

Natural Language Processing Lab

Exercise 1:

Take your own input document with multiple sentences. Identify the important sentence among all the sentences in the document.

Steps to be applied:

- 1) Pre-process the input document.
- 2) Encode each token using the following techniques:
 - i. One-hot
 - ii. Bag of words
 - iii. Tf-idf
 - iv. Count vectorization
- 3) Apply distance metrics to identify sentences that are closer to each other.
- 4) Identify the important sentence
- 5) Compare each encoding techniques and analyze their performance.

Submission: A PDF file containing the code and several input with which model is evaluated should be submitted on the LMS page.