Kenneth Chuson  
CSC 620-01 Natural Language Processing Spring 2022

HA#11 Write-Up

Between Program\_0 and Program\_2 they have different document rankings. Which means they have different documents of text. It depends on their frequency (how many each word occurs in each document) for the term frequency. So, the Program\_0 has more frequencies because there are some repeated words in each document than Program\_2.

Also, it is similar when you normalize term frequency which calculates the number of terms and divided by normalize document.

For the Inverse Document Frequency (IDF), is a bit different, we are trying to find out relevant documents matching the query. Using logarithms helps to weigh down the effects when the frequency is occur too much. So, Program\_2 has lower probability of each most of the words from each of the document than Program\_1. I believe Porgram\_2 has better ranking because it does not have repeated words (less occur of frequent words).