CS214: Systems Programming, Spring 2018
Assignment 2: Keyspace Construction – Readme

Xiao Yan(xy184)
Satita Vittayaareekul (sv439)

Struct & Linked list:
The data structure we use in this project is linked list because we don't know how many tokens or filenames we need to keep track of.
We use two structs to keep track of the token and each token's file appearance.

Tokenizer:
The tokenizer we write first separate the input into a small token struct. Then it will be checked if the token has existed in linked list, if it has, then we checked if the file names linked with the token has repeated or not, if yes, we increase the count(rating) of the filenames, if not, we add a new node of new filename behind of the old filenames. If the token has never existed, we add the new token at the end of the linked list.
So our tokenizer not just separate string into token but also linked them together so that we could keep track of it and use it later.
The linked list it creates is not sorted but has no repeated elements.

Sort:
We have a sort function which will sort the existed linked list, both token and file name, by using qsort..

Qsort:
Adding the address of each node into an array, we use qsort function to sort the member of the array based on the comparison function we write by ourselves. We create our own string_compair functions because filename sorting and token sorting are slightly different based the requirements. After we sort the array, we need to relink the member in the array into a new linked list.