

Apache Log Parser App

The `Apache Log Parser App` or `ALPA` for short is a command line application that can be used to parse old [Apache HTTP Server](#) logs from 1995 to generate a report with a number of statistics.

How to build the app

To build the app you will need to download and install (if not already available on your system):

- [Java SE 12 JDK](#)
- [Maven](#)

Open a terminal (in UNIX-like systems) or command prompt (in Windows) and do the following steps:

Clone the project locally:

```
git clone git@github.com:velissarious/ApacheLogParserApp.git
```

Move to the directory:

```
cd ApacheLogParserApp/
```

Use Maven with the `package` target to generate the executable jar file which will also contain all the dependencies. Like this:

```
mvn package
```

This will produce a number of jar files. Specifically:

```
stefanos-0.9.0-SNAPSHOT.jar
stefanos-0.9.0-SNAPSHOT-jar-with-dependencies.jar
```

Use the second jar file as listed for convenience, as it is the one that contains all dependencies.

How to use the app

To use the app you must start a terminal or a command prompt at its location and type the following to execute it:

```
java -jar stefanos-0.9.0-SNAPSHOT-jar-with-dependencies.jar
```

By default the program will try to find input file called `access_log_Aug95` which is available from [NASA-FTP](#) and produce a plain text file report called `report.txt`.

Input

To use another input file you must specify it with the `-i` or `-input` option, for example:

```
java -jar stefanos-0.9.0-SNAPSHOT-jar-with-dependencies.jar -i small
```

In this example the program will attempt to parse a file called `small`.

The program is designed to work only with Apache log files.

Report

The program will output a file report called `report.txt` in the directory the program is located in. You can specify a different report file with the `-r` or `-report` option like this:

```
java -jar stefanos-0.9.0-SNAPSHOT-jar-with-dependencies.jar -r test.txt
```

In this example, the program will generate the report with the name `test.txt` instead of `report.txt`.

Verbose

By default, the program will not display any output if successful, it will only display errors and warnings if need be, e.g. it will display an error for each malformed log entry encountered while parsing the log file.

To make the display output more verbose you need to specify the `-v` or `-verbose` option like this:

```
java -jar stefanos-0.9.0-SNAPSHOT-jar-with-dependencies.jar -v
```

Option Parsing

Using option parsing you can to generate a report that contains one of the following points:

0 - Default option, all of the bellow:

1 - Top 10 requested pages and the number of requests made for each.

2 - Percentage of successful requests (anything in the 200s and 300s range).

3 - Percentage of unsuccessful requests (anything that is not in the 200s or 300s range).

4 - Top 10 unsuccessful page requests.

5 - The top 10 hosts making the most requests, displaying the IP address and number of requests made.

7 - For each of the top 10 hosts, show the top 5 pages requested and the number of requests for each page.

You can use this feature by specifying the the `-o` or `-option` option:

```
java -jar stefanos-0.9.0-SNAPSHOT-jar-with-dependencies.jar -o 1
```

This example will display item 1 from the list above.

Technical Details

Dependencies

The project has the following dependencies that are also defined in Maven's `pom.xml` file.

- [SQLite](#) jdbc - a file based database.
- [JCommander](#) - a command line parsing library.
- [JUnit 4](#) - for unit testing.

Portions of the program are written as SQL queries.

Assumptions

- All possible input are log files are Apache HTTP Server log file from 1995.
- The input log will always fit in memory.
- Pages are only `html` pages.

Sample Output

This is `report.txt` contents for the `access_log_Aug95` file.

1. Top 10 requested pages and the number of requests made for each

```
-----  
/ksc.html 43379  
/shuttle/missions/sts-69/mission-sts-69.html 24544  
/shuttle/missions/missions.html 22339  
/software/winvn/winvn.html 10059  
/history/history.html 10039  
/history/apollo/apollo.html 8957  
/shuttle/countdown/liftoff.html 7836  
/history/apollo/apollo-13/apollo-13.html 7157  
/shuttle/technology/sts-newsref/stsref-toc.html 6451  
/shuttle/missions/sts-69/images/images.html 5258
```

2. Percentage of successful requests (anything in the 200s and 300s range)

```
-----  
99.3460084667921 %
```

3. Percentage of unsuccessful requests (anything that is not in the 200s or 300s range)

```
-----  
-----  
0.65399153320789 %
```

4. Top 10 unsuccessful page requests

```
-----  
/shuttle/missions/STS-69/mission-STS-69.html  
/elv/DELTA/uncons.htm  
/software/winvn/winvn.html.  
/shuttle/missions/technology/sts-newsref/stsref-toc.html  
/shuttle/missions/sts-79/mission-sts-79.html  
/shuttle/missions/sts-69-mission-sts-69.html  
/shuttle/technology/sts-newsref/stsref-to.html  
/software/winvn/winvn.html/wvsmall.gif  
/software/winvn/winvn.html/winvn.gif  
/software/winvn/winvn.html/bluemarb.gif
```

5. The top 10 hosts making the most requests, displaying the IP address and number of requests made.

```
-----  
-----  
edams.ksc.nasa.gov 6530  
piweba4y.prodigy.com 4846  
163.206.89.4 4791  
piweba5y.prodigy.com 4607  
piweba3y.prodigy.com 4416
```

www-d1.proxy.aol.com 3889
www-b2.proxy.aol.com 3534
www-b3.proxy.aol.com 3463
www-c5.proxy.aol.com 3423
www-b5.proxy.aol.com 3411

7. For each of the top 10 hosts, show the top 5 pages requested and the number of requests for each page requests made.

Host: edams.ksc.nasa.gov
/ksc.html 1020
/shuttle/missions/sts-69/mission-sts-69.html 28
/shuttle/missions/sts-69/liftoff.html 17
/shuttle/missions/missions.html 16
/whats-new.html 15

Host: piweba4y.prodigy.com
/shuttle/missions/sts-69/mission-sts-69.html 111
/shuttle/missions/missions.html 96
/ksc.html 70
/shuttle/countdown/liftoff.html 42
/history/apollo/apollo.html 34

Host: 163.206.89.4
/ksc.html 251
/shuttle/missions/sts-69/mission-sts-69.html 54
/shuttle/countdown/liftoff.html 44
/shuttle/missions/missions.html 40
/shuttle/countdown/countdown.html 26

Host: piweba5y.prodigy.com
/shuttle/missions/sts-69/mission-sts-69.html 115
/shuttle/missions/missions.html 88
/ksc.html 71
/history/history.html 38
/shuttle/countdown/liftoff.html 32

Host: piweba3y.prodigy.com
/shuttle/missions/sts-69/mission-sts-69.html 119
/shuttle/missions/missions.html 100
/ksc.html 76
/shuttle/technology/sts-newsref/stsref-toc.html 31
/history/apollo/apollo-13/apollo-13.html 31

Host: www-d1.proxy.aol.com
/ksc.html 72
/shuttle/missions/sts-69/mission-sts-69.html 71
/shuttle/missions/missions.html 57
/history/history.html 39
/history/apollo/apollo.html 29

Host: www-b2.proxy.aol.com
/ksc.html 69
/shuttle/missions/sts-69/mission-sts-69.html 66
/shuttle/missions/missions.html 50
/history/history.html 33
/history/apollo/apollo.html 24

```
Host: www-b3.proxy.aol.com
/ksc.html 77
/shuttle/missions/missions.html 73
/shuttle/missions/sts-69/mission-sts-69.html 51
/history/history.html 39
/history/apollo/apollo.html 29
```

```
Host: www-c5.proxy.aol.com
/ksc.html 79
/shuttle/missions/sts-69/mission-sts-69.html 61
/shuttle/missions/missions.html 51
/history/apollo/apollo.html 38
/history/history.html 33
```

```
Host: www-b5.proxy.aol.com
/shuttle/missions/sts-69/mission-sts-69.html 59
/shuttle/missions/missions.html 51
/ksc.html 51
/history/history.html 25
/shuttle/countdown/liftoff.html 24
```

The display output would contains the warnings for malformed output. There are no malformed logs in this example. If there were it would look like the snippet bellow:

```
...
Malformed log entry in line 1333474 !
Malformed log entry in line 1333505 !
Malformed log entry in line 1333551 !
...
```