ASSIGNMENT THREE

PROCESSING STRINGS, ARRAYS, POINTERS

CS2263, Fall 2021

LEARNING OUTCOMES

Develop and test a utility program, `htags` that analyses an HTML file and prints of the list of the HTML tag names together with the number of occurrences of each tag.

HTML FILE STRUCTURE

Here is a simple html file:

```
<!DOCTYPE html>
<html lang = "en"> <!-- attribute that the language is english -->
    <head>
        <meta charset="utf-8"/>
        <title>Hello (Suessian) world!</title>
    </head>
    <body>
        <p>Hello</p>
        <p>World</p
    </body>
</html>
```

Elements of the file are enclosed with tags describing their purpose (tag names). There are three forms of tags:

1. Those that have opening and closing tags of the form <tag-name></tag-name>
2. Those that have no content (single tag) of the form <tag-name/>
3. Those that have no content (single tag) and begin with “<!”. Your program should ignore these tags.

Note that tags can contain other text (attributes). These should be ignored for our purposes. For example, `<html lang = "en">` is an html tag.

YOUR TASK

Your program should read the html file from the standard input (use the input redirection to read the entire HTML file) read the entire HTML file into a character array, called here the input array. *This input array is the only data structure for storing text* (i.e. strings of characters) in your program. The new implementation must not store the HTML tags directly in a separate data structure (array). Instead, `htags` should use pointers stored in an index table, pointing to the first occurrence of each tag identified in the input array. A separate (parallel) array of integers should then be used to keep track of the number of occurrences of each identified tag type. Output from the program should list each tag name, followed by the number of occurrences, per line. The following would be reported for the html above:

```
$ htags < hello.html
html  1
head  1
meta  1
title 1
body  1
p     2
```

- Your program must consist of at least TWO functions, with at least one of them compiled and tested separately.
- The program is to be implemented using pointers ONLY to access array elements. The square brackets "[" and "]" can be used to specify the array dimensions ONLY. In all other cases you must use pointers.
- Only one char array, the "input" array, can be used to store the text from the entire input HTML file.
- You are not allowed to store the detected tags in a char array: you must use the index table to store pointers to the tags detected in the input char array.
- Assume the input file contains less than 100000 characters.
- Assume there are less than 100 different tag types in the HTML file.
- Your program will be marked against a different input file.

THE REPORT

- In a few sentences describe the design of your program. Focus on what each of the data structures holds and how each of the functions acts on them.
- Show the testing of one of the functions using a test program.
- Show the output from running your program on the included HelloWorld.html file.
- Show the output from running your program on the included Sample.html file.
- There are other html files to try as well

SUBMISSION FORMAT

Before the due date for this assignment, students should submit a single zip or tar file (named *LastName_FirstName_A3.zip or LastName_FirstName_A3.tar*) online to the lms containing:

- Your report as a pdf file
- Your source code directory:
    - This should include all of your source files, including test programs and makefile.
    - This should not include object (.o) files and executables. Nobody needs to see those.