Assignment 7

The Great Firewall of Santa Cruz: Bloom Filters, Linked Lists, Binary Trees and Hash Tables

Mahyar Vahabi

Assignment purpose:

- In a similar form of a *1984* dystopian world, my job was to design a program that would censor certain words and translate them into “acceptable” form of speech. The process of translating oldspeak to newspeak took a series of functions and implementations of filtering out words.
- In this week’s project, we learned about and created different ADTs that would help us fix bad words and translate them into good words.
- A Bloom filter is a data structure designed to check whether an element is present in a list. It tells us that the element either definitely is not in the list or may be in the list.
- A hash table is a data structure that maps keys to values. It store translations from oldspeak to newspeak, but also as a way to store all prohibited oldspeak words without newspeak translations
- A binary search tree is a rooted binary tree data structure whose keys less than a node’s value goes under its left subtree, and keys greater than a node’s value go under its right subtree.

Errors:

- **Unfortunately**, for some reason, my code gets caught in a while loop when I read and write my files in banhammer. However, it returns the correct output for testing mixedspeak, goodspeak, and badspeak.
- For that reason, my statistics never print sadly, so I have no idea what to compare in my graphs, but I can do my best to explain my predictions of the results of what I think I would want to compare.

Analysis:

- The things I would have wished to compare, is how the hash table's average bst size/height and Average bst tree traversal are affected if the hash table and bloom filter sizes are changed as an experiment.
- In a very confusing attempt to understand the unknown data that I have, when the size of the hash table is larger, the average statistics *should* decrease. On another note, more memory is used, and less cache hits, which slows things down.
- If the size of the bloom filter is larger, so should be the average statistics.In addition, the more items added, the larger the probability of false positives.