**Politiko - A Profession Specific Information Retrieval System**

### **Aim** Created a profession specific information retrieval system pertaining to professions such as journalism, politics and law.

### **Scope** Given a corpus from the Prime Minister’s speeches and a search query, we are meant to extract the speeches referencing several issues addressed in the query. We are also to indicate if a particular issue has never been addressed by the Prime Minister before.

### **Order of execution** Run Jupyter notebook. Execute each code block (cell by cell) in code.ipynb notebook.

$ sudo python3 code.py

### **Installing nltk**

$ pip3 install nltk

$ python3

>>> import nltk

>>> nltk.download()

Packages: all

**DATA STRUCTURES USED:**

**Document\_tokens\_list**Contains lists enclosed within a list It will contain the stemmed tokens from each file in the corpus as individual lists. All are appended to make a list. Example:

[[‘i’,’am’,’good’],[‘isro’,’nasa’],[‘india’,’is’,’best’]]

**Vocabulary**

Will contain a dictionary of all the unique words in the corpus. Example:

{‘i’: 1, ‘am’:2, ‘good’:3, ‘isro’:4, ‘nasa’ :5, ‘india’:6 , ‘is’ :7, ‘best’:8]

**Members**

Pratik (2016B4A70549H)

Anshuman Pati (2016B4A70470H)

Piyush Jain (2016A3PS0885H)