Assignment 1:Text Summarization using Python & NLTK: TF-IDF Algorithm

10 Points Possible

8/15/2024

Attempt 1

In Progress

NEXT UP: Submit Assignment

Add Comment

Unlimited Attempts Allowed

7/30/2024 to 8/15/2024

Details

Problem Statement

Build. A summary of long pieces of text keeping key information content and overall meaning. The summary must represent the most important or relevant information within the Text.**Process Steps**

Steps involved to create the text summary

- Tokenize sentences
- Create frequency matrix of words in each sentence
- Calculate Term Frequency and Generate matrix
- Create a table for documents per words
- Calculate IDF and generate matrix
- Calculate TF-IDF and generate matrix
- Score the sentences
- Find the threshold
- Generate the summary

Perquisites

- Python 3
- NLTK Toolkit
- IDE or Text Editor

Submission Instructions

A PDF document has to be uploaded on Canvas under ‘Assignment’ covering following:

- Overall process description & solution approach
- Tool used and reasons to use this specific tool
- Source code snippets
- Final output results and analysis of results

Note: Each document page should have student's BITS Id.

References

Refer following for detailed steps and examples of text summarization case studies.

https://towardsdatascience.com/text-summarization-using-tf-idf-e64a0644ace3?gi=ebc5e81b9984 (https://towardsdatascience.com/text-summarization-using-tf-idf-e64a0644ace3?gi=ebc5e81b9984)

https://medium.com/@ashins1997/text-summarization-f2542bc6a167 (https://medium.com/@ashins1997/text-summarization-f2542bc6a167)

Evaluation Criteria

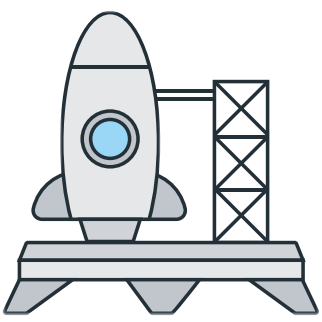
This assignment is of 10 marks

S.No.	Evaluation Task	Marks
1	Overall solution design and process architecture	3
2	Tool used and reasons to use this specific tool	2
3	Final output results and analysis of results	3
4	Document quality (structure, detailing, presentation etc)	2

Choose a submission type.

Upload

Arc



Drag a file here, or

or

Canvas Files