МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

НИЖЕГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ

УНИВЕРСИТЕТ им. Р.Е.АЛЕКСЕЕВА

Институт радиоэлектроники и информационных технологий

Кафедра информатики и систем управления

Лабораторная работа №3

по дисциплине

Управление Данными

РУКОВОДИТЕЛЬ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Балашова Т. И. \_\_\_

(подпись) (фамилия, и.,о.)

СТУДЕНТ:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Калинин Д. А.\_\_\_\_\_

(подпись) (фамилия, и.,о.)

17-АС

(шифр группы)

Работа защищена «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_

С оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

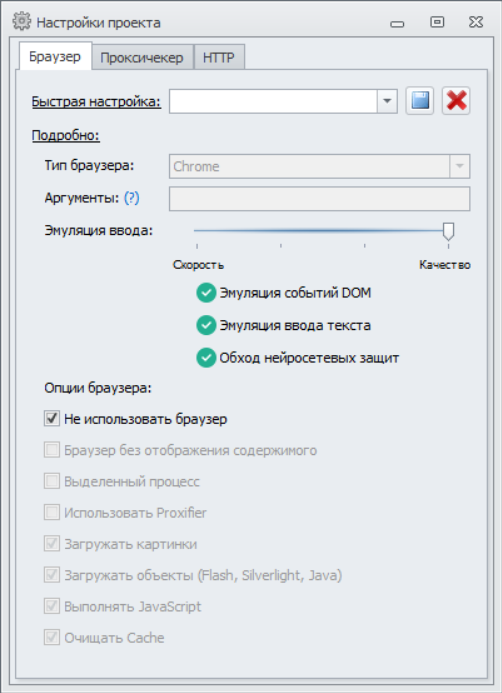
Нижний Новгород 2020

**Задание**

Изучить использование программного обеспечения ZennoPoster на примере парсинга сайта <https://github.com/> с целью получения ссылок на открытые issues репозитория <https://github.com/opencv/cvat>.

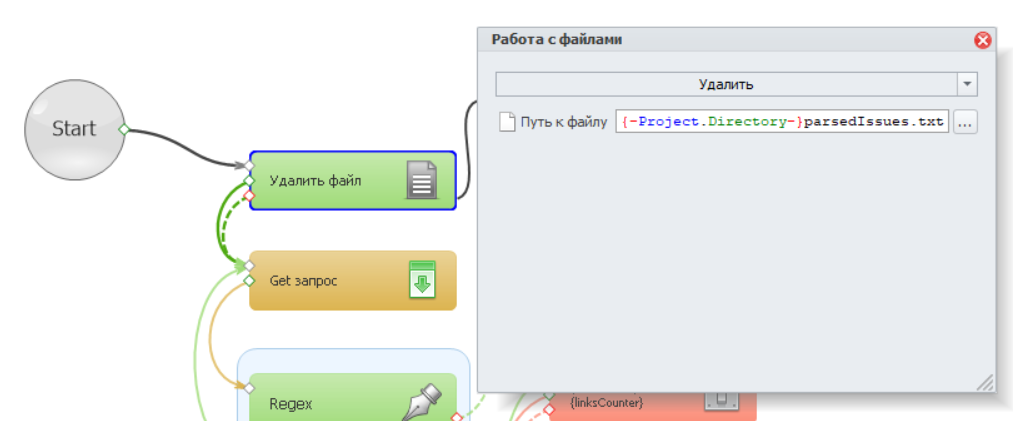
**Выполнение работы**

Так как нам не нужно эмулировать действия пользователя, отключим поддержку браузера.



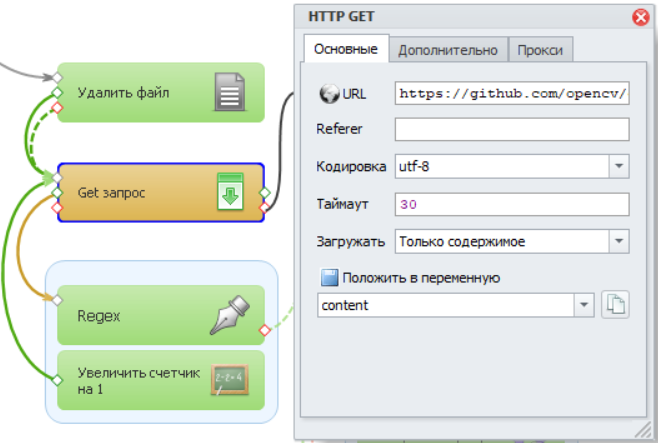
*Рисунок 1. Настройки проекта.*

Очистим предыдущие результаты парсинга.



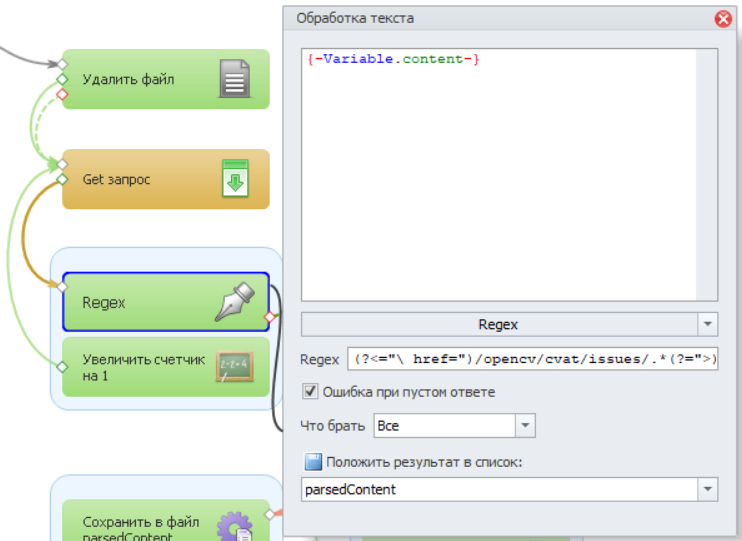
*Рисунок 2. Очистка файла*

В цикле будем получать страницы с issues.



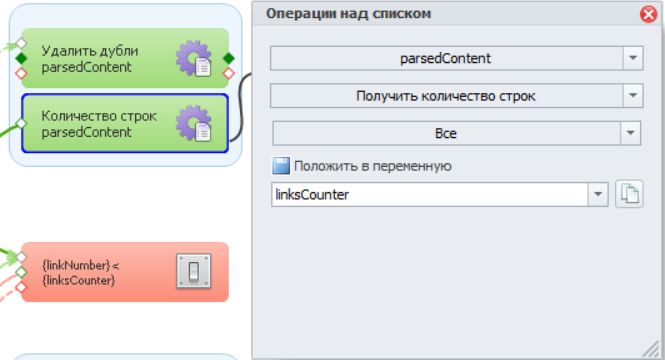
*Рисунок 3. Получение страницы GET запросом.*

Распарсим страницу с помощью регулярного выражения.



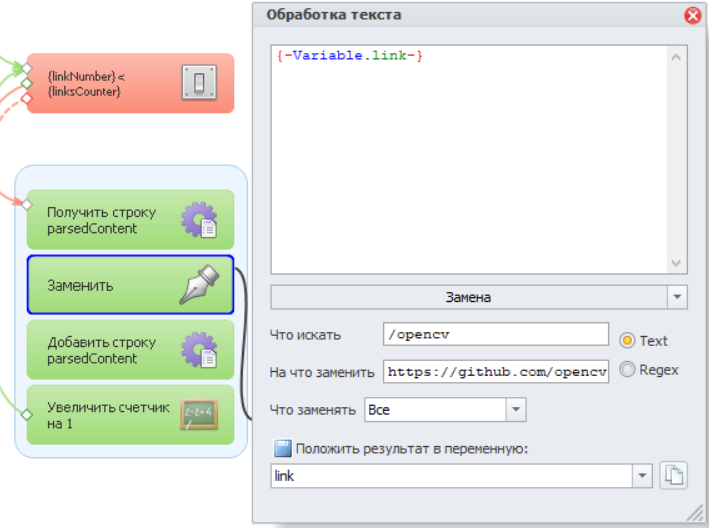
*Рисунок 4. Обработка страницы регулярным выражением*

В случае успеха увеличим счетчик и получим следующую страницу. В противном случае означает, что мы получили все ссылки и можем переходить к обработке списка. В частности, удалим повторения (особенности структуры HTML документа) и посчитаем количество строк.



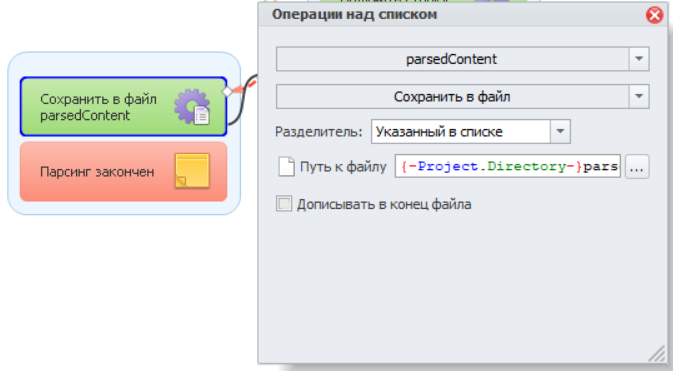
*Рисунок 5. Подсчет количества строк.*

Для каждой строки в цикле добавим домен github.com.



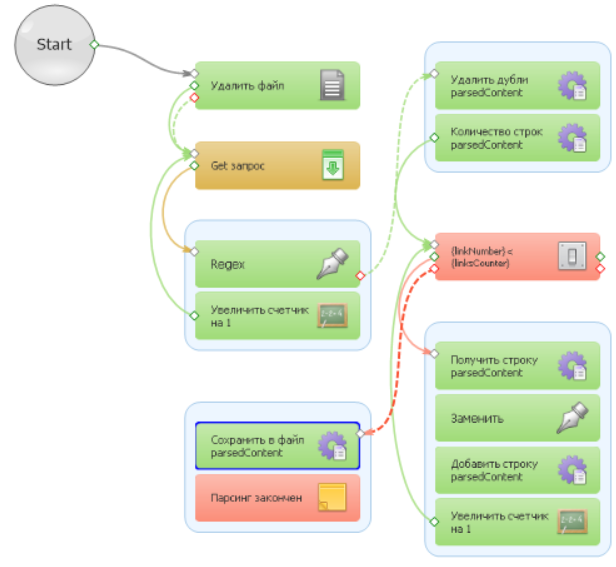
*Рисунок 6. Обработка строк списка.*

В конце запишем получившийся список в файл.

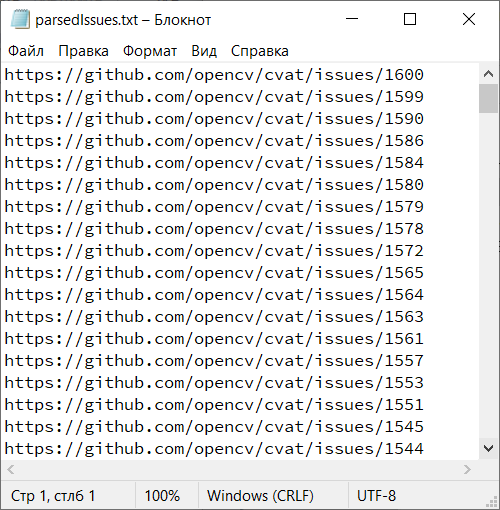


*Рисунок 7. Сохранение списка в файл.*

Блок схема программы парсера целиком и результат его работы.



*Рисунок 8. Блок схема программы.*



*Рисунок 9. Результат работы парсера.*

**Выводы**

По результатам выполнения данной работы мы познакомились с программой ZennoPoster и создали простой парсер, который создает список ссылок на открытые issues.