Data Structures Lab: Project 2

Introduction

This is the second of two projects in this class. The purpose of these projects is for you to explore a topic we have covered in lab in more detail. This will require you to do some research on your own; however, if you get stuck, please ask me for help. You should choose ONE of the following projects, complete it, and put your results in the git repository for this project. Please make a note of which project you have chosen in the `README.md` file and your submitted PDF. If you would like to do another project that is not on this list, but is related to the topics we have discussed so far in class, feel free to ask.

For each project, you must submit a PDF file containing answers to questions and screenshots documenting your project, as well as any configuration files or scripts you wrote for the assignment. You PDF should fully document the project, including how it was run, the results, and a discussion of the approach you took in the code. You may use LaTeX to generate the PDF file if you like, but are not required to do so; if you submit a doc or docx instead, you will lose 15 points. Your submission for this project should be similar to the lab assignments you have submitted so far.

Project A: Regular Expressions

Write a program (in C++ or another language of your choice) that, given an HTML file, does the following:

1. Prints the text of every hyperlink (`<a>` tag) and the URL it links to.
2. Prints the alt-text of every image (`<img>` tag) and the URL of the image.

Hints:

- `wget` or `curl` can download webpages for you.
- Use regular expressions! (C++ 11 has a regular expression library. Make sure to use `g++ -std=c++11` to get all the C++11 goodies!)

Project B: C++ Standard Library

Using data structures from the C++ standard library, implement a program to count the number of occurrences of each word in a file. Your program should output a table of words and their corresponding occurrence count. You should ignore punctuation from the beginning and end of words and case, so 'Bob,' and 'bob' are both the same word. The book to count words in is the thought-provoking classic:

`$ wget` `https://www.gutenberg.org/ebooks/4507.txt.utf-8`

Hint: use a map or unordered_map.

Project C: Unit Testing

Using Catch, Boost, or the bash simple_cpp_test provided in class, write unit tests for a project or assignment of your own. If you choose this option, please check with me that your chosen project will suffice for testing!

Project D: GUIs

Use Qt to make a 4-function calculator app. Make sure to also include thorough documentation in your submitted PDF, including screenshots of its functionality.