

## Python In-class Programming Assignment

### Log File Analyzer

Update your repo if necessary (using `git pull`) and copy all the files in the following directory to a location of your choosing: `~/repo201/classwork/cw41/`. This directory contains an executable file called `logScanner` to allow you to see how the finished code should work. You can run `logScanner` by first copying it to your working directory and then adjusting its permissions with the following commands:

- a. `cp ~/repo201/classwork/cw41/logScanner .`  
(the trailing space and period in the command above are important)
- b. `chmod 755 logScanner`

You only have to perform steps (a) and (b) once. After that, you can run the program any time you want by typing: `./logScanner <log filename>`

*The Apache HTTP Server, colloquially called Apache, is free and open-source cross-platform web server software, released under the terms of Apache License 2.0. Apache is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation.*<sup>1</sup>

Included in `cw41` are two log files for a real Apache web server. Write a Python program that takes the name of a log file from the command line and generates the output shown below:

```
File statistics:
    Unique IP addresses: 3,291
    Largest hit count: 12,500
IPs matching largest hit count: 171.34.22.9
```

Once you have the IP address (or addresses) with the largest hit count, point a browser to: <https://www.infobyip.com/> and see if you can find anything interesting ☺

*You may get multiple IP addresses with the same max hit count, which you should be able to handle, but for the two test files I've given you, there's only one IP address in each log file with the max number of hits.*

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Apache\\_HTTP\\_Server](https://en.wikipedia.org/wiki/Apache_HTTP_Server)