Программа для ЭВМ «Spider – Распределённая автоматизированная клиент-серверная система для сбора и систематизации информации»

Описание функциональных характеристик программного обеспечения и информации, необходимой для установки и эксплуатации программного обеспечения

Оглавление

[1 Общие сведения 3](#_Toc175735080)

[2 Функциональные возможности 4](#_Toc175735081)

[2.1 Шаблоны 4](#_Toc175735082)

[2.1.1 Создание/редактирование шаблона 4](#_Toc175735083)

[2.2 Форматы 5](#_Toc175735084)

[2.2.1 Теги формата 6](#_Toc175735085)

[2.2.2 Значения по умолчанию 6](#_Toc175735086)

[2.3 Администрирование 7](#_Toc175735087)

[3 Информация, необходимая для установки и эксплуатации программного обеспечения 9](#_Toc175735088)

# Общие сведения

Программа предназначена для:

* Сбор данных с источника по заданному шаблону выбранным методом (API, FTP, парсинг, напрямую из БД, скачивание файлов);
* Создание и настройка шаблонов сбора данных с источника.

# Функциональные возможности

Программа состоит из следующих функциональных разделов:

* **Шаблоны**. Содержит список доступных данному пользователю шаблонов сбора; позволяет выбирать шаблоны из списка, применять фильтр/поиск по списку и др.
* **Форматы**. Отображает форматы сбора с соответствующей каждому формату коллекцией тегов; позволяет просматривать теги выбранного формата, создавать новый формат, редактировать выбранный формат.
* **Администрирование**. Содержит перечень настроек для управления Системой и характеристиками основных сущностей.

## Шаблоны

Данный раздел носит в себе функцию отображения всех шаблонов Системы.

Шаблоны представляют собой совокупность настроек и логической схемы получения и преобразования данных. Другими словами, это система визуальных скриптов (блоков схемы), графическое представление процесса получения и преобразования данных.

Шаблоны сбора предназначены для многократного использования, они перезапускаются в зависимости от необходимости обновления данных (настраивается расписание запусков). При изменении содержания информации на интернет-страницах шаблон продолжает функционирование, если не поменялась структура сайта и он был написан универсально.



Рисунок 1. Страница списка шаблонов

### Создание/редактирование шаблона

Данный раздел содержит форму для создания нового или редактирования существующего шаблона. Содержит:

* Перечень параметров, описывающих шаблон (наименование, статус, ссылка на источник данных, график запуска и т.д.);
* Блок-схема, описывающая процесс сбора данных с источника. Блок-схема позволяет использовать различные технологии парсинга в нужной пользователю последовательности. В схемах используются заранее разработанные функциональные блоки, отвечающие своей цели (блок скачивания, блок команд, блок экспорта в xml, блоки для арифметических, логических операций или операций конвертирования данных, и др.).



Рисунок 2. Создание шаблона

## Форматы

Данный раздел отображает форматы сбора с соответствующей каждому формату коллекцией тегов; позволяет просматривать теги выбранного формата, создавать новый формат, редактировать выбранный формат. Создание форматов является обязательным при сборе данных, так как это ключевой фактор для понимания, что и как собирать.

От формата данных зависит то, как и какую информацию собирать с источника и как впоследствии она будет заливаться и храниться в базе данных. От формата данных зависит сложность структуры базы данных, а также сложность сбора данных с источника.



Рисунок 3. Страница списка форматов

### Теги формата

В данной вкладке задается структура данных соответственно источнику, какие данные требуется с него собрать, чтобы соответствовать полноте данных собираемой сущности.



Рисунок 4. Список тегов выбранного формата

### Значения по умолчанию

В данной вкладке присваиваются дополнительные свойства для тегов:

* Имя – обозначение тега, свойства которого задаются.
* Многократный – возможность множественного использования указанного тега.
* Закрывающий – свойство тега, определяющее необходимость его закрытия. Работает следующим образом: по умолчанию для любого формата в значение поля тега берутся все данные от «>» и до «<» (от открытия тега до любого ближайшего следующего тега), но когда требуется брать всю HTML-структуру, то используется свойство «Закрывающий», и в данные поля будут взяты все вложенные теги от открывающегося тега до закрывающегося, и при выгрузке в XML вся HTML-структура в данном поле будет экранирована для избежания ошибок.
* Обязательный –обязательность присутствия тега в выгружаемой XML. Если данное свойство активировано и при формировании XML значения данного поля не будет, то запись не будет выгружена в конечную XML, так как не прошла проверку (т.е. если обязательные поля пустые, то Система выдаст предупреждение).
* Хэш – свойство, определяющее, от каких тегов (полей) будет генерироваться Хэш в тег «Хэш».
* Текстом – экспортирует значение тега так, как оно указано на источнике, без приведения к какому-то определенному типу.



Рисунок 5. Свойства тегов выбранного формата на вкладке «По умолчанию»

## Администрирование

Данный раздел содержит перечень настроек для управления Системой и характеристиками основных сущностей. Список настроек определяется настройками доступа к интерфейсу для авторизованного пользователя:

* **Пользователи** – виден список всех пользователей Системы.
* **Планировщик** – виден список всех доступных планировщиков. Также возможно создать свой планировщик.
* **Потоки** – виден список всех потоков. Также возможно создать свой поток.
* **Маркеры** – виден список всех доступных меток, которые можно проставить шаблону. Здесь же показано количество шаблонов с той или иной меткой. Также здесь доступно создание свой метки.
* **Блоки** – виден список всех блоков, которых можно использовать для создания блок-схемы. Для каждого блока есть краткое описание его функционала, а также описание типов входных и выходных сокетов блока.



Рисунок 6. Список настроек раздела «Администрирование»

# Информация, необходимая для установки и эксплуатации программного обеспечения

Установка программного обеспечения не требуется. Доступ к программе предоставляется по ссылке.

Эксплуатация программы осуществляется посредством веб-интерфейса.