Ho Yin Cheng

**Find The Longest Word Report**

**Summary of the program:**

I develop a Trie class to store all the chars that was inserted from the data. Each char was stored in a Node object. I used ‘#’ to indicate the end of string. When I inserted a word into the Trie, the Trie will search a char from its children until a matched is found. Otherwise, it will add a new char node in its children list. If it found the “#” node at the first time, then it will add the original word with suffix to the queue and continue searching.  
  
After finishing the insert, it will start finding the longest word made of other words.

At this point, the algorithm will not do any modification of the Trie. We start polling the pair(original, suffix) from the queue, and do the same procedure as above instead of adding node, if there is no matched char during searching. If there is no matched char, we will discard this case. If all the char is matched, it will insert the original word to the solution list. The algorithm will finish until the queue is empty

After that, we will sort the solution list in descending order by using string’s length as a comparator. Finally, the longest word made of other words will the 1st element from the list.

**Solution:**

The total inserted words : 173530

The first longest word: ethylenediaminetetraacetates 28letters

The second longest word: electroencephalographically 27letters

The first shortest word: lion 4 letters

The second shortest word: lipa 4 letters

The total longest words count: 97107

**Side Test case:**

Add a to z into the provide large data set in sorted order.

This case is the easiest case to analyze for the large data set.

Total words: 173554

The first longest word: ethylenediaminetetraacetates 28letters

The second longest word: ethylenediaminetetraacetate 27letters

The first shortest word: yo 2 letters

The second shortest word: ya 2 letters

The Total count: 173528

The children size is 26

**Complexity Analysis:**

Let n to be length of a word

The complexity will be O(), because the worst case of a word will be z\*.

Then we will divided the words into (n(n+1))/2 cases and add them to the queue. Then, we do searching to find matched char in Node class. The search function’s complexity is O(27). Since the total alphabets will be 26 plus ‘#’, the total children node will be 27.

Since the search function is inside the findthelongestword function, the complexity will be . After simplification, it will be O().

For the addwords function, it contains search function, then the complexity will be O(27n). Since dominate 27n, the complexity of the whole program will be O()