Information Retrieval 8130

Assignment 3

Submitted by: Hasan Mashrique

**Due Date:** 9/28/2022

**Problem 1:**

Answer in **problem1.md** file (view in preview mode)

**Problem 2:**

The code for this problem extracts all the vocabulary from the course website main page and other related links from the main page.

For each word,

* it counts the frequency of the word from all pages.
* shows how many documents it occurs in.

So, the code outputs 2 dictionaries (hash table):

1. Frequency of each unique word in the corpus/document collection.
2. Occurrence count of each word in the documents.

Here, not all the links are used from the main page. Some links are not working, and some are not allowed/forbidden to enter. So, the working links are copied in a **workinglinks** list for ease of parsing the correct links.