

Homework 6

problem1A

```
[yg336@wind ~/inf503/Homework6]$ ./main problem1A SARS.txt 50000
The number of arguments passed: 4
The first argument is: ./main
The second argument is: problem1A
The third argument is: SARS.txt
The forth argument is: 50000
1
1
0
0
1
```

The fragment 1,2 and 5 can be found in the genome sequence.

problem1B

```
[yg336@wind ~/inf503/Homework6]$ ./main problem1B SARS.txt 5000
The number of arguments passed: 4
The first argument is: ./main
The second argument is: problem1B
The third argument is: SARS.txt
The forth argument is: 5000
found 5000 in 5000 sequences
Total Nodes: 446931003
[yg336@wind ~/inf503/Homework6]$ ./main problem1B SARS.txt 50000
The number of arguments passed: 4
The first argument is: ./main
The second argument is: problem1B
The third argument is: SARS.txt
The forth argument is: 50000
found 50000 in 50000 sequences
Total Nodes: 446931003
[yg336@wind ~/inf503/Homework6]$ ./main problem1B SARS.txt 100000
The number of arguments passed: 4
The first argument is: ./main
The second argument is: problem1B
The third argument is: SARS.txt
The forth argument is: 100000
found 100000 in 100000 sequences
Total Nodes: 446931003
[yg336@wind ~/inf503/Homework6]$
```

There are 446931003 nodes in the trie.

We can find all the randomly generated 36-mers in the genome sequence. Because all the 36-mers are generated from the genome sequence, we can at least find one perfect match.

The search time for K queries of size n and subject of size G is: $O(K*n)$