CS173 - Dr. Kretchmar
Group members: Linh Nguyen & Nguyen (Nicolas) Nguyen

Project 1 Report

I. Python routines

In this project, we will create 2 functions other than main(). One will be responsible for splitting the text into sentences and stripping whitespace from each one. The other will build a dictionary classifying where each word appears in the text.

II. Approach

Create a List where each element is a sentence from the text:

1. First, we will read the text and split it into sentences, taking advantage of the fact that sentences are only separated by periods.
2. Next, we create an empty list to store our sentences.
3. Using the strip() method, we will loop through each sentence in the list created in (1) and remove the leading and trailing whitespaces in each one.
4. Then, we append each sentence to the empty list.

Build a dictionary that has words and the indices of sentences they appear in:

In order to keep track of where each word appears in the text, we want to create a dictionary where the key is the word itself, and each corresponding value is a list of indices indicating the position of the sentences (that the word appears in) in the sentence list (e.g. {'jane':[1,2,3,8,11], ...}). This way, we can check if a word is in the text by checking if it is in the dictionary keys. Then, we can easily print the sentences it appears in by printing sentences from the list with the corresponding index.

To match the formatting of the sample output, we use the lower() method to convert all characters into lowercase, then sort the dictionary keys using the sort() method before printing them to the terminal. To print out the correct sentence number, we add one to each sentence index since we're counting from one, not zero.