

# Project 1- text parser

Saba Yousefian Jazi

This program is written with python 3 and jupyter notebook.

To run it, you have to have jupyter notebook installed. (A .py version is provided beside the .ipynb file in case there was any problem in running the notebook file.)

To run the program, the files that should be parsed and the stop\_word file have to be located in the same folder as the code.

The output will be made as a text file named "parser\_output" in the same folder the code file is located.

The code consists of 2 blocks in the notebook.

First block is for importing the libraries used in the code and the second block consists of the main functions.

These blocks should be run one after another, in the order they are placed.

The libraries that were used are:

ElementTree : to extract each part of the file between the tags in a tree data structure.

Re: to split the text(in words) in every non alphabetic character.

PorterStemmer: to use the porter stemmer algorithm for stemming process

The output file is provided with the code.