Assignment 1

Delivery Notes:

- This is a group assignment of 3 members (at most)
- All students should work and fully understand everything in the code.
- Due date is on 17/3/2024.

Assignment Details:

In this assignment, you are required to extract text from Html page and apply processing in the text then get the unique words from the text:

- **Data:**
  1. USE any URL (e.g. Wikipedia),
  2. Extract HTML from URL
  3. Extract TEXT from HTML page e.g. (p or headings)
  All 3 steps in python
- **Processing on Data:**
  1. Cleaning data from each symbol or character doesn't contain to the data.
  2. Normalization: make all the data to lower case
  3. Tokenization: split the data to words
  4. Lemmatization or Stemming: return each word to origin.
  5. Stop words: remove stop words from the data.
- **Unique Word:** the output of the assignment to get all the unique words on the data.