

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 Program_2 has better ranking because it does not have repeated words (less occur of frequent words).