I haven’t decided on a name yet, although I have drawn up a rough outline of how my program will be laid-out. My project is going to be centered around the website [Genius](https://genius.com/), which is the most popular lyric platform for music released primarily in the US. For my outline, I’ve split the program into two parts, the scraping side, and the visualization side.

For the scraping portion of my program, I’d like to first go over the two input modes, automatic and manual. Automatic will designate the building of the ‘data structure’ with an input of the lyrics of the top 50 songs of genius at the time of query. Manual will prompt the user for a search term, scrape and output the potential songs related to the search term, and prompt the user to select the song from the output(Numerically, and not “Never Gonna Giv…”). This is quite a fledgling concept, and since this is the final project(Being a summation of all previous works), I may add more functionality here if I have time(Possibly a more generic search(grab all songs for given artist, genre, etc). For additional libraries, I'm planning on using those that were displayed in-class and that we have used previously.

For the visualization portion, I'm going to be keeping the data in a single-use data structure, which will be an array of dictionaries. The dicts will contain the song name, artist, producer, lyrics:[], a counter object generated from the lyrics, release date, and probably quite a bit more. I definitely know that I'm going to try to have visual as well as text outputs, however, I'm planning on outputting both through matplotlib pyplots. This would probably look like :

| Song Name  Artist Name…  … | <bar graph containing lyric word occurrences> |
| --- | --- |
| <Image of song/album cover> | <bar graph containing lyric letter occurance> |

For regular song outputs, or a longer running output of all the words that have been indexed. I don’t presume the transfer of data between the data structure to pyplots will be too difficult, and a potential addition I may incorporate is a pyplot with an output of the song lyrics, each individually colored on a gradient(Basically a lyric heatmap). I will also try to incorporate some sort of stop word prompt as well(either via .txt or User Input). As I said previous, I’d like to make this program pretty robust, so If you can add any input of further visualization or classification techniques I could use for this type of data, it’d be much appreciated. Everything stated in this proposal is proof of concept, although I plan on sticking to what I have(Unless you disapprove of this Idea). The most free form aspect of this assignment is the pyplot outputs, I am one who likes maximalist design philosophies, so expect quite a bit more user interaction with output methods.