

Приложение к заданию на выполнение курсового проекта

Краткое описание программы:

Разработать парсер HTML на языке Python

Полное описание задачи:

Разработать для приложения модули:

1. Лексер

В лексер подается на вход содержимое HTML файла. Он по определенным паттернам читает этот файл и создает лист кортежей, состоящих из содержимого этого элемента и его название. Возвращает этот лист.

2. Парсер

Получает на вход лист элементов, сгенерированный в лексере. Последовательно читает эти элементы, создает подходящие к ним объекты соответственных классов(Tag, Text) и рекурсивно добавляет их в объект DOM(Document Object Model).

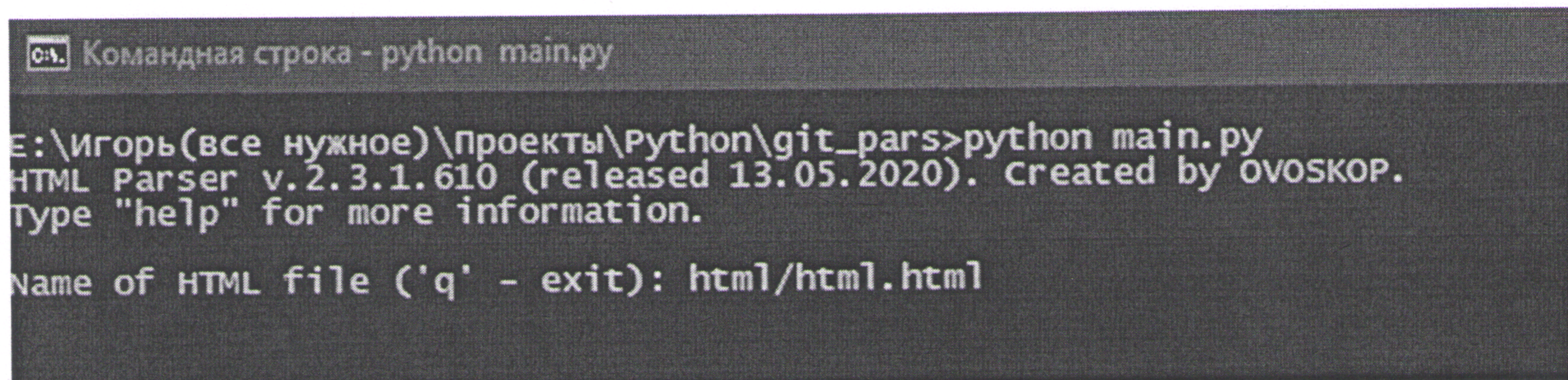
3. Главный модуль

Получает на вход имя HTML файла, считывает с него данные, передает в лексер, после этого передает полученные из лексера данные в парсер и получает DOM. После выводит DOM в виду AST(Abstract Syntax Tree) и если в HTML есть ошибки, то выводит их как warning.

Входные данные:

На вход подаётся имя HTML файла.

Пример ввода имени HTML файла:



```
Командная строка - python main.py
E:\игорь(все нужное)\проекты\Python\git_pars>python main.py
HTML Parser v.2.3.1.610 (released 13.05.2020). Created by OVOSKOP.
Type "help" for more information.
Name of HTML file ('q' - exit): html/html.html
```

Рис. 1. Ввода имени HTML файла

Выходные данные:

Выходные данные представлены AST деревом.

Пример AST дерева: