Sasha Smolyansky
Professor Narahari
Computer Architecture
6 December 2018

CSCI 2461 Project 5 Questions:

*** Reference flowchart for cyclic explanation of program. Reference program comments for step by step description of program. ***

a. When searching for a search query inside a linked list, the time it would take would be O(N * m_i * K). The function is N number of documents multiplied by m_i words per document (particular to each document) multiplied by K number of words in the search query input.
b. If the hash function maps all the words evenly across the buckets, then the time complexity will be essentially be the same as the big O notation for the linked list, however, when using buckets (a more efficient method), the entire function will be divided by the number of buckets (b). Thus, the big O notation for this solution would be O((N * m_i * K)/b).
c. The elimination of stop words within the program leads to a more efficient search process because it eliminates words that are common to all the documents in order to focus on the unique words which distinguish specific documents from others.
d. The first method you cycle through all the nodes and calculate the ‘idf’ which is needed to remove the desired stop word, but for the second method you find the specific node which already would contain the desired information. The second method is faster and more efficient due to the pre-calculated information that is needed. I implemented the first option.