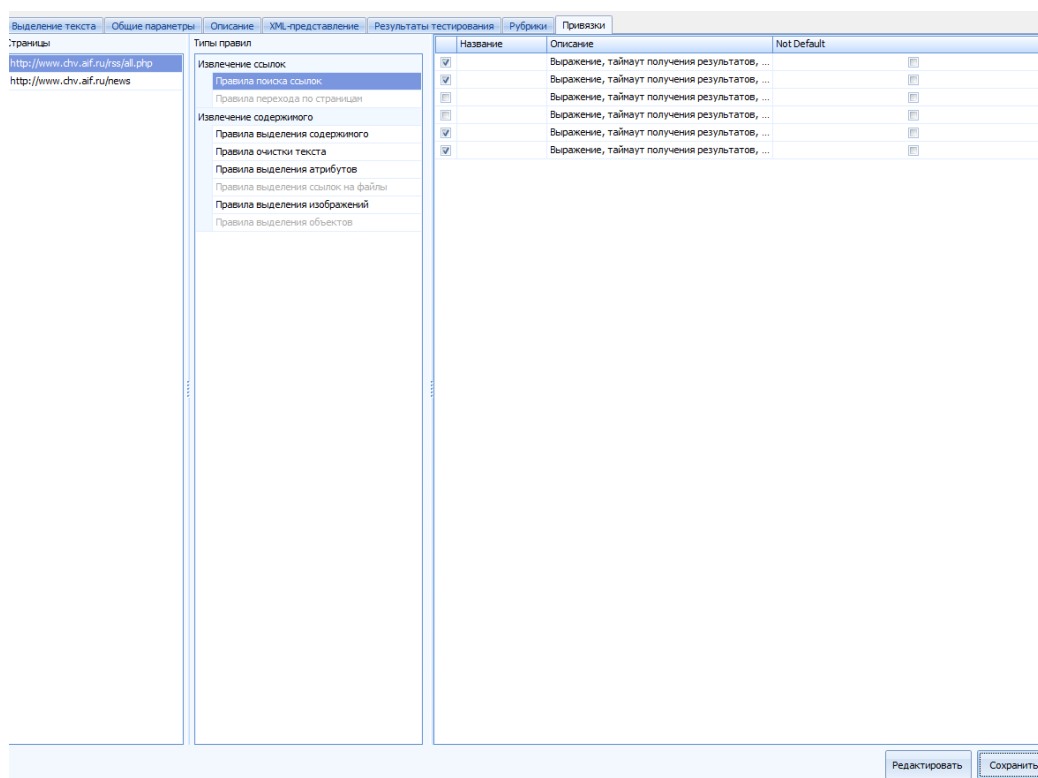


Рисунок 24 – Вкладка «Привязки»