

CS2263 Assignment 4

Student:

Name: Will Ross

Number: #3734692

Email: will.ross@unb.ca

Date: March 25th, 2024

Due Date: March 27th, 2024

Contents

- [CS2263 Assignment 4](#)
 - [Contents](#)
 - [Question 1](#)
 - [Answer Q1](#)
 - [Question 2](#)
 - [Source code Htag3](#)
 - [Question 3](#)
 - [Question 4](#)
 - [htag3.c output](#)
 - [Question 5](#)
 - [htag3.c D2L Output](#)

Question 1

In a few sentences describe the design of the htags3 program. Focus on the description of the algorithm used and of any data structures needed to complete the task.

Answer Q1

For my htag3s program, we will iterate through the input given by a file. It will then check if the input is a valid file. If the file is valid, it enters a while loop until we reach the end of the file `EOF`. Then checks if the `currChar` is equal to `<`. If true, we then get the next char in the file and see if it equals `!`, if true that means we are inside an HTML comment tag so we set the variable `inComment` to true. If not, that means we are in a tag and we set `newAllocat = 2` to reset its value and we `totalAllocat += 2`. Then set `inTag` to true and use `malloc` to add the new total allocated memory to the word array. Then we store the `currChar` into the `words` array for later. If the `currChar` didn't equal `<` then we check if `currChar` equals `>` meaning the end of a tag. If true, we check if we are in a comment if so set `inComment` to false. If we were not in a comment, we set `newAllocat += 2` and `totalAllocat += 2` and then use `realloc` to add the total allocated memory onto the word array then the `inTag` to false and set the last char on the `word` array to `\0` to symbolise the end of chars in the array. Then print the current memory allocation that was used on that tag. We then set `found` equal to zero. `found` represents if we find the same tag in the array. We then loop through all the characters stored in our array checking if `word` and `unique[i]` are the same. If they are the same we then set `found` to true and break from the for loop. One broke we check if `found` does not equal true. If found is false we string copy the contents of `word` to `unique[uniqueCount]`, then reset the `index` to zero. Then we check if the following are true `inTag && !inComment && isalpha(currChar)`, if so we then set `currChar` equal to `word[index]`. Once the while loop hits `EOF` we then pass the unique array of char and the count into a `printArr` function.

Question 2

Show the complete source code.

Source code Htag3

```
#include <stdio.h>
#include <stdlib.h>
#include <stdbool.h>
#include <ctype.h>
#include <string.h>

void printArr(char **arr, int size){

    printf("Unique words between HTML tags:\n");
    for (int i = 0; i < size; i++)
    {
        printf("<%s>\n", arr[i]);
        free(arr[i]);
    }
}

void readFile(char *filename)
{
    FILE *file = fopen(filename, "r");

    if (file == NULL)
    {
        printf("ERROR: Could not open file\n");
        return;
    }

    //If we are in a tag
    bool inTag = false;

    //If we are in a comment
    bool inComment = false;

    //The index for storing tags
    int index = 0;

    //The count of unique chars
    int uniqueCount = 0;

    //Array for all words stored
    char * word = NULL;

    //to add to the 2d array
    int newAllocat = 0;
```

```
//total allocation to the 2d array
int totalAllocat = 0;

//Array for unique tags
char ** unique = NULL;

//The current value read in
int currChar;

//If we found duplicate
bool found = false;

int open = '<';
int close = '>';

while ((currChar = fgetc(file)) != EOF)
{
    //check if we are at a tag start
    if(currChar == open){
        //get next char
        currChar = fgetc(file);

        //check if we are at a comment tag
        if(currChar == '!')
        {
            //in a comment
            inComment = true;
            continue;
        }
        else
        {
            newAllocat = 2;
            totalAllocat += 2;

            //in a tag and not a comment

            inTag = true;
            word = malloc(totalAllocat * sizeof(char));

            //add the current word to be returned
            word[index++] = currChar;
            continue;
        }
    }

    //if the current is at the end of a tag
    else if (currChar == close)
    {
        //if was in a comment change to false
        if(inComment)
```

```
{
    inComment = false;
    continue;
}

//if not in comment
else{

    //set the array to be 2 more for last char and close tag
    newAllocat += 2;
    totalAllocat += 2;
    word = realloc(word, totalAllocat * sizeof(char));

    // Set not in tag
    inTag = false;

    // Set the end of the word array
    word[index] = '\\0';

    printf("Current Allocation of current tag: %d\\n", newAllocat);

    // Check if the word is unique
    found = false;

    // Loop through the word array to make sure all unique
    for (int i = 0; i < uniqueCount; i++)
    {
        // Check if they are the same
        if (strcmp(word, unique[i]) == 0)
        {
            found = true;
            break;
        }
    }

    // If the word is unique
    if (!found)
    {
        // Add the word to the unique array
        unique = realloc(unique, (uniqueCount + 1) * sizeof(char *));

        // Allocate memory for the word
        unique[uniqueCount] = malloc(totalAllocat * sizeof(char));

        // Copy the word to the unique array
        strcpy(unique[uniqueCount++], word);
    }

    // Reset the index
    index = 0;

    // Free the word allocation
```

```
        free(word);
    }
}
// If we are in a tag and not in a comment
if (inTag && !inComment && isalpha(currChar))
{
    // Add a new heap memory allocation
    newAllocat++;
    totalAllocat++;

    // Reallocate the word array
    word = realloc(word, totalAllocat * sizeof(char));

    // Add the current char to the word array
    word[index++] = currChar;
}
}
printArr(unique, uniqueCount);
fclose(file);

printf("Total allocated memory: %d\n", totalAllocat);
free(unique);
}

int main(int argc, char **argv)
{
    if (argc != 2)
    {
        printf("ERROR: Provide a file to read.\n");
        return EXIT_FAILURE;
    }

    readFile(argv[1]);

    return EXIT_SUCCESS;
}
```

Question 3

Show the output from running htags3 program on the following input HTML code (store it in a file):

```
<html> <b> TEST1 </b> <b> TEST2 </b> </html>
```

```
~/wross$ gcc htags3.c -o htags3.exe
~/wross$ ./htags3.exe index.html
Current Allocation of current tag: 7
Current Allocation of current tag: 4
Current Allocation of current tag: 5
Current Allocation of current tag: 8
Unique words between HTML tags:
<html>
<b>
</b>
</html>
Total allocated memory: 24
~/wross$
```


Question 4

Show the output from running htags3 program on this HTML file, i.e. the very file describing this assignment: you need to download this file separately from D2L to your computer (do not download the entire D2L web page!). The file name is A4W2024.html.

htag3.c output

```
~/wross$ ./htags3.exe A4W2024.html
Current Allocation of current tag: 7
Current Allocation of current tag: 7
Current Allocation of current tag: 73
Current Allocation of current tag: 61
Current Allocation of current tag: 8
Current Allocation of current tag: 9
Current Allocation of current tag: 8
Current Allocation of current tag: 73
Current Allocation of current tag: 25
Current Allocation of current tag: 20
Current Allocation of current tag: 4
Current Allocation of current tag: 80
Current Allocation of current tag: 8
Current Allocation of current tag: 5
Current Allocation of current tag: 5
Current Allocation of current tag: 47
Current Allocation of current tag: 62
Current Allocation of current tag: 4
Current Allocation of current tag: 62
Current Allocation of current tag: 5
Current Allocation of current tag: 57
Current Allocation of current tag: 37
Current Allocation of current tag: 5
Current Allocation of current tag: 51
Current Allocation of current tag: 162
Current Allocation of current tag: 4
Current Allocation of current tag: 51
Current Allocation of current tag: 5
Current Allocation of current tag: 164
Current Allocation of current tag: 62
Current Allocation of current tag: 52
Current Allocation of current tag: 52
Current Allocation of current tag: 52
Current Allocation of current tag: 72
Current Allocation of current tag: 170
Current Allocation of current tag: 78
Current Allocation of current tag: 53
Current Allocation of current tag: 150
Current Allocation of current tag: 5
Current Allocation of current tag: 89
Current Allocation of current tag: 7
Current Allocation of current tag: 8
```

Current Allocation of current tag: 8

Unique words between HTML tags:

```
<html>
<head>
<meta http-equiv=Content-Type content="text/html; charset=windows-1252">
<meta name=Generator content="Microsoft Word 15 (filtered)">
<style>
</style>
</head>
<body lang=EN-CA link=blue vlink="#954F72" style='word-wrap:break-word'>
<div class=WordSection1>
<p class=MsoNormal>
<b>
<span lang=EN-US style='font-size:14.0pt;font-family:
"Times New Roman",serif'>
</span>
</b>
</p>
<p class=MsoNormal style='text-align:justify'>
<span lang=EN-US
style='font-family:"Times New Roman",serif'>
<u>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
</u>
<a href="https://www.educba.com/types-of-tags-in-html/">
<span style='text-decoration:
none'>
</a>
<span style='font-family:"Times New Roman",serif'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:0cm;margin-bottom:.0001pt;text-indent:0cm;line-height:12.0pt'>
<i>
<span
style='font-family:"Times New Roman",serif'>
</i>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:7.1pt;margin-bottom:.0001pt;text-indent:0cm;line-height:12.0pt'>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
<span lang=EN-US style='font-family:"Courier New"'>
<span
lang=EN-US style='font-family:"Courier New"'>
<span lang=EN-US
style='font-family:"Courier New"'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;line-height:12.0pt'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:43.1pt;margin-bottom:.0001pt;text-indent:-18.0pt;line-height:
12.0pt'>
<span
lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>
<span lang=EN-US style='font-family:
"Courier New"'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
12.0pt;margin-left:43.1pt;text-indent:-18.0pt;line-height:12.0pt'>
```

```
<br>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;text-indent:0cm;line-height:
12.0pt'>
</div>
</body>
</html>
Total allocated memory: 1967
~/wross$
```

Question 5

Show the output from running htags3 program on another HTML file, of your choice.

For the last test I used this websites source code [Assignment 1 Report](#)

htag3.c D2L Output

```
~/wross$ ./htags3.exe A1W2024.html
Current Allocation of current tag: 7
Current Allocation of current tag: 7
Current Allocation of current tag: 68
Current Allocation of current tag: 61
Current Allocation of current tag: 8
Current Allocation of current tag: 9
Current Allocation of current tag: 8
Current Allocation of current tag: 47
Current Allocation of current tag: 25
Current Allocation of current tag: 20
Current Allocation of current tag: 4
Current Allocation of current tag: 80
Current Allocation of current tag: 8
Current Allocation of current tag: 5
Current Allocation of current tag: 5
Current Allocation of current tag: 47
Current Allocation of current tag: 62
Current Allocation of current tag: 162
Current Allocation of current tag: 62
Current Allocation of current tag: 4
Current Allocation of current tag: 5
Current Allocation of current tag: 164
Current Allocation of current tag: 5
Current Allocation of current tag: 4
Current Allocation of current tag: 51
Current Allocation of current tag: 5
Current Allocation of current tag: 51
Current Allocation of current tag: 150
Current Allocation of current tag: 67
Current Allocation of current tag: 41
Current Allocation of current tag: 170
Current Allocation of current tag: 67
Current Allocation of current tag: 41
Current Allocation of current tag: 9
Current Allocation of current tag: 33
Current Allocation of current tag: 10
Current Allocation of current tag: 42
Current Allocation of current tag: 33
Current Allocation of current tag: 170
Current Allocation of current tag: 89
Current Allocation of current tag: 62
Current Allocation of current tag: 25
```

```

Current Allocation of current tag: 7
Current Allocation of current tag: 8
Current Allocation of current tag: 8
Unique words between HTML tags:
<html>
<head>
<meta http-equiv=Content-Type content="text/html; charset=unicode">
<meta name=Generator content="Microsoft Word 15 (filtered)">
<style>
</style>
</head>
<body lang=EN-CA style='word-wrap:break-word'>
<div class=WordSection1>
<p class=MsoNormal>
<b>
<span lang=EN-US style='font-size:14.0pt;font-family:
"Times New Roman",serif'>
</span>
</b>
</p>
<p class=MsoNormal style='text-align:justify'>
<span lang=EN-US
style='font-family:"Times New Roman",serif'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:0cm;margin-bottom:.0001pt;text-indent:0cm;line-height:12.0pt'>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
<u>
</u>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:7.1pt;margin-bottom:.0001pt;text-indent:0cm;line-height:12.0pt'>
<br>
<i>
<span style='font-family:"Times New Roman",serif'>
</i>
<span
style='font-family:"Times New Roman",serif'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
12.0pt;margin-left:43.1pt;text-indent:-18.0pt;line-height:12.0pt'>
<span
style='font-size:7.0pt;font-family:"Times New Roman",serif'>
<span style='font-family:"Courier New"'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:43.1pt;margin-bottom:.0001pt;text-indent:-18.0pt;line-height:
12.0pt'>
<span style='font-size:7.0pt;font-family:"Times New Roman",serif'>
<span
style='font-family:"Courier New"'>
<strong>
<span style='font-family:Times'>
</strong>
<span style='font-family:
"Courier New"'>
<span

```

```
style='font-family:Times'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:79.1pt;margin-bottom:.0001pt;text-indent:-18.0pt;line-height:
12.0pt'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;text-indent:0cm;line-height:
12.0pt'>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
<span style='color:red'>
</div>
</body>
</html>
Total allocated memory: 2016
~/wross$
```