

Project 4 Report

Ryan Everett

Describe the data structure(s) you used to store the information in synsets.txt. Why did you make this choice?

The synset data is stored in a Symbol with the form of Java's default HashMap. In the HashMap, the synset ID is used as the Key and the synset definition is used as the value in the form <Integer, String>. I use a Symbol Table because the WordNet needs to store a lot of data and retrieve a value quickly. The Symbol table allows quick retrieval in the matter of $\log(n)$ time.

Describe the data structure(s) you used to store the information in hypernyms.txt. Why did you make this choice?

The data structure is similar to the one used for synsets.txt. I use a Symbol Table with the synset ID being the Key and a Bag of hypernyms being the value. The bag of hypernyms allows the storage of multiple hypernyms as a single object. Using a symbol table allows for storing a large amount of objects and retrieving an object very quickly, roughly $\log(n)$ at the worst time.

Describe your algorithm to compute the SAP. What is the worst-case running time as a function of the structure of the graph (height, number of vertices, number of edges, etc)? Best case running time?

My algorithm for SAP utilizes BreadthFirstSearch to iterate through the vertices in the graph and compute the shortest path between its neighbors. It first constructs two BFS objects, and iterates through the vertices of the graph to find the shortest path. The worst case is if there are no paths and the best case is if the path has already been calculated. Worst Case is $O(E+V) * N$ and the Best Case is $O(1)$.