**Implementation:**

Retrieve the inverted index for the corpus from the file created in HW3.

Retrieve the file containing document tokens dictionary created in HW3.

Retrieve the document Id dictionary created in HW3.

Retrieve the file containing document Id and their terms created in HW3.

Read the file containing queries and put them into a dictionary.

* Calculate the average document length using the doc\_tokens\_dict which contains the number of tokens in each document.
* For each query:
* Calculate the BM25 score. The score can be calculated as follows:

Calculate the weight of each term in the query and store it.

For each document in the corpus:

Get the document length (number of tokens in document)

For each term in the query,

Retrieve its frequency in the document and calculate its bm25 score as:

Where K is:

Since there is no relevance information available, ri and R will be 0.

ni is the number of documents in which the term appears.

N is the number of documents in corpus

fi is the frequency of the term in the document

qfi is the frequency of the term in the query

k1 = 1.2

k2 = 100

b = 0.75

In order to avoid error due to absence of query term in a document, smoothening has been done to the component.

* Sort the documents based on their BM25 scores and write the ranked documents to a file