Alexander Dean

COS 285

1 November 2017

Report: Program 7

The purpose of Program 7 was to begin implementing a class that can search the SongCollection for a word in the lyrics of a song. This particular assignment dealt with creating the Map of Sets to hold all of the words in all of the songs. The Data structure, thus far, consists of two parts: the TreeMap of each word, and the sets each word maps to containing all of the songs containing that word. The results and analysis given by the program can be found in the below figure.

Figure : Results from Program 7

Number of keys in map: 31385

N: 637601

Size of map: 188310

Size of all Tree Sets: 3825606

Size of Data Structure: 4013916

Size usage: 6N

The extra credit portion of this assignment dealt with determining the ten most common words in the SongCollection. This was implemented by creating a new class called RankedItem that tracked the name and count of an item. The count, in this case, was the size of the TreeSet that a word mapped to. This allowed the use of another TreeSet to store the ten most common words. The program iterates through the TreeMap and checks if the count for the current word is greater than the least-common word in the current set of top words. If it is, it is added, and the least common word is removed. Because this iterates over the entire map, it compares 31385 times to check every item. If N is 637601, then it runs (1/20)N. This value will change based on how many different words are in a song. So, if K is the number of words in the map, O(k) is the runtime of this method. The results of determining the ten most common words can be found below.

TOP WORDS:

Word Count

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

up 3544

down 3607

ll 3614

out 3644

re 3885

love 3905

like 3986

know 4318

don 4393

just 4748