DOCUMENTATION

1. *read\_all\_docs*( ) : This function takes no arguments and updates the *all\_doc\_vectors* list with term-frequency vectors of each document in the dataset.
2. *input\_vector*( query ) : This function takes the user query as input and returns a term-frequency dictionary.
3. *inv\_index\_all\_docs*( ) : This function takes no arguments and creates the inverted index for all documents in the *inv\_index* defaultdict.
4. *tf\_idf\_vectorize*( ) : This function takes no arguments and converts all the term-frequency vectors in *all\_doc\_vectors* to TF-IDF vectors.
5. *tf\_idf\_query*( query\_vector ) : This function takes the term-frequency vector of the query as input and converts it to a TF-IDF vector.
6. *tf\_idf\_score*( word, frequency ) : This function takes a word and its frequency as inputs and returns its TF-IDF score.
7. *dot\_product*( vector\_a, vector\_b ) : This function takes two unit normalized vectors as input and returns their dot product.
8. *stem\_and\_tokenize*( doc\_text ) : This function takes a string of document text matter as input and returns a stemmed and tokenized list of words in the document.
9. *create\_vector*( token\_list ) : This function takes an input list and creates a document frequency dictionary in *doc\_freq* dict.
10. *doc\_string*( doc\_id ) : This function takes a document’s ID as input and reads it and returns its text content as a string.
11. *query\_result*( query\_vector ) : This function takes unit normalized TF-IDF query vector as input and returns a list of DocIDs and their weighted cosine scores in sorted order as output.