Title: SPAM CHECKER USING TRIE

Submitted By:

1. Ananya Chopra (20103127)
2. Drishtant Maurya (20103140)
3. Nandini Agarwal (20103144)
4. Vivek Arora (20103145)

Submitted To:

Ms. Sherry Garg

ACKNOWLEDGEMENT

We would like to express my special thanks of gratitude to my teacher Ms. Sherry Garg, Our Lab Teacher and who gave us the golden opportunity to do this wonderful project on the topic " Spam Checker", which also helped us in doing a lot of Research and We came to know about so many new things and we are really thankful to them. Secondly, we would also like to thank my parents and friends who helped us a lot in finalizing this project within the limited time frame.

Contents

Introduction

Spam emails are one of the most irritating things, and it eats up a lot of space, so to deal with that, we have a spam checker.
Spam checker checks for some specific words in a file, which is considered spam when exceeding a certain number of repetitions in a file

The project checks for spam mails. There is a file where you can add the words/sentences which are or can be categorized as spam, the code basically searches for these spam words in the content sent for checking, if the content contains those words, it is considered to be a spam, else given a clean chit

The above task is achieved using trie., it searches Trie is an efficient information re***trie***val data structure. Using Trie, search complexities can be brought to optimal limit (key length). If we store keys in binary search tree, a well-balanced BST will need time proportional to **M * log N**, where M is maximum string length and N is number of keys in tree. Using Trie, we can search the key in O(M) time.

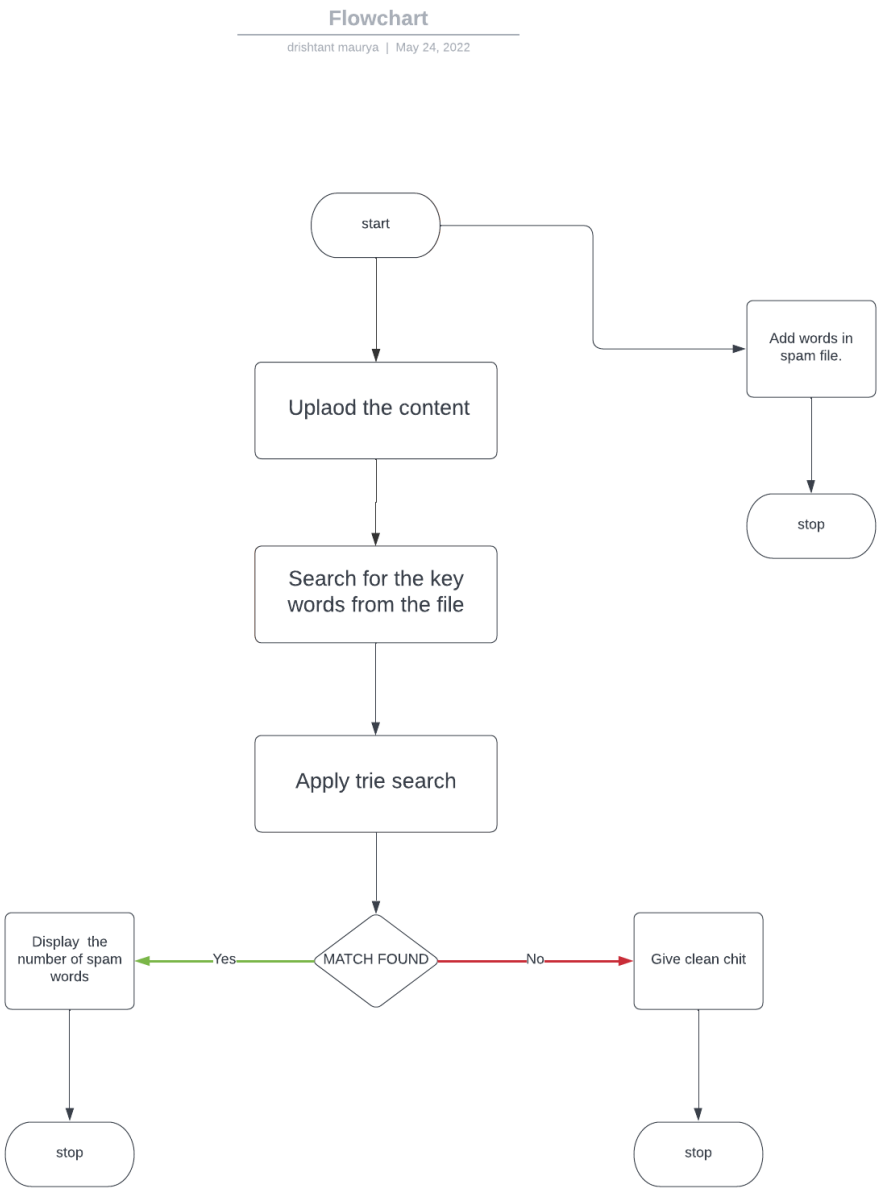
Data Structures

- Trie
- Vector
- Array
- string

Algorithms Used

- **Trie Implementation**

Flow Chart


Flowchart
drishtant maurya | May 24, 2022
start
Add words in spam file.
Uplaod the content
stop
Search for the key words from the file
Apply trie search
Display the number of spam words
Yes
MATCH FOUND
No
Give clean chit
stop
stop

Software Used

1. Visual Studio Code:

Visual Studio Code is a free source-code editor made by Microsoft for Windows, Linux and macOS.

2. GitBash:

Git Bash is a source control management system for Windows. It allows users to type Git commands that make source code management easier through versioning and commit history.

3. MinGW:

It formerly mingw32, is free and open-source software development environment to create Microsoft Windows application

4. GitHub:

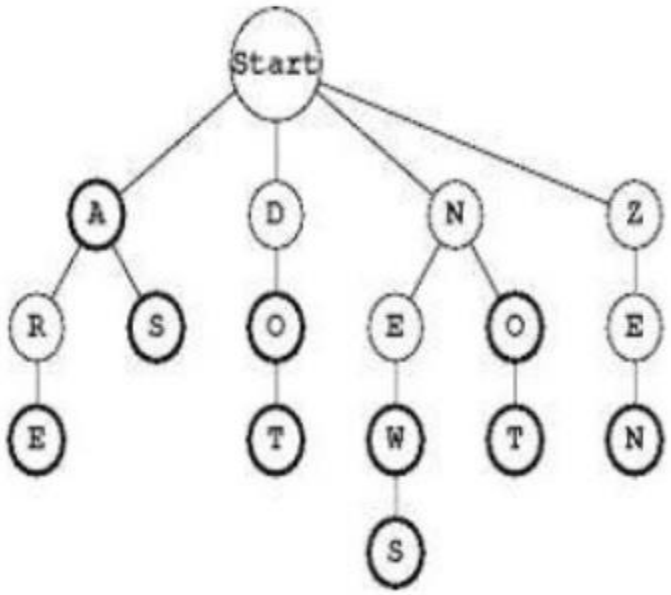
It makes it easy to contribute to your group work projects,
It helps in documentation; it tracks changes in your code across versions

ALOGORITHM USED FOR SPAM CHECKER:

Trie is an efficient information re*trie*val data structure. Using Trie, search complexities can be brought to optimal limit (key length). If we store keys in binary search tree, a well-balanced BST will need time proportional to **M * log N**, where M is maximum string length and N is number of keys in tree. Using Trie, we can search the key in O(M) time.

PSEUDO CODE:

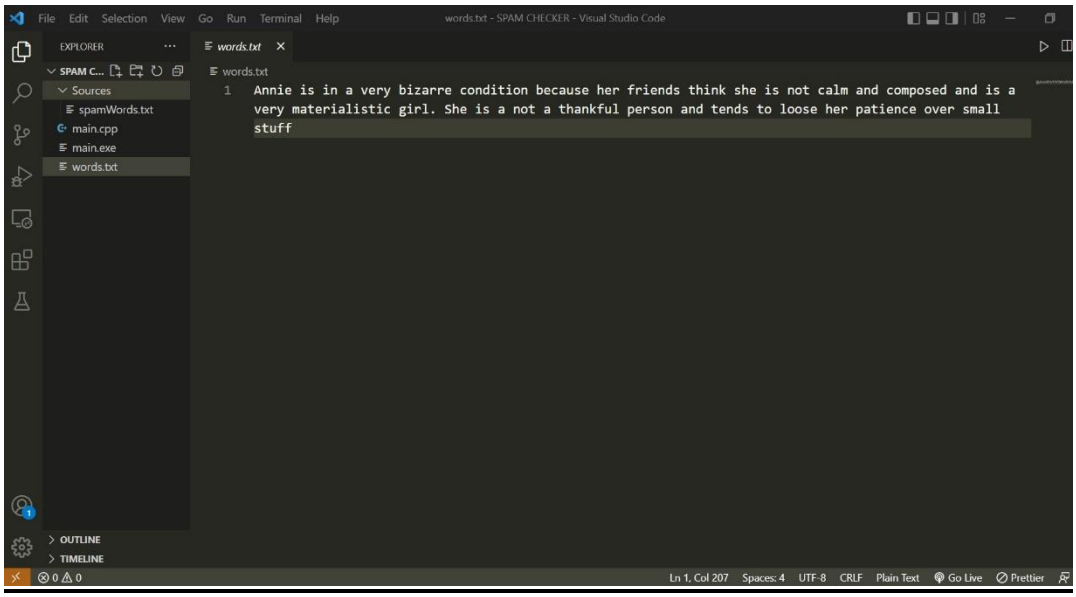
```
Boolean check (String s)
{
        for (every char in String s)
    {
        If (child node of current char is null)
          {
                Return false;
        }
    }
    Return true;
}
```



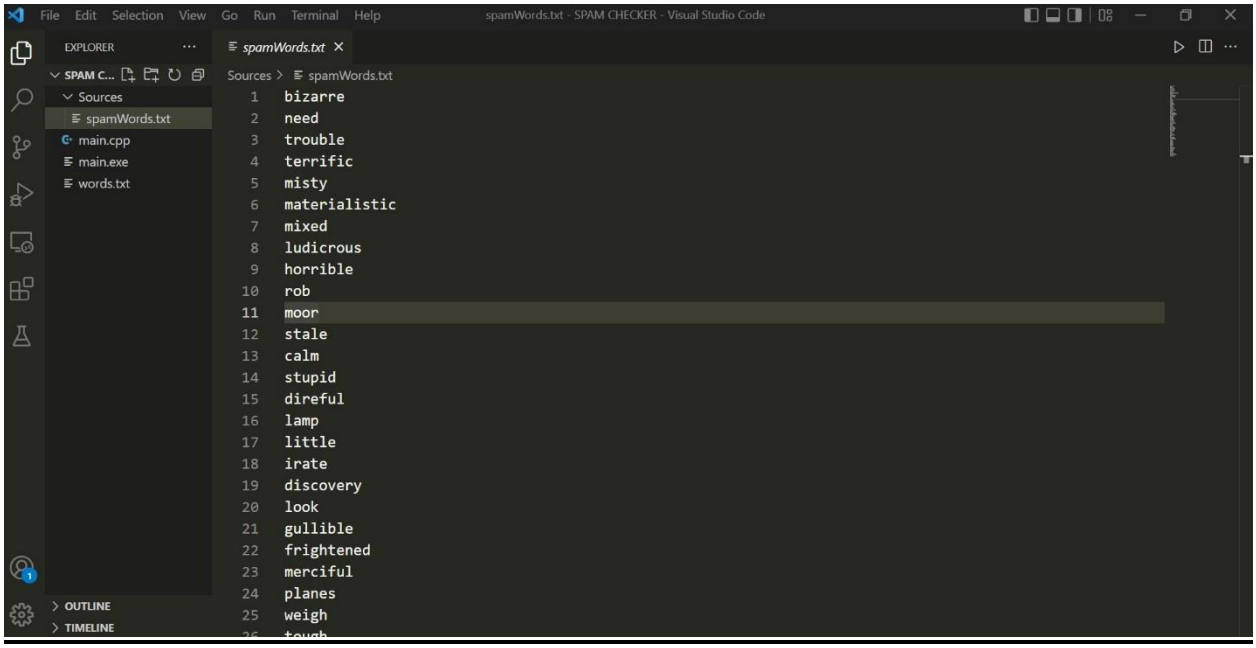
SAMPLE CODE:

```
bool search (string key)
   {
      TrieNode *pCrawl = root;

      for (int i = 0; i < key.length(); i++)
      {
         int index = key[i] - 'a';
         if (!pCrawl->children[index])
            return false;

         pCrawl = pCrawl->children[index];
      }

      return (pCrawl != NULL && pCrawl->isEndOfWord);
   }

} Spam;
```

INPUT FILE :



SPAM DATABASE:



OUTPUT: