CS771 Assignment 2

Group: Chaggi
Abhinav Kumar Singh (200017)
Harshit Gupta (200429)
Krishan Kumar (200521)
Donga Harika (190321)

1 Implementation

We have used the following algorithm for pruning:

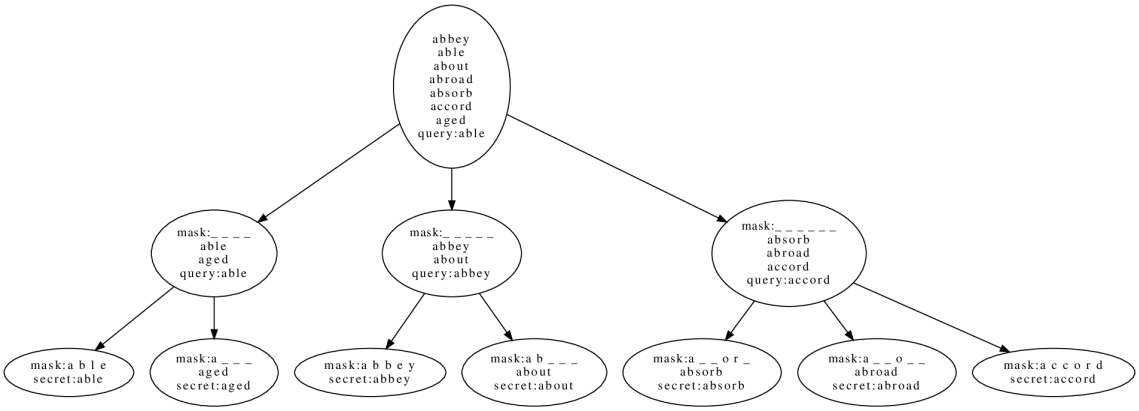
```
1) If the node is root:
        a) Select a random query from the list of all words
        b) Mask the query with each word
        c) Split the dictionary based on the word lengths.
        d) Create children as per the number of splits.
           Since we are given that the word lengths vary from 4
           to 15, the root node has a total of 12 children.
        e) If there exists a split with only one entry, make
           it a leaf.

2) Else if the node is not leaf:
        a) Select a random word from the list of words associated
           with the current node.
        b) Mask it with all the words in our current list of
           words, and for each distinct mask, create a new
           split.
        c) Create children as per the number of splits.
        d) If there exists a split with only one entry, make
           it a leaf.
3) Else if the node is leaf
        The node contains only a single word which is the
        secret word.
```

For example, for the following dictionary,

```
abbey
able
about
absorb
accord
aged
```

A possible decision tree is:



2 Performance Statistics

Upon running the script on the given google collab notebook and with "dict_secret" file:

Metric	**Value**
Average Time To Finish Training	0.15 seconds
On-Disk size	1010940 bytes (1 mb)
Average Number of Queries	4.95
Win Rate	1.0

We also tried running the code against a dictionary of size 100000 which gave the following results:

Metric	**Value**
Average Time To Finish Training	5.21 seconds
On-Disk size	20365099 bytes (19.5 mb)
Average Number of Queries	5.366
Win Rate	0.998

The win rate was not 1.0 this time because a few words took more than 15 queries. It was observed that even after increasing "max queries" upto 100, the accuracy did not reach 100% though a slight increase was observed.

3 References

[1] https://www.poirrier.ca/notes/wordle/