

Gabriel Martinica  
ID: 40120255

## Modules Used:

NTLK:

As described by the assignment. This module is the one used for tokenizing the text, using a regex pattern to separate each document in the file and porter stemmer the tokens

Tabulate:

Module used to represent table information as described in subproject 3.

In this project I go into more details in my demo file.

Creating an inverted index was an interesting thing. I know there were some design choices which could be optimized by creating a more efficient algorithm or using a better data structure for the data given. I realize how document retrieval consumes so much memory and takes so much time. Maybe some algorithms that had a complexity of  $O(n^2)$  could be reduced to linear time by using a more efficient pointer system.

I think that the fact that my table results were similar to the ones in the book means I did a good job and went in the right path. I think as everything there is a compromise and deciding something instead of another leads to different types of problems that we are never aware of.