LAB EXERCISE 1 (17 Aug 2020)

Explore the working of the Bag of Words IR model. Implement the concepts like basic text preprocessing, vocabulary construction, and n-gram modelling (available in NLTK Python)

You may consider different document collections available in Python NLTK (e.g. Brown corpus, Project Gutenberg Corpus) and explore how to use these for retrieving documents.

Submit a detailed report on the findings of your analysis, supported by the necessary code snippets w.r.t your program and results.

Note:

- Lab Exercises are meant for exploring basic concepts, and will not be evaluated. However, the concepts learnt will be helpful in completing lab assignments will be considered towards 15% of the continuous evaluation component.