**ОПИСАНИЕ ПРОГМАММЫ**

Программа запускается командой:

python main.py <URL>

Например:

python main.py https://moslenta.ru/news/lyudi/formula-poleznogo-zavtraka-01-08-2022.htm

Для каждого новостного сайта необходимо указать поисковые параметры в файле settings.txt. Первая строка в этом файле это ключевой тег поиска (например div), вторая строка это css класс этого тега, можно пропустить если нужно искать только по тегу.

**ОПИСАНИЕ АЛГОРИТМА**

Алгоритм основан на рекурсивном спуске по DOM дереву. Символы <,>,/ - позволяют определить начало и конец искомого элемента. Для улучшения поиска и исключения ошибок содержимое страницы очищается от стилей, одинарных тегов и javascript вкраплений.

**ВОЗМОЖНОЕ УЛУЧШЕНИЕ**

Так как программа делалась в сжатые сроки не реализованы многие возможности. Например можно добавить поиск по нескольким CSS классам сразу, можно улучшить отображение выводимого в файл текста, тем самым расширив список сайтов по которым можно вести поиск.

Так же можно сделать код «чище» и «читабельнее» используя например, регулярные выражения. Думаю в определенных местах можно провести оптимизацию, рефакторинг всего приложения.

**СПИСОК URL, НА КОТОРЫХ ПРОВЕРЯЛОСЬ РЕШЕНИЕ**

https://lenta.ru/news/2022/08/01/monkeypox/

div

topic-body\_\_content

https://ria.ru/20220802/kulikovo\_pole-1806469827.html

div

article\_\_text

https://www.interfax.ru/russia/855185

p

https://tass.ru/ekonomika/15392023

https://tass.ru/mezhdunarodnaya-panorama/15368327

https://tass.ru/ekonomika/15395389

p

https://aif.ru/sport/football/spartak\_na\_svoem\_pole\_razgromil\_orenburg

div

article\_content

https://moslenta.ru/news/lyudi/renata-litvinova-31-07-2022.htm

https://moslenta.ru/news/lyudi/simptomatika-01-08-2022.htm

https://moslenta.ru/news/lyudi/formula-poleznogo-zavtraka-01-08-2022.htm

div

\_39FNd9SD

***Пояснение***: под url адресами указаны параметры которые нужно вписать в файл «settings.py» (тег поиска и css класс поиска) Результаты поиска находятся в папке «результаты»