**Search Engine for Song Lyrics**

The application consists of 2 files, build\_index.py and ranking.py.

Build\_index.py takes as input word 5000.txt and test.csv and stores the inverted index in a pickle file for consequent retrievals.

Data Structures :-

* List
* Dictionary
* Tuple

build\_index.py :-

This file takes as input 5000 word.txt which is the list of all words in the set of songs in the corpus, test.csv contains songID and their (word,frequency) pairs. Using these an inverted index is build and is stored using pickle(object serialization) to be further used in ranking.py

ranking.py :-

Takes as input a query string and returns top 10 songs with lyrics matching that query.

The input query is converted to lowercase, whitespace is removed and for each word in its inverted index calculates the score for a document w.r.t that query.

Calculating score for ranking documents :-

The score is calculated by adding (1+log10(term frequency)) of a specific term in that document, and is done for all individual words in the query.