Tarnowskie Góry, 22.04.2016

Semester: II

Group: 3

Computer Programming Laboratory

Author: Marek Paśnik

Tutor: Jakub Nalepa

1. **Task topic**

**HTML**

The program should allow analysing and formatting HTML files. Existing tags, attributes should be counted; the most frequent tag should be found etc. The input file should be finally formatted (by adding or removing whitespaces).

1. **Project analysis**

2.1. Data structures:

* Binary tree – used to store tags taken from the file

2.2.Algorithms used (the most interesting ones):

- malloc – allocating the memory for node

- strcmp – compares the tag names taken from an input file

- getc – reading characters from an input file

2.3 Problem analysis:

Program has to read the input file character by character, analyse the code and detect the tags and attributes inside them. Program should count the amount of tags and attributes and later show them to the user. Program is also formatting the input file code, inserts the cuts and break lines and prints the formatted output to separate file.

1. **Internal specification**

3.1 Variables:

**char input\_file\_name[255**] – variable that stores the name of the input file given by user in command line

**char output\_file\_name[255]** - variable that stores the name of the input file given by user in command line

**char \*buff** – universal buffer

**struct Tag\* root –** root of tag tree

**struct Attribute\* root\_a –** root of attribute tree

**3.2 Functions:**

**int get\_tag(FILE\*, char\*,FILE\*); //get tag from Input HTML file**

**struct Tag \*add\_tag(struct Tag\*, char\*); //return address of a node and adds the tag to the tree**

**struct Attribute \*add\_attribute(struct Attribute\*, char\*); //return address of a node and adds the attribute to the tree TODO**

**struct Tag \*create\_node(); //make tree node**

**void ptree(struct Tag\*,FILE \* output\_file); //prints a tree of tags**

**void ptree\_a(struct Attribute\*,FILE \* output\_file); //prints a tree of attributes**

**void show\_help(); //prints help with usage of the program**

**void print\_error(int code\_of\_error); //error handling function, prints error**

**int len(char\*); //return string length without '\0'**

**int mostFrequent(struct Tag\*); //checks the most frequent tag**

**void print\_error(int code\_of\_error)** – this function receives the integer parameter named

`code\_of\_error`. Depending which code of error it receives from -1 to -7 it prints the specified error message (for example. If the parameter is -1 then it prints `Error: There is no input file with that name. Check the name of the file you want to read and try again.`).

**void show\_help()** – as the name says this simple function just prints the help page with proper usage of parameters and theirs descriptions.

**void insert\_space(int x, FILE \*output)**

Function is inserting number of spaces to output file.

1. **Testing**

Input file:

<html>

<head>

<title>Input HTML File</title>

</head>

<body>

<h1>Input HTML File Header</h1>

<p style="color:red">Testing attribute counter</p>

</body>

</html>

Output file:

<html>

<head>

<title>Input HTML File </title>

</head>

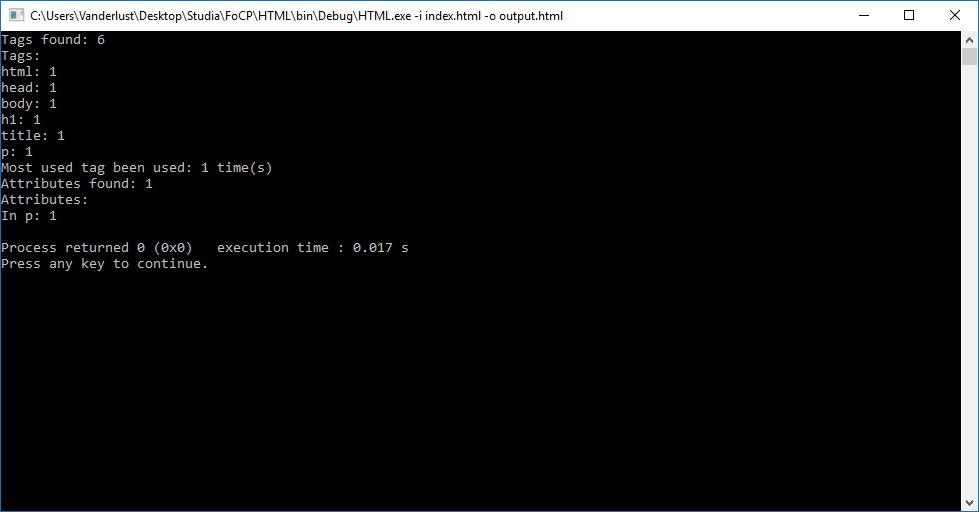
<body>

<h1>Input HTML File Header </h1>

<p style="color:red">Testing attribute counter </p>

</body>

</html>



# Bugs

5.1 Counting attributes but not show attribute itself

5.2 Spaces are inserted between tags