CMP2003 Data Structures and Algorithms (C++)
Term Project
Instructor : Assistant Prof. Tevfik Aytekin

-- Log Analyzer --

1. Main Requirements

You are expected to write a c++ console application which reads a text file that consists of HTTP requests to a web server. An example part is given below from access_log file :

Host	TimeStamp	Filename	HTTP Reply Code	Bytes in Reply
local - -	[24/Oct/1994:13:41:41 -0600]	"GET index.html HTTP/1.0"	200	150
local - -	[24/Oct/1994:13:41:41 -0600]	"GET 1.gif HTTP/1.0"	200	1210
local - -	[24/Oct/1994:13:43:13 -0600]	"GET index.html HTTP/1.0"	200	3185
local - -	[24/Oct/1994:13:43:14 -0600]	"GET 2.gif HTTP/1.0"	200	2555
local - -	[24/Oct/1994:13:43:15 -0600]	"GET 3.gif HTTP/1.0"	200	36403
local - -	[24/Oct/1994:13:43:17 -0600]	"GET 4.gif HTTP/1.0"	200	441
local - -	[24/Oct/1994:13:46:45 -0600]	"GET index.html HTTP/1.0"	200	3185

…

Your program should be able to store individual filenames from the log file and insert it into a suitable data structure, if it is a filename. If that filename is already encountered before its counter should be increased by 1. Try to find the most efficient data structure since we will also take into consideration the time efficiency of your code. The log file will be provided in the itslearning system. Note that we can use a different log file (with the same format) during the demonstrations. Note that filenames, in the above example, are the strings between GET and HTTP strings such as "index.html", "1.gif", "2.gif", etc.

After reading and processing is over, your program should list the "top 10" most visited web pages in decreasing order of # of total visitors .

Sample output :

Filename1 # of total visitors
Filename2 # of total visitors
Filename3 # of total visitors
Filename4 # of total visitors
Filename5 # of total visitors
.
.
.
Filename10 # of total visitors

Total Elapsed Time : X seconds

Whole application can be implemented with console facilities (you do not need advanced GUI elements). The project consists of two parts.

A. Implementation of the data structure.

This will be a proper C++ class. (Please use no third party libraries including C++ STL, Boost etc.). However you can use, **iostream**, **ctime**, **fstream**, **string** like IO and string related classes.

B. The main program itself. The program should create an instance of the data structure, put filenames from the text file into the data structure and update their visitor counts there. Your program should calculate the elapsed time starting from reading "access_log" text file until the end of the printing top 10 web pages. There is no need to ask user input.

2. Submission

You are expected to submit

a. A working copy of your program executable.
b. The Visual Studio project directory (config files) and source code.
c. A small report on sample execution, description of data structures/algorithms used and how to build the application and run. Even better, use a screen capture program like Camtasia Studio, to record a sample run of your executable in video, to make our life easy.

The project is at most 3 PERSON size.
You are encouraged to work in groups, 1 person project grades will be decreased by 20%.

The deadline is set 22 December 2017 11:59 pm. Submit your files from itslearning system.
Demos will be scheduled and announced later.
Late submissions will get lower grade by 10% for each day.

3. Cheating Policy.

You are not supposed to use each other's source code. Also do not use source code from internet, another person or your book's examples.

All the source codes will be filtered through a similarity analysis tool, which is known to be effective against many types of code copying and changing tricks. These projects will be graded as 0.

4. Evaluation

70% of the grade will be the amount of your work you did, correctness of your results, and your understanding of the concepts involved. We will order all projects according to their running times, and you will get the remaining 30% grade from this gradation.

Any lack of 3 items mentioned above (source code with visual studio config files, executable and project report), crashing executables and existence of viruses may cause you to get very low grades.

In the demonstrations all the group members must be present. All the members will be asked questions and a common grade will be given to all group members. Therefore wrong or inconsistent answers will affect all group member's grade.

5. Bonuses

You can get bonuses for extra efforts :

* Good coding styles and OO programming skills

* Or any other nice feature you can think of.

Please mention such extra efforts.