Twitter Keyword Search

Members: Khang Le, Aryan Gupta, Ehsan Sabouni, Roman Velez, Mayank Yadav

Input/Output

Inputs:

- Database of tweets (.txt file)
- User inputs keywords

Output:

- Returns top 10 most similar tweets to the user's keywords in order

Approach

Goal: Speed of search

Design choices: Data structures (inverted index), algorithms

Plan: Create simple UI that allows user to input query keyword(s). Find all the tweets that have at least 1 similar word to the query search, compute the similarity for all those tweets, and rank the top 10 out of all those tweets. This will be done with the inverted index data structure as it is widely used in information retrieval system to to efficiently retrieve documents containing specific terms.