1. Your NDCG@5 for 1) and 2)

2. Explain what you did in order to improve NDGC@5.

3. Include any other information that may be relevant for this milestone

Improvements

* Added Weights based on Title and Anchor Text. These weights will add more weight to documents that contain the query terms in their Title and Anchor texts.
  + We noted that the first few results returned would be results containing the title or anchor text of queries. The remaining results might not contain the query terms in either the title or anchor text.
* Add 2-Gram, 3-Gram, and 4-Gram weights per documents. Documents that contain a permutation of the query terms in either: 2-gram, 3-gram, or 4-gram will have their weights adjusted. The weights of the N-grams are based on the TFIDF of the N-grams among documents.
* Every file we searched (had a different file for each n-gram type, a file for all titles, a file for all anchor texts, a file for TFIDF numbers, and a file for all positions) was run on a separate thread. Since each file was independent, we could speed up search by doing file searches on all different threads.
* The positions were used to find the relevant text snippets related to the query terms
* Modified Tokenizer to treat Capitalized letters as a split between words, in case the user or a document forgot to put a space to separate out the words
* Used the hyperlinks from each file as a guide for a hub for each file. Each link that is targeted will have an authority associated with how many links point from each hub file. The top links will only be files that meet an authority threshold.

Information

* Time to pull all ten documents with text snippets
  + 1874 ms
* Time to pull all ten documents without text snippets
  + 776 ms
* Time to pull a one word query with text
  + 240 ms
* Time to pull a one word query without text
  + 210 ms
* Time to pull a four word query with text
  + 384 ms
* Time to pull a four word query without text
  + 302 ms