Project Initial Proposal

Team Name: GroupOf4

Team Member

Ying Sun(yingsun3) - Captain Chang li(changli8)

Yuteng Zhuang(yutengz2) Qinhan Xia(qinhanx2)

1. **What is your free topic? Please give a detailed description. What is the task? Why is it important or interesting? What is your planned approach? What tools, systems or datasets are involved? What is the expected outcome? How are you going to evaluate your work?**
    - Free topic: Lyrics Smart Search

    - Description: Sometimes, people want to search for the song they want by keywords, but most of these searches are based on the name of the song. We want to build a search engine for lyrics, where people can search for the songs by lyrics. Also, we want to implement a keywords based search engine, but its search results are accompanied by sentiment analysis of lyrics(happy, sad, etc.), which makes it easy for users to understand. At the same time, we hope to build other search engines based on emotion keywords, so that users could search for songs based on their mood.

    - Why important: The need to listen to songs according to one's mood is common, but there is no well known existence sentiment based song search engine available.

    - Tool: Rasa, metapy, flask, Nltk, pandas, matplotlib, seaborn, WordCloud, re, sklearn
      System: MacOS, Windows
      Dataset: **Billboard Hot-100 Songs 2000-2018 w/ Spotify Data + Lyrics**

    - Expected Outcome:
      Keywords based Search: Top 10 rank songs according to the correlation from the input sentiment
      Sentiment based Search: Top 10 rank songs with descending score and their emotion from lyrics.

- Test with User-Based Evaluation: Sample dozens of people around, record their reviews and evaluate the accuracy of search results.

2. **Which programming language do you plan to use?**
    - Python
    - React

3. **Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.**
    - Front end: lyrics frequent visualization - 20 hours
    - Implement keywords based search algorithms - 10 hours
    - Implement sentiment based search algorithms - 20 hours
    - Implement sentiment analysis - 10 hours
    - Set up server & connect front end- 10 hours
    - Testing - 20 hours