Rose Lauture & Wendy Posada
Problem Solving and Software Design
30 October 2017

Project Overview:

The magazine Rolling Stone provides their readers with lists of top popular music, movies, and TV shows. Therefore, it provided the perfect source for text mining. We used rollingstone.com to compile the data for the all-time top 5 hip-hop, country, Motown, R&B, and Boy-band songs. We were able to navigate through Rolling Stone to find this data in 5 different articles. Through compiling this data we set out to find out why these songs were the top songs. We would conclude why they are the top songs due to the familiarity in lyrics.

Implementation:

When it comes to working with the data we needed to import the lyrics. Therefore, instead of manually inserting the lyrics we imported a function called Py-Lyrics where we were able to import the lyrics for each songs. This was a crucial part when it came to analyzing the data. Also to keep our data organized we decided to add a function where we were able to retrieve the lyrics on another file. This allowed the code just to output our findings in one document and keep the lyrics in another file.

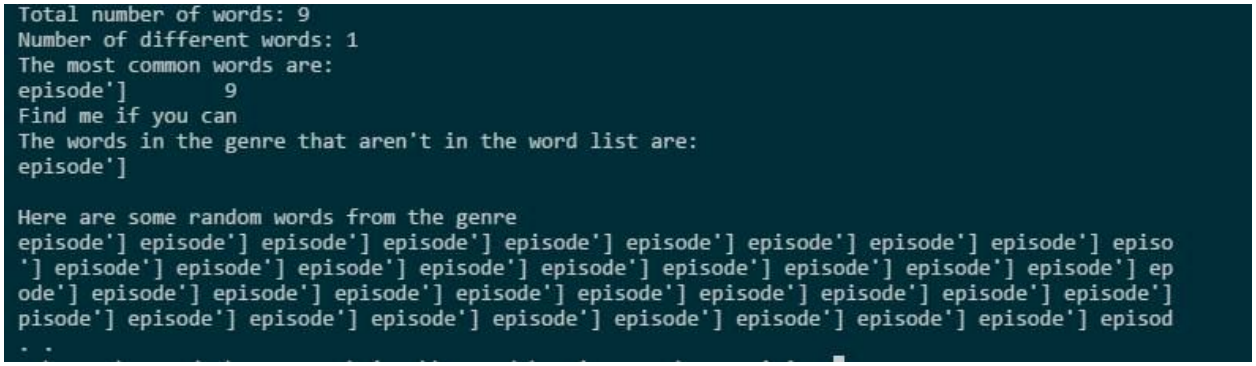
The second part of our implementation process was the data analyzing component. We used the following functions most_common, total_words, different_words and random_words. Total_words - returns the total of the frequencies in a histogram, different_words - returns the number of different words in a histogram, most_common - makes a list of word frequency pairs in descending order of frequency, random_word - shows us the probability of each word in proportion to its frequency for a random given word. These functions allow us dive into our data and she the following results.

Results:

Once we ran the code, we received interesting results. We were assuming that the songs that were popular were going to be because of love or money, but it turns out the most common word in the data was episode. Looking at the figure below, episode came up 9 different times in the songs. Therefore, we can assume that there is not any one particular reason that the songs are popular.

Figure 1

Rose Lauture & Wendy Posada
Problem Solving and Software Design
30 October 2017



Reflection:

Looking back at the project, we believe we learned a lot but there are a couple of things we would do differently. The first part of the project would be to consult with Professor Zhi with our thinking process therefore, it would help us work quicker and more efficient. As well as using a site where they explicitly gave us a definite list instead of us using, “Top ten R&B Artists” and trying to pull top five songs. That would have given us a more solid answer. We also originally planned to analyze 5 different genres but we were unable to get all the different files to open. Therefore, we had to only analyze on the hip-hop genre. Besides our setbacks, we believed we worked well as team. We split the work evenly and assigned each other tasks depending on our strengths. When an issue arise we worked together to solve the problem. As for what we learned through this assignment, we learned that python has many different features that allows us create different modules. For an example the import Py Lyrics function. Also we learned how that it takes many trial and errors until you are able to create a functioning code. In conclusion, we believe we are able efficient text mine and analyze a set of data.