**Comp 5970 Assignment 2**

**Haden Stuart**

**1. N-Gram Language Model**

1. The central assumption we are making in N-Gram Language Models is that each word is independent in the sense that a word n is only dependent on the previous n-1 words.

1. This is a reasonable assumption because in order to calculate the probability of the next word, we will only need to see the words that come before it. Words that will take place after a word have no effect on the word itself, so we really only need to look at the previous words to calculate the probability of a word occurring.

1. TBD

**2. Unigram Language Model**







**3. Bigram Language Model**



**4. Vector Space Model**

1. Using dot product similarity for the given document and query, we get a score of 7.

1. For all words that are present in the document, this would decrease the IDF value. For all words that are not present in the document, this would increase the IDF value.



**5. TF-IDF Heuristic**