

# CS2263 Assignment 3

---

**Student:**

Name: Will Ross

Number: #3734692

Email: [will.ross@unb.ca](mailto:will.ross@unb.ca)

**Date:** February 16th, 2024

**Due Date:** February 21st, 2024

# Contents

- [CS2263 Assignment 3](#)
  - [Contents](#)
  - [Question 1](#)
    - [Answer Q1](#)
  - [Question 2](#)
    - [Source code Htag](#)
  - [Question 3](#)
    - [Answer Q3](#)
  - [Question 4](#)
    - [Source code Htext1](#)
  - [Question 5](#)
    - [Input](#)
    - [Htext1.c output](#)
    - [Htag1.c output](#)
  - [Question 6](#)
    - [Htext1.c D2L output](#)
    - [Htags.c D2L Output](#)
  - [Question 7](#)
    - [Htext1.c Output](#)
    - [Htag.c Output](#)

## Question 1

In a few sentences describe the design of the htags1 program. Focus on the description of the algorithm used and of any data structures needed to complete the task.

### Answer Q1

For my Htag program we will iterate through the input given by a file, it will then check if the input is a `<` if its true it will then check for `!` if that's true that means we are in a html comment. Doing this we get the current characters and then next char due to whitespace and invisible characters if its true we set `inComment` to `true`, if not we print a open tag `<` followed by the two chars we checked for. Once we enter the while loop for reading inside a tag we check if we are in a comment we don't use `putchar()`. After we leave the while loop we check if we in a comment if not we will print the closing tag `>` followed by a new line character `\n`.

## Question 2

Show the complete source code for htags1.

Source code Htag

```
#include <stdio.h>
#include <string.h>
#include <stdbool.h>
void printArr(char a[], int n){

    for(int i = 0; i < (n-1); i++){

        putchar(a[i]);

    }

    putchar('\n');
}
void readTags()
{

    int currChar;
    int first = '<';
    int second = '>';
    int newline = '\n';
    bool inComment = false;
    int check;

    while((currChar = getchar()) != EOF)
    {
        inComment = false;
        if(currChar == '<')
        {

            check = getchar();
            check = getchar();
            if(check == '!')
            {

                inComment = true;
            }
            else
            {
                putchar(first);
                putchar(check);
            }

            while((currChar = getchar()) != '>' && currChar != EOF){
                if(!inComment)
                {
```

```
                putchar(currChar);
                continue;
            }
        }
        if(!inComment)
        {
            putchar(second);
            putchar(newline);
        }
    }
}

int main()
{
    char start[] = "Start of output: ";
    char end[] = "End of output: ";

    printArr(start, strlen(start));
    readTags();
    printArr(end, strlen(end));

    return 0;
}
```

## Question 3

In a few sentences describe the design of the htext1 program. Focus on the description of the algorithm used and of any data structures needed to complete the task.

### Answer Q3

For my Htext1 program we will iterate through the input given by a file, it will then check if the input is a `>` if its true it will then we enter the while loop for plain text outside the html tags. We check if the currChar doesnt equal a starting tag `<`. After we leave the while loop we will print a new space character `_`.

## Question 4

Show the complete source code for htext1.

Source code Htext1

```
#include <stdio.h>
#include <string.h>

void printArr(char a[], int n)
{
    //just a for loop to go through char array
    for(int i = 0; i < (n-1); i++)
    {
        putchar(a[i]);
    }

    //new line after complete
    putchar('\n');
}

void readText()
{
    int currChar;
    int index = 0;
    int space = ' ';

    while((currChar = getchar()) != EOF)
    {
        if(currChar == '>')
        {
            while((currChar = getchar()) != '<' && currChar != EOF)
            {
                putchar(currChar);
            }
            putchar(space);
        }
    }
}

int main()
{
    char start[] = "Start of output:";
    char end[] = "End of output:";

    printArr(start, strlen(start));
    readText();
    printArr(end, strlen(end));

    return 0;
}
```

```
}
```



## Question 5

Show the output from running htags1 and htext1 programs on the following input HTML code (store it in a file and use input redirection to read):

Input

```
<html> <b> TEST </b> </html>
```

Htext1.c output

```
Start of output  
TEST
```

Htag1.c output

```
Start of output:  
<html>  
<b>  
</b>  
</html>  
End of output:
```

## Question 6

Show the output from running `htags1` and `htext1` 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 `A2W2024.html`.

### Htext1.c D2L output

Start of output

CS 2263

Assignment 2

Due: Feb. 21, 2024 (before midnight)

---

Assignments  
are to be completed individually ( not as a group work).

&nbsp;

Sample  
reference: <https://www.educba.com/types-of-tags-in-html/>

&nbsp;

Processing strings

Develop and test two  
utility programs called `htags1` and `htext1` to process an HTML file. The  
`htags1` program should &nbsp;print all  
HTML tags, except the HTML comment tags `<!-- -->`, one per line of  
output. The `htext1` program should print  
all plain text extracted from the source HTML file, i.e. just the text, no  
tags. The utility programs should read from standard input and may use the  
input redirection to access  
a file. (This means that your program code should read from standard input,  
and  
NOT perform any file operations, like `fopen()` etc. You may capture the  
output  
from your program using output redirection, or you may just output to the  
terminal window and print the screen, your choice.)

Assume an  
HTML tag is any text enclosed between `<` and `>`, including these  
delimiter  
characters. The end tags, like `</p>`, are valid tags. For example, in

sample input HTML file containing:

we find the following tags:

and the plain text in this sample input is:

When outputting the plain text you may keep new line characters present in the input, or you may replace them with space characters.

The data to your program should be read from standard input (use the input redirection to read from the HTML file) using the `getchar()` function. The output should be done using the `putchar()` function and/or using the `printArr()` function from Assignment 1. You must use the `printArr()` function at least once in each utility program developed for this assignment. The Assignment Report should consist of the following parts/exercises:

1.      In a few



&nbsp;

## Htags.c D2L Output

Start of output:

```
<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 link=blue vlink=purple 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>
<b>
<span lang=EN-US style='font-size:14.0pt;font-family:
"Times New Roman",serif'>
</span>
</b>
</p>
<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'>
</span>
</p>
<p class=MsoNormal>
<u>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
<b>
</b>
</span>
</u>
</p>
<p class=MsoNormal>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
```

```
</span>
</p>
<p class=MsoNormal>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
<a href="https://www.educba.com/types-of-tags-in-html/">
<span
style='text-decoration:none'>
</span>
</a>
</span>
</p>
<p class=MsoNormal>
<span style='font-family:"Times New Roman",serif'>
</span>
</p>
<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'>
</span>
</i>
</p>
<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>
<span
lang=EN-US style='font-family:"Courier New"'>
</span>
<span lang=EN-US
style='font-family:"Times New Roman",serif'>
<b>
</b>
</span>
<span lang=EN-US style='font-family:"Courier New"'>
</span>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
<a
href="https://linuxcommand.org/lc3_lts0070.php">
</a>
</span>
</p>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-
bottom:
0cm;margin-left:7.1pt;margin-bottom:.0001pt;text-indent:18.0pt;line-height:
12.0pt'>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
```

/

```
style='font-family:"Courier New"'>
</span>
</b>
<span lang=EN-US
style='font-family:"Times New Roman",serif'>
</span>
<b>
<span lang=EN-US style='font-family:"Courier New"'>
</span>
</b>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
<b>
<span lang=EN-US style='font-family:"Courier New"'>
</span>
</b>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
<span lang=EN-US style='font-family:"Courier New"'>
</span>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
<span lang=EN-US style='font-family:"Courier New"'>
</span>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
</p>
<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-family:"Times New Roman",serif'>
</span>
<span
lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
</p>
<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-family:"Times New Roman",serif'>
</span>
<span
lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
<span style='font-family:"Times New Roman",serif'>
<u>
</u>
</span>
</p>
```



`<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-family:"Times New Roman",serif'>`  
`</span>`  
`<span`  
`lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>`  
`</span>`  
`<span lang=EN-US style='font-family:"Times New Roman",serif'>`  
`</span>`  
`</p>`  

`<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-family:"Times New Roman",serif'>`  
`</span>`  
`<span`  
`lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>`  
`</span>`  
`<span style='font-family:"Times New Roman",serif'>`  
`<u>`  
`</u>`  
`</span>`  
`</p>`  

`<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`  
`lang=EN-US style='font-family:"Times New Roman",serif'>`  
`</span>`  
`<span`  
`lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>`  
`</span>`  
`<span style='font-family:"Times New Roman",serif'>`  
`<br>`  
`<br>`  
`</span>`  
`</p>`  

`<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-family:"Times New Roman",serif'>`  
`</span>`  
`<span`  
`lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>`  
`</span>`  
`<span style='font-family:"Times New Roman",serif'>`  
`</span>`

```
<span lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
</p>
<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-family:"Times New Roman",serif'>
</span>
<span
lang=EN-US style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
<span style='font-family:"Times New Roman",serif'>
</span>
<span
lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
</p>
<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>
</p>
<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>
</p>
<p class=MsoNormal>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
</p>
</div>
</body>
</html>
End of output:
```

Show the output from running htags1 and htext1 program on another HTML file, of your choice. For the last test I used this websites source code [Assignment 1 Report](#)

## /





```
</p>
<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>
<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'>
</span>
</p>
<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>
<b>
</b>
</u>
</span>
</p>
<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'>
<br>
</span>
<i>
<span style='font-family:"Times New Roman",serif'>
</span>
</i>
</p>
<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
style='font-family:"Times New Roman",serif'>
<i>
</i>
<i>
```

```
</i>
</span>
</p>
<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
style='font-family:"Times New Roman",serif'>
</span>
</p>
<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>
<i>
</i>
<br>
<br>
<br>
<span style='font-family:"Courier New"'>
</span>
</p>
<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>
<i>
</i>
<span style='font-family:"Courier New"'>
</span>
<i>
</i>
<i>
</i>
<br>
<u>
</u>
<br>
<br>
<span style='font-family:"Courier New"'>
</span>
<span
style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
</p>
<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'>
</p>
<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>
<i>
</i>
<br>
<span
style='font-family:"Courier New"'>
</span>
<span
style='font-family:"Courier New"'>
</span>
<strong>
<span style='font-family:Times'>
</span>
</strong>
<span style='font-family:"Courier New"'>
</span>
<br>
<br>
<span style='font-family:"Courier New"'>
</span>
<span
style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
</p>
<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'>
</p>
<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>
<i>
</i>
<br>
<span
style='font-family:"Courier New"'>
</span>
<span style='font-family:"Courier New"'>
</span>
<span
style='font-family:"Times New Roman",serif'>
```



```
</span>
<span style='font-family:
"Courier New"'>
</span>
<strong>
<span
style='font-family:Times'>
</span>
</strong>
<span
style='font-family:"Courier New"'>
</span>
<span style='font-family:"Courier New"'>
</span>
<span style='font-family:"Courier New"'>
</span>
<span style='font-family:"Courier New"'>
</span>
<br>
<br>
<span style='font-family:"Courier New"'>
</span>
<span style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
</p>
<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'>
</p>
<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>
<i>
</i>
<br>
<i>
<u>
</u>
</i>
<i>
<u>
</u>
</i>
</p>
<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'>
<span style='font-size:7.0pt;font-family:"Times New Roman",serif'>
```

```
</span>
</p>
<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'>
<span style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
<i>
</i>
</p>
<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'>
<span style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
</p>
<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'>
<span style='font-size:7.0pt;font-family:"Times New Roman",serif'>
</span>
</p>
<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'>
</span>
</p>
<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'>
<i>
</i>
</span>
</p>
<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'>
<b>
<u>
<span style='color:red'>
</span>
</u>
</b>
```

```
</span>
</p>
<p class=MsoNormal>
<span lang=EN-US style='font-family:"Times New Roman",serif'>
</span>
</p>
</div>
</body>
</html>
```

End of output: