

String Search

Presented by Angel Bencosme, Osman
Barrie, Isaac Gbaba, Tyler McNamara

START

Table of contents



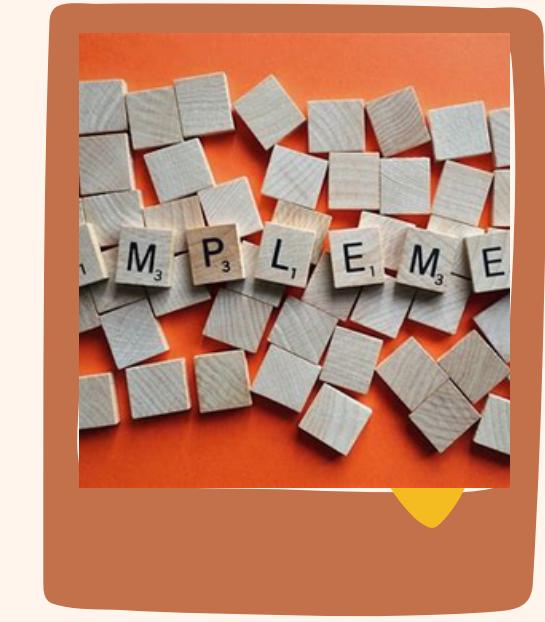
Introduction



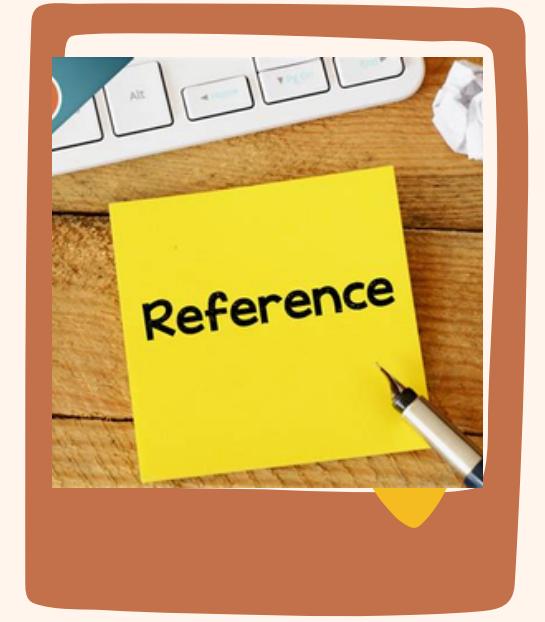
Project Overview



Methods,



Implementation,



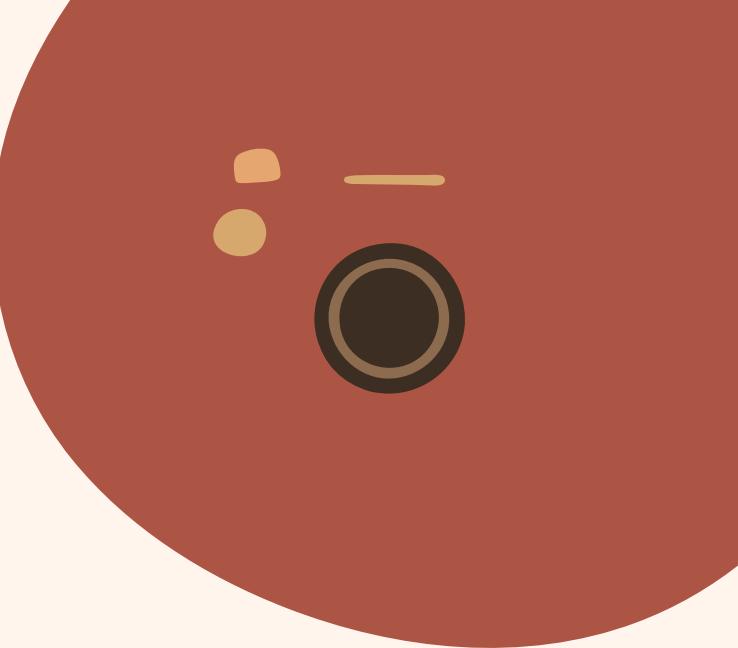
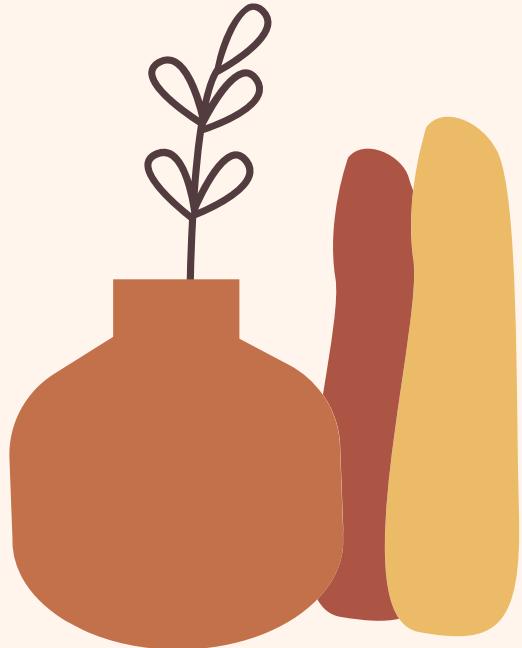
References



Contributions



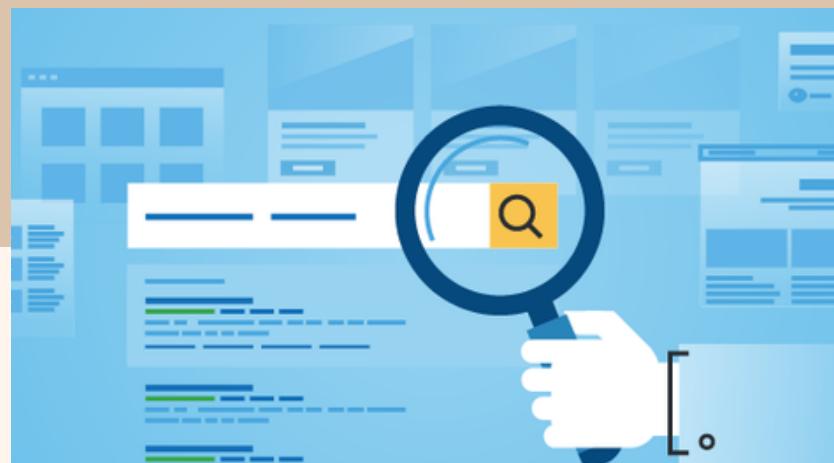
Conclusion.

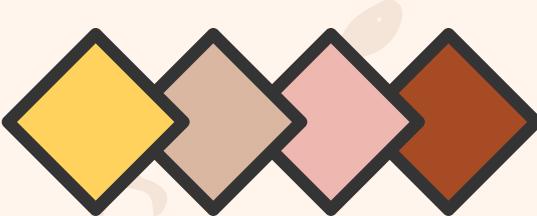


context & purpose



In computer science, string search algorithms stand as important tools in many applications in our daily lives. A string search algorithm is a method designed to find occurrences of a substring (a sequence of characters) within a larger string(or body of text.) In this project, we spotlight two algorithms: the Boyer-Moore and Knuth-Morris-Pratt algorithms. The Boyer-Moore algorithm stands out for its efficiency in searching large texts.





Introduction to our project

” what is it?

-a variation of control find

” why did we choose it?

We wanted to recreate a real life example of when the string search algorithms would be used in the best way for our audience to understand both the boyer moore and the Knuth morris pratt algorithms

” how does it mesh with our topic?

control find is a shortcut to find words, phrases, numbers in a document which can be grouped together under one word . Before anything is written to a file we convert it into a stringwhen you use control find you are essentially just searching for a string or a group of strings.

The Boyer-Moore-Horspool Algorithm (intuition)

string											pattern		
6	7	8	9	0	a	b	c	d	e	f	g	h	i
a	b	c	d										

'9' is not in the pattern, shift 4 places

'c' occurs as the penultimate character, shift 1 place

'd' matches 'd', check for match

Methods

Basic string search

Boyer-Moore

KMP (Knuth Morris Pratt) search



Basic string search

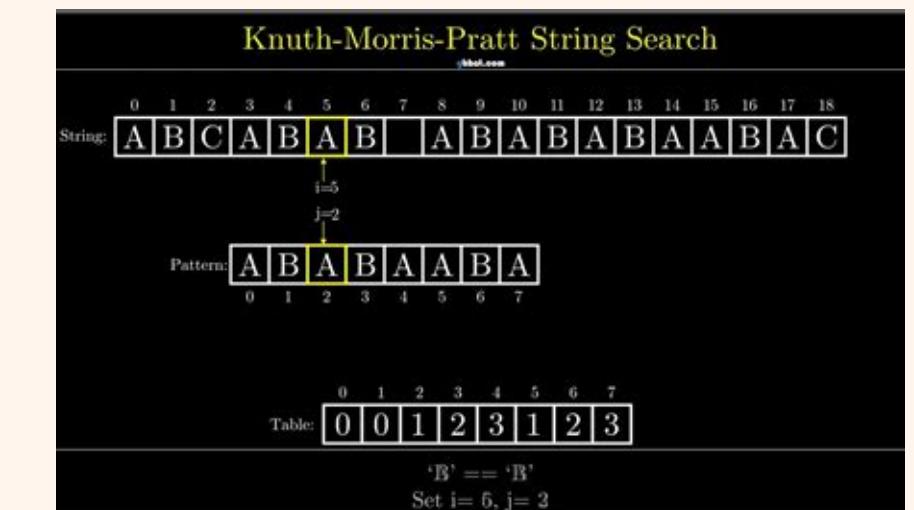
a brute force search that moves through every element of the string until it finds an instance of the key.

Boyer-Moore

Similarly to the KMP method, the Boyer-Moore method preprocesses the string to avoid unnecessary comparisons. But Boyer-Moore does this in a more efficient way. It uses two different preprocessing methods to move the key through the string.

KMP (Knuth Morris Pratt) search

once we detect a mismatch in the process of comparing the key to the string, the algorithm jumps to the next instance of the previous pattern before it was mismatched. It does this to avoid useless comparisons and preprocessing the key into a “partial match” table



Implementation

For our implementation we decided to recreate the CTRL + F / Command + F function. We wanted to recreate a real life example of when the string search algorithms would be used. To create the CTRL + F function we implemented the different string search functions into one file. There are 3 different string search methods in the program , the basic function , the Knuth-Morris-Pratt function and the Boyer-Moore function.



Implementation

To create the CTRL + F function we implemented the different string search functions into one file. There are 3 different string search methods in the program , the basic function , the Knuth-Morris-Pratt function and the Boyer-Moore function. To run the program it takes in a file name and key in the command line. The main function then calls the “callFunction” function.

```
bool callFunctions(const std::string& s, const std::string& key , std::string function) {  
  
    if(function == "Basic" || function == "basic"){  
        basic(s,key);  
    }  
    else if(function == "KMP" || function == "kmp"){  
        KMPSearch(s,key);  
    }  
    else if(function == "Boyer-Moore" || "Boyer Moore"){  
        BoyerMoore(s,key);  
    }  
  
    else{  
        std::cout << "Invalid function!" << std::endl;  
    }  
}
```

Implementation

The function then asks the user to enter which string search function they want to use and prints out the amount of times the key was found and the row and column of the key. The functions did require some edits so that they would be useful for our implementation of CTRL + F.

- PS C:\Users\Angel\.vscode\project1\mc3.zip\output> g++ -std=c++11 .\controlf.cpp -o controlf
- PS C:\Users\Angel\.vscode\project1\mc3.zip\output> ./controlf .\email.txt "[date]"
Pick which function to use : Basic , KMP , or Boyer-Moore
"Boyer Moore"

```
[date] found 1 time(s)
[date] found at :
Row 6 Column 73
```

Implementation

```
✉ email.txt
1  Hi team,
2
3  The [meeting name] meeting scheduled for this [time frame] has been canceled.
4
5  I apologize for the late notice, but I know everyone will welcome the extra time
6  back in their day. We will reconvene at our regularly scheduled time on [date].
7
8  Thanks,
9  [Your name]
10 [Your title]
```

This is an example of what a test would look like. All the blank spots are filled with spaces and each line contains exactly 80 characters. If the code is compiled and the user is looking for “[date]” this is what the function is going to return .

Contributions!

Angel Bencosme

CTRL F, BASIC CODE, KMP CODE,
AND WROTE SECTION OF REPORT

Osman Barrie

GRAPHS, RECORD TIME AND
WROTE SECTION OF REPORT

Tyler McNamara

BOYER MOORE, WROTE
SECTION OF REPORT

Isaac Gbaba

SLIDES, WROTE SECTION OF
REPORT

Conclusion

The Boyer-Moore Algorithm is a more efficient and used algorithm compared to the KMP algorithm because of its speed and effectiveness when searching for dictionary words. The KMP algorithm is used for DNA matching and for text with a lot of patterns. Chrome can use something like the boyer-moore algorithm as it is generally the fastest algorithm for string searching as most likely when the user is going to use the control+f function when they are looking for dictionary words. But the company is Google, so they might even have their own variation of the string search function which might be even faster than the Boyer-Moore algorithm.

A	T	A	A	T	T	A	C	C	A	A	C	A	T	C
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

A	T	C
---	---	---



Resources

REFERENCES

Explore scientific, technical, and medical research on ScienceDirect. ScienceDirect.com | Science, health and medical journals, full text articles and books. (n.d.).

<https://www.sciencedirect.com/>

Written in c++ using Visual Studio Code

Github- <https://github.com/angelbencosme2022/CSC-212-Final/blob/main/README.md>

