1585 Data Structures Lab
Spring 2023

Project #2

This is the second project for this class. The purpose of this project is for you to explore a topic we have covered in the lab in more detail. This will require you to do some research on your own; however, if you get stuck, please ask for help. You should choose **ONE** of the following options, complete it, and put your results in the git repository for this project. Please make a note of which option you have chosen in the `README.md` file.

For each option, you must enter answers to questions in the `README.md`, as well as add any configuration files or scripts you wrote for the assignment to this repository. Your submission for this project should be similar to the lab assignments you have submitted so far.

Option 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!)

Option 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`.

Option C: Unit Testing

Using **Catch**, **Boost**, or the bash simple cpp test provided in class, write unit tests for one of your **CS-1575** assignments.