Damietta University
Faculty of Computers and Artificial Intelligence

Information Storage and Retrieval Course, Year 2022/2023

Task (2)

1- Write a python program than takes a text from a user and ask the user to choose one of three choices:

Choice number 1: print tokenized **words**

Choice number 2: print tokenized **sentences**

Choice number 3: print original text.

Example (1):

Input: " NLTK is a leading platform for building Python programs to work with human language data. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project."

Input: Choice number 1

Output: `['NLTK', 'is', 'a', 'leading', 'platform', 'for', 'building', 'Python', 'programs', 'to', 'work', 'with', 'human', 'language', 'data', '.', 'NLTK', 'is', 'available', 'for', 'Windows', ',', 'Mac', 'OS', 'X', ',', 'and', 'Linux', '.', 'Best', 'of', 'all', ',', 'NLTK', 'is', 'a', 'free', ',', 'open', 'source', ',', 'community-driven', 'project', '.']`

Example(2):

Input: " NLTK is a leading platform for building Python programs to work with human language data. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project."

Input: Choice number 2

Output: `['NLTK is a leading platform for building Python programs to work with human language data.', 'NLTK is available for Windows, Mac OS X, and Linux.', 'Best of all, NLTK is a free, open source, community-driven project.']`

2- creat an account of Kaggle https://www.kaggle.com/

3- search for a dataset in a preferred topic on Kaggle

https://www.kaggle.com/datasets

Bonus:

Read the csv file of the dataset and print first 5 rows.

Remarks:
- use nltk library
- Due Data is Sunday 12/3/2023.
- Solve this task **individually**.