LAB - 4 Exercises

Problem- 1:

Collect any 10 documents (English text documents) from the web and create inverted index by doing necessary preprocessing steps using python.

Challenging Exercise-1:

Collect any 10 documents (Indian Language text Documents in Unicode) from the web and create inverted index by doing necessary preprocessing steps using python.

Challenging Exercise-2:

Collect any 10 documents (Documents in different formats such as PDF, DOC, ODF) from the web and create inverted index by doing necessary preprocessing steps using python.

Resource link

https://github.com/REMitchell/python-scraping