**Documentation**

Pre-processed files:

1. File idf.csv has all the idf scores written in it.

2. File output.csv has all the tf-idf scores.

document vectors.py:

1. Check():

Takes a list as input and checks if a term is present in the list, if it is present returns the term frequency otherwise returns -1.

1. Idf():

This function is used to calculate the IDF values of each document with respect to input term number. It adds all the term frequencies of a term for the entire corpus and returns the IDF value by the formula IDF=log(N/Df).

1. tf-idf():

Takes the keys in train dictionary as input, calculates term frequencies by 1+log tf and multiplies it by idf score from the file idf.csv and returns the tf\*idf score.

isc\_calculation.py:

1. q-vector DS:

Contains the tf-idf scores of the query generated using the same technique as used for the documents in the corpus.

1. isc\_score DS:

It is a list created to store and sort in descending order the similar documents(songs) out of which top 10 are returned.

1. Music.txt and words.txt contain the data which are handled using standard file handling procedures of python.