

CS2263 Assignment 5

Student:

Name: Will Ross

Number: #3734692

Email: will.ross@unb.ca

Date: April 9th, 2024

Due Date: April 11th, 2024

Contents

- [CS2263 Assignment 5](#)
 - [Contents](#)
 - [Question 1](#)
 - [Answer Q1](#)
 - [Question 2](#)
 - [Source Code list.h](#)
 - [Source Code list.c](#)
 - [Source code Htag4](#)
 - [Question 3](#)
 - [Htags4.c without -m output](#)
 - [Question 4](#)
 - [htag4.c -m output](#)
 - [Question 5](#)
 - [Output without -m A4W2024](#)
 - [Output with -m A4W2024](#)
 - [Question 6](#)
 - [Output without -m A5W2024](#)
 - [Output with -m A5W2024](#)

Question 1

In a few sentences describe the design of the htags4 program. Focus on the description of the algorithm used and of any data structures needed to complete the task.

Answer Q1

For my htag4s program, we will have 3 files, list.h, list.c, htags4.c. list.h is my header file with all my functions for our linkedlist functionality to be called, list.c is my c file used for all of my linked list algorithms for the htags4 program. htags4 is the main file of my program. In list.c i have 5 functions, `*Node_constructor` used for constructing nodes from char inputs from htags. `*list_add` is a function used to add a node onto my linked list of htags htags. `List_search` is a function used to find if the input tag already exists somewhere in the linked list so there is no duplicates. `List_print` is used to print our tag inputs, and `List_free` is used for freeing memory on our linked list.

In htags.c I made quite a bit of changes compared to htags3.c due to the linked list data structure. We read in the users input from the terminal line now checking if the user wants the memory option. Then passing the file and the memory option into a readfile function. Once then we go line to line checking the input for tags. Once we read `<` we add the current char to the word array and set currentAllocat to 1 for memory allocation. Cycling through the while loop if we read a `>` and inTag is true, we add current char and `\0` to the word array and then construct a new Node if we arent in a comment and the tag isnt in the linked list already. Then we get the size of the node and add it to our current memory allocation and add total memory allocation from the current variable. If in tag is true we will add the tags onto the current word array and make sure incomment isnt true. Then we close the file scanner, print the linked list, free the memory and print the total memory if the user wanted to see the memory of the program.

Question 2

Show the complete source code.

Source Code list.h

```
// header file
#ifndef LIST_H
#define LIST_H

#include <stddef.h>
#include <stdio.h>

typedef struct listnode {

    struct listnode *next;
    char *tag;
} Node;

// Constructor
Node *Node_construct(char *tag);

// Add a new node on the linked list
Node *List_add(Node *head, Node *n);

// search for a node in the linked list
int List_search(Node *head, char *tag);
// int List_search(Node * head, Node * n);

// print the linked list
void List_print(Node *head);

// free memory in the list
void List_free(Node *head);

#endif
```

Source Code list.c

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

```
Node *Node_construct(char *tag)
{
    Node *n = malloc(sizeof(Node));

    if (n == NULL)
    {
        return NULL;
    }
    size_t tagSize = sizeof(strlen(tag) + 1);

    n -> tag = strdup(tag);

    if ((n->tag) == NULL)
    {
        free(n);
        return NULL;
    }
    n->next = NULL;

    return n;
}

// this just adds at the end due adding tags at the end to be read later
Node *List_add(Node *head, Node *n)
{
    if (head == NULL)
    {
        return n;
    }
    Node *current = head;

    while (current->next != NULL)
    {
        current = current->next;
    }

    current->next = n;

    return head;
}

// int List_search(Node * head, Node * n)
int List_search(Node *head, char *tag)
{
    Node *current = head;

    while (current != NULL)
    {
        if (strcmp(current->tag, tag) == 0)
        {
            return true;
        }
    }
}
```

```

        current = current->next;

    }

    return false;
}

void List_print(Node *head)
{
    Node *current = head;

    while (current != NULL)
    {
        printf("%s\n", (current->tag));
        current = current->next;
    }
}

void List_free(Node *head)
{
    Node *current = head;

    while (current != NULL)
    {
        Node *temp = current;
        current = current->next;
        free(temp->tag);
        free(temp);
    }
}

```

Source code Htag4

```

#include "list.h"
#include <ctype.h>
#include <stdbool.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

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

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

```

```

    return;
}

bool inTag = false, inComment = false;
int index = 0, currChar, prevChar, open = '<', close = '>';

char *word = malloc(1);
Node *head = NULL;

int totalAllocat = 0;
int tagCount = 0;

while ((currChar = fgetc(file)) != EOF) {

    if (currChar == '!' && prevChar == '<')
    {
        inComment = true;
        inTag = false;
        index = 0;
        continue;
    }
    if (inComment && prevChar == '-' && currChar == '>')
    {
        inComment = false;
        continue;
    }
    if (inComment)
    {
        prevChar = currChar;
        continue;
    }

    if (!inTag && currChar == '<')
    {
        inTag = true;
        word = realloc(word, index + 2);
        word[index++] = currChar;
    }
    else if (inTag && currChar == '>')
    {
        word[index++] = currChar;
        word[index] = '\0';
        if (!List_search(head, word))
        {
            Node *n = Node_construct(word);
            head = List_add(head, n);
            int currentTagMem = sizeof(Node) + strlen(n->tag) + 1;
            totalAllocat += currentTagMem;
            if (memory)

```

```
        {
            printf("Current allocated memory: %d bytes\n", currentTagMem);
            tagCount++;
        }
    }
    word[0] = '\\0';
    index = 0;
    inTag = false;
}
else if (inTag)
{
    word = realloc(word, index + 2);
    word[index++] = currChar;
}

prevChar = currChar;
}

fclose(file);
List_print(head);
List_free(head);
free(word);

if (memory)
{
    printf("Total allocated memory: %d bytes\n", totalAllocat);
    printf("Number of tags found: %d\n", tagCount);
}
}

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

    bool memory = false;
    if (argc == 3 && strcmp(argv[2], "-m") == 0)
    {
        memory = true;
    }

    readFile(argv[1], memory);
    return EXIT_SUCCESS;
}
```


Question 3

Show the output from running htags4 program on the following input HTML code (store it in a file) without using the -m option:

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

Htags4.c without -m output

```
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5> ./htags4
index.html
<html>
<b>
</b>
</html>
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5>
```

Question 4

Show the output from running the htags4 program on the following input HTML code with the -m option specified as the command line parameter (i.e. \$ htags4 -m myInputFile.html):

htag4.c -m output

```
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5> ./htags4
index.html -m
Current allocated memory: 23 bytes
Current allocated memory: 20 bytes
Current allocated memory: 21 bytes
Current allocated memory: 24 bytes
<html>
<b>
</b>
</html>
Total allocated memory: 88 bytes
Number of tags found: 4
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5>
```

Question 5

Show the output from running the htags4 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.

Output without -m A4W2024

```
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5> ./htags4
A4W2024.html
<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>
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5>

```

Output with -m A4W2024

```

PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5> ./htags4
A4W2024.html -m
Current allocated memory: 23 bytes
Current allocated memory: 23 bytes
Current allocated memory: 89 bytes
Current allocated memory: 77 bytes
Current allocated memory: 24 bytes
Current allocated memory: 25 bytes
Current allocated memory: 24 bytes
Current allocated memory: 89 bytes
Current allocated memory: 41 bytes
Current allocated memory: 36 bytes
Current allocated memory: 20 bytes
Current allocated memory: 96 bytes
Current allocated memory: 24 bytes
Current allocated memory: 21 bytes
Current allocated memory: 21 bytes
Current allocated memory: 63 bytes
Current allocated memory: 78 bytes
Current allocated memory: 20 bytes
Current allocated memory: 78 bytes
Current allocated memory: 21 bytes
Current allocated memory: 73 bytes
Current allocated memory: 53 bytes
Current allocated memory: 21 bytes
Current allocated memory: 67 bytes
Current allocated memory: 178 bytes

```

Current allocated memory: 20 bytes
Current allocated memory: 67 bytes
Current allocated memory: 21 bytes
Current allocated memory: 180 bytes
Current allocated memory: 78 bytes
Current allocated memory: 68 bytes
Current allocated memory: 68 bytes
Current allocated memory: 68 bytes
Current allocated memory: 88 bytes
Current allocated memory: 186 bytes
Current allocated memory: 94 bytes
Current allocated memory: 69 bytes
Current allocated memory: 166 bytes
Current allocated memory: 21 bytes
Current allocated memory: 105 bytes
Current allocated memory: 23 bytes
Current allocated memory: 24 bytes
Current allocated memory: 24 bytes

```
<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: 2655 bytes
Number of tags found: 43
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5>

```

Question 6

Show the output from running the htags4 program on another HTML file, of your choice. I chose Assignment 5 as the test input as we were asked to test A4 not A5

Output without -m A5W2024

```

PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5> ./htags4
A5W2024.html
<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>
<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>
<i>
<span style='font-family:"Times New Roman",serif'>
</i>
<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"'>
<span lang=EN-US style='font-family:
"Courier New"'>
<span lang=EN-US style='font-family:
"Times New Roman",serif'>
<span
lang=EN-US style='font-family:Courier'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:61.1pt;margin-bottom:.0001pt;text-indent:-18.0pt;line-height:
12.0pt'>
<p class=MsoNormal style='margin-left:61.1pt'>
<span lang=EN-US
style='font-size:8.0pt;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'>
<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-family:"Times New Roman",serif'>
<span style='font-family:
"Courier New"'>
<span style='font-family:"Courier New"'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;text-indent:0cm;line-height:
12.0pt'>
</div>
</body>
</html>
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5>

```

Output with -m A5W2024

```
PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5> ./htags4
A5W2024.html -m
Current allocated memory: 23 bytes
Current allocated memory: 23 bytes
Current allocated memory: 89 bytes
Current allocated memory: 77 bytes
Current allocated memory: 24 bytes
Current allocated memory: 25 bytes
Current allocated memory: 24 bytes
Current allocated memory: 89 bytes
Current allocated memory: 41 bytes
Current allocated memory: 36 bytes
Current allocated memory: 20 bytes
Current allocated memory: 96 bytes
Current allocated memory: 24 bytes
Current allocated memory: 21 bytes
Current allocated memory: 21 bytes
Current allocated memory: 63 bytes
Current allocated memory: 78 bytes
Current allocated memory: 20 bytes
Current allocated memory: 78 bytes
Current allocated memory: 21 bytes
Current allocated memory: 180 bytes
Current allocated memory: 78 bytes
Current allocated memory: 21 bytes
Current allocated memory: 20 bytes
Current allocated memory: 67 bytes
Current allocated memory: 21 bytes
Current allocated memory: 68 bytes
Current allocated memory: 68 bytes
Current allocated memory: 68 bytes
Current allocated memory: 69 bytes
Current allocated memory: 79 bytes
Current allocated memory: 62 bytes
Current allocated memory: 186 bytes
Current allocated memory: 63 bytes
Current allocated memory: 84 bytes
Current allocated memory: 88 bytes
Current allocated memory: 186 bytes
Current allocated memory: 94 bytes
Current allocated memory: 166 bytes
Current allocated memory: 67 bytes
Current allocated memory: 58 bytes
Current allocated memory: 57 bytes
Current allocated memory: 105 bytes
Current allocated memory: 23 bytes
Current allocated memory: 24 bytes
Current allocated memory: 24 bytes
```



```

<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>
<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>
<i>
<span style='font-family:"Times New Roman",serif'>
</i>
<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"'>
<span lang=EN-US style='font-family:
"Courier New"'>
<span lang=EN-US style='font-family:
"Times New Roman",serif'>
<span
lang=EN-US style='font-family:Courier'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;margin-right:0cm;margin-bottom:
0cm;margin-left:61.1pt;margin-bottom:.0001pt;text-indent:-18.0pt;line-height:
12.0pt'>
<p class=MsoNormal style='margin-left:61.1pt'>
<span lang=EN-US
style='font-size:8.0pt;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'>
<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-family:"Times New Roman",serif'>
<span style='font-family:
"Courier New"'>
<span style='font-family:"Courier New"'>
<p class=MsoBodyTextIndent style='margin-top:6.0pt;text-indent:0cm;line-height:
12.0pt'>
</div>
</body>
</html>
```

Total allocated memory: 2919 bytes

Number of tags found: 46

PS C:\Users\willr\Documents\GitHub\Cs2263\Assignments\assignment5>