

Assignment 4

Write a Python programme that will do analysis of an Apache access log file. There are two example log files on the moodle site, `apache.log` and `apache_short.log`. Use these to test your script, however keep your programme general as it will be corrected against a third access log from a similar server.

Your programme should do **all** of the following:

1. `-n` Number of unique IP addresses in log:

```
$ ./assign4.py -l apache.log -n
15559
```

2. `-tN` List top N IP addresses in the log file by the number of requests (where N is a number).

```
$ ./assign4.py -l apache.log -t 4
IP                      Requests
95.108.240.252          30362
137.43.92.119           10093
130.149.198.12          7545
66.249.76.189           6839
```

3. `-v` Number of visits by an IP address:

Given an IP address, calculate the number of visits to the site. A visit starts when the address first requests a page, and ends when the address has not requested a page for more than 1 hour.

```
$ ./assign4.py -l apache.log -v 192.168.1.1
4
```

4. `-L` List all of the requests made by an IP address

```
$ ./assign4.py -l apache.log -L 92.112.13.66
92.112.13.66 ... "GET /style.css HTTP/1.1" 200 934 ...
92.112.13.66 ... "GET /calc HTTP/1.1" 200 64305 ...
92.112.13.66 ... "GET /favicon.ico HTTP/1.1" 404 507 ...
```

5. `-d Date` list number of visits of all the requests on a specific date (Here Date is date in ddMMMyyyy format)

```
$ ./assign4.py -l apache.log -d 09Feb2013
178.151.143.247        1054
111.118.171.103        784
188.44.56.155          673
```

Notes:

- Use functions (even better if you can use classes, but not mandatory) within your programme and reuse as much code as possible when performing the different tasks.

- The Python module *getopt* is useful for parsing command line arguments:
<https://docs.python.org/3.2/library/getopt.html>
- See the Apache documentation for more detail on the access log file
<http://httpd.apache.org/docs/current/logs.html#accesslog>
- Use regular expressions to extract fields from each line
<https://docs.python.org/3.2/howto/regex.html>
- Name the file assign -XXX.py where XXX is your student number.
- Upload a single *.py script to moodle.
- In Python there is a lot of existing code available on the internet, and you are of course welcome to look at this, however your programming should be entirely written by you.
- At the head of it put your name, email, student number and what the program does in a comment block