**Pre-requirements**

These are the libraries that are used in the script:

**To read from PDF:**

PyPDF2

**For Pre-processing of text**

NLTK

**Summary of the script**

PyPDF2 was used to read and iterate through the pages of the pdf document.

NLTK was used to pre- process the text like tokenizing, using stop words to eliminate useless words, and stemming.

Frequency was calculated by NLTK and hence the relevant keywords were found.