

HTTP Log File Analyzer

Input File: weblog.txt

Programming Language: Java

Free & Open Source Library Used: JFreeChart (Java Library for Charts, Graphs, Plots and Piecharts, etc.)[<http://www.jfree.org/>]

Java Concepts Used: File I/O, String Tokenizer, Vector and Enumeration.

Operating System: Any OS(Linux, Windows, etc.) with Java Support.

Scope of Assignment:

The assignment is divided into **6 modules**:

1. BROWSER STATS Module:

It gives details about distribution of Browsers like Firefox, Internet Explorer, Opera, Safari, Chrome, Netscape.

2. BOT STATS Module:

It gives details about distribution of Search Engine Spiders like GoogleBot, Yahoo Slurp, MSNBot, Twiceler, Exabot.

3. OS STATS Module:

It gives details about distribution of Operating Systems like Windows, Macintosh, Linux.

4. WEBSITE STATS Module:

It gives details the most Popular Website based on URL count.

5. SUCCESS STATS Module:

It gives the percentage of ERROR hits to respective domains.

6. LOG VIEWER Module:

It shows logs of User Activity, Accessed URLs and User Agent Log.

Theory:

User Agent String Format

Source: http://en.wikipedia.org/wiki/User_agent#Format

Introduction:

The User-Agent string format is currently specified by Section 14.43 of [RFC 2616 \(HTTP/1.1\)](#)

String Format:

The format of the User-Agent string in HTTP is a list of product tokens (keywords) with optional comments. For example if your product were called WikiBrowser, your user agent string might be *WikiBrowser/1.0 Gecko/1.0*. The "most important" product component is listed first. The parts of this string are as follows:

- Product name and version (*WikiBrowser/1.0*)
- Layout engine and version(*Gecko/1.0*). In this case, this indicates the Layout engine and version.

Unfortunately, during the browser wars, many web servers were configured to only send web pages that required advanced features to clients that were identified as some version of Mozilla.

For this reason, most Web browsers use a User-Agent value as follows:

Mozilla/[version] ([system and browser information]) [platform] ([platform details]) [extensions]

Example:

**Mozilla/5.0 (iPad; U; CPU OS 3_2_1 like Mac OS X; en-us)
AppleWebKit/531.21.10 (KHTML, like Gecko) Mobile/7B405**

The components of this string are as follows:

Mozilla/5.0: Previously used to indicate compatibility with the Mozilla rendering engine

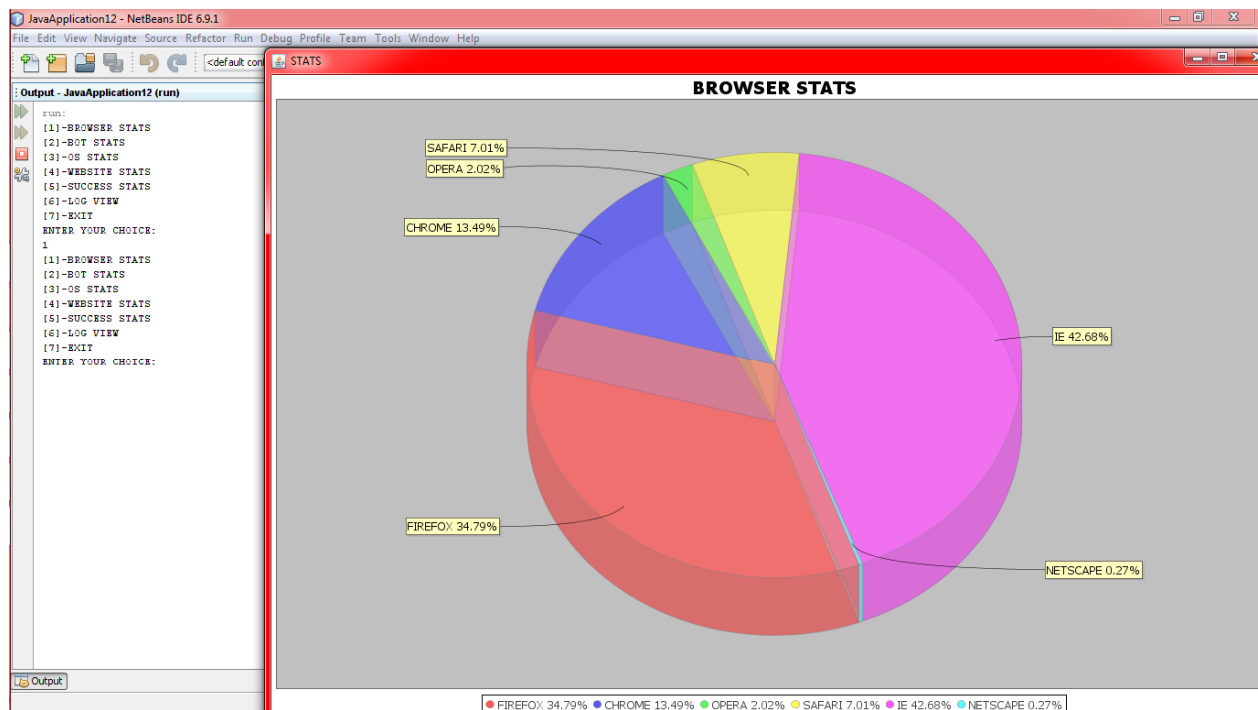
(iPad; U; CPU OS 3_2_1 like Mac OS X; en-us): Details of the system in which the browser is running

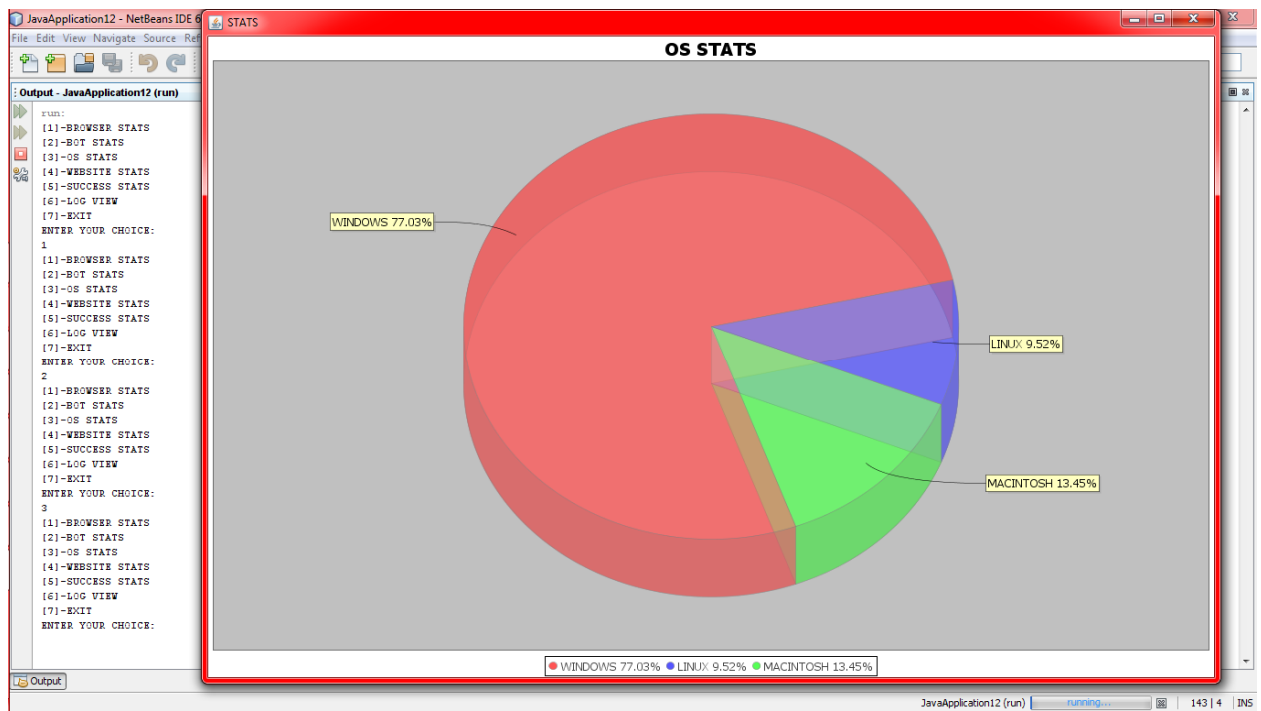
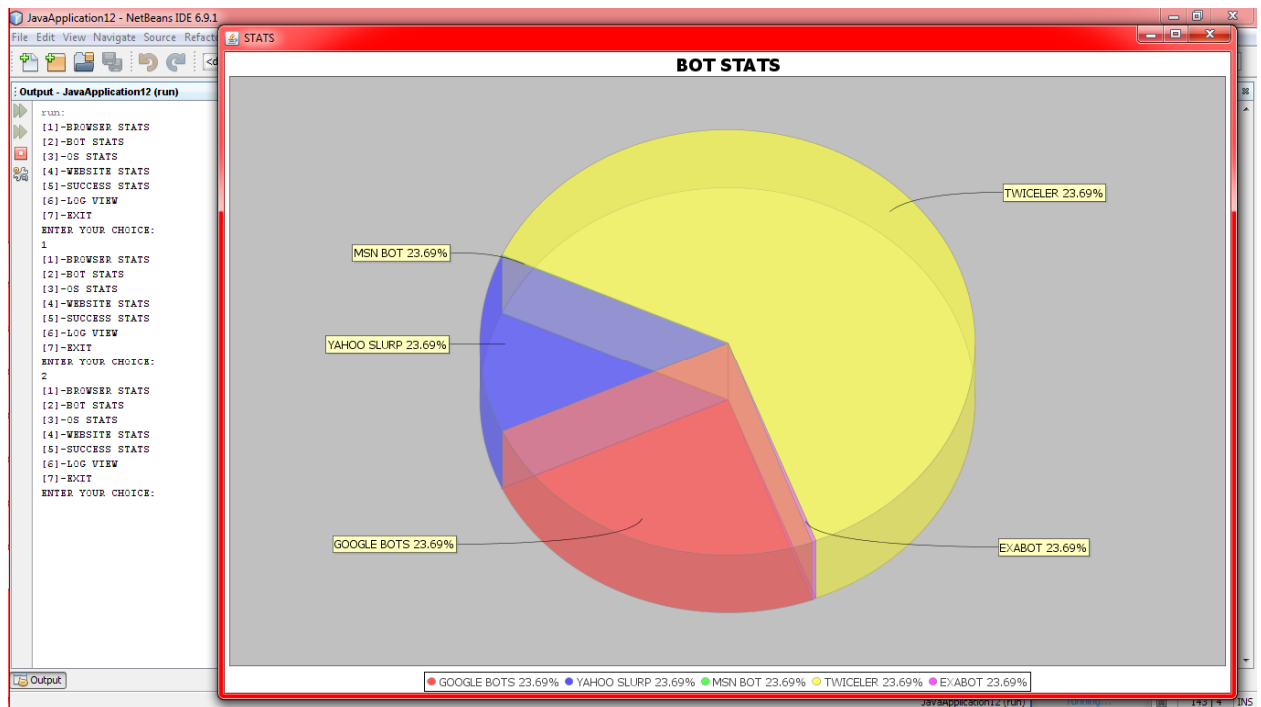
AppleWebKit/531.21.10: The platform the browser uses

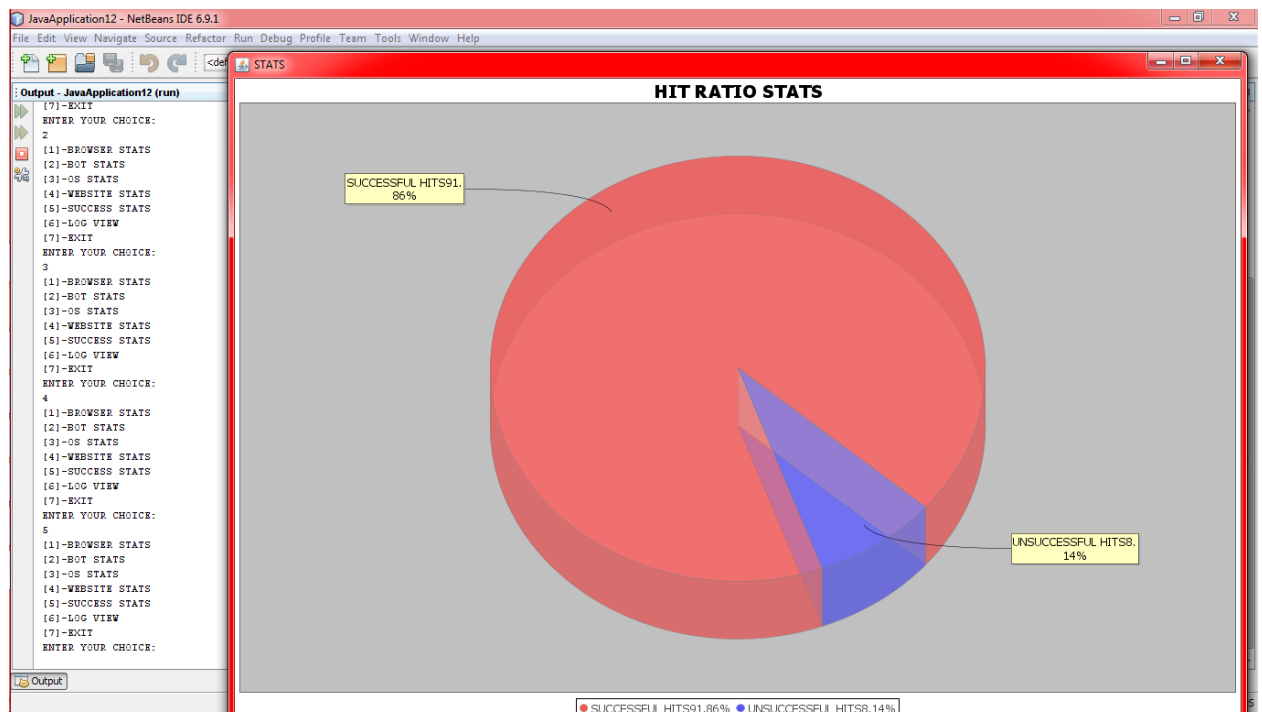
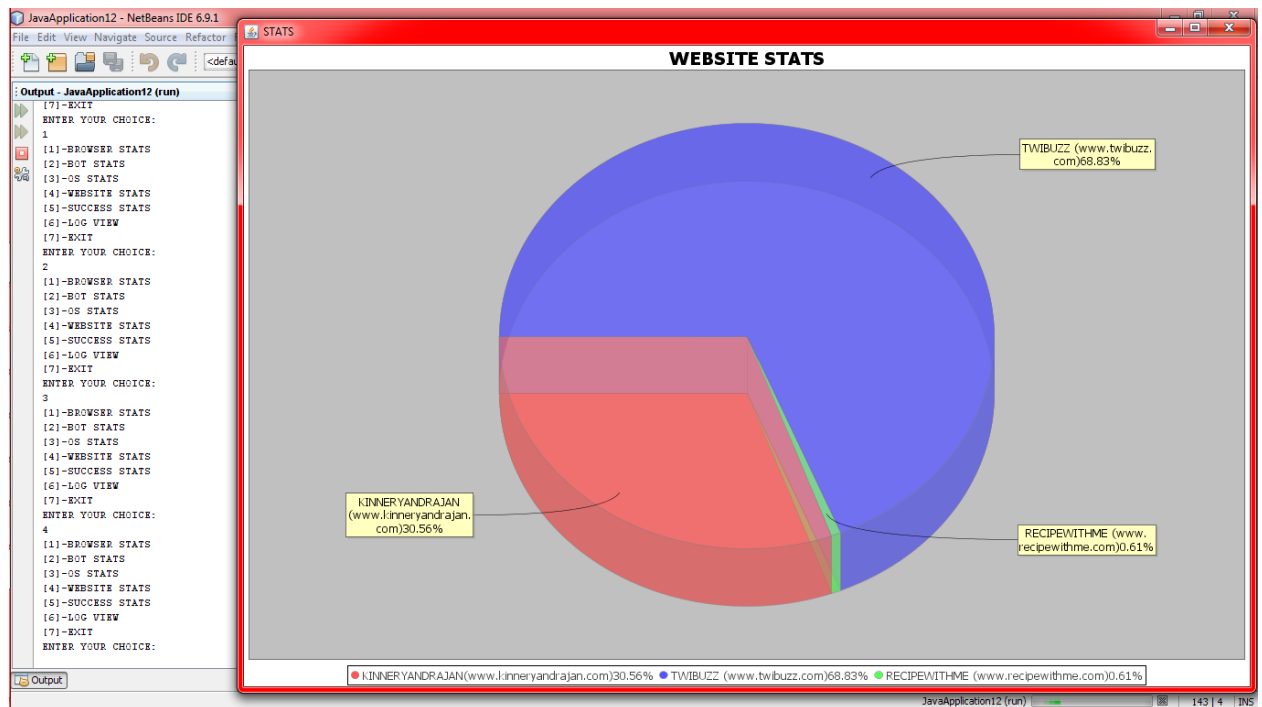
(KHTML, like Gecko): Browser platform details

Mobile/7B405: This is used by the browser to indicate specific enhancements that are available directly in the browser or through third parties. An example of this is Microsoft Live Meeting which registers an extension so that the Live Meeting service knows if the software is already installed, which means it can provide a streamlined experience to joining meetings.

Output:







```
JavaApplication12 - NetBeans IDE 6.9.1
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
<default config> Search (Ctrl+I)

Output - JavaApplication12 (run)
[5]-SUCCESS STATS
[6]-LOG VIEW
[7]-EXIT
ENTER YOUR CHOICE:
5
[1]-BROWSER STATS
[2]-BOT STATS
[3]-OS STATS
[4]-WEBSITE STATS
[5]-SUCCESS STATS
[6]-LOG VIEW
[7]-EXIT
ENTER YOUR CHOICE:
6
[1]-USER ACTIVITY LOG
[2]-ACCESSED URL LOG
[3]-USER AGENT LOG
[4]-BACK
ENTER YOUR CHOICE:
1

Sno IP ADDRESS DATE URL BYTES
0 82.95.120.160 14/May/2009:00:10:34 www.twibuzz.com 112813
1 82.95.120.160 14/May/2009:00:10:35 www.twibuzz.com/style.cs 639
2 82.95.120.160 14/May/2009:00:10:36 www.twibuzz.com/logo.png 5410
3 82.95.120.160 14/May/2009:00:10:43 www.twibuzz.com/favicon.ic -
4 66.249.70.102 14/May/2009:00:35:33 www.twibuzz.com/cgi/tweets.py?q=denre 7322
5 66.249.70.102 14/May/2009:00:36:20 www.twibuzz.com/cgi/tweets.py?q=nb 5898
6 66.249.70.102 14/May/2009:00:37:08 www.twibuzz.com/cgi/tweets.py?q=may120da 7755
7 72.30.142.50 14/May/2009:00:40:56 www.kinneryandrajan.com/robots.tx -
8 88.30.142.50 14/May/2009:00:40:56 www.kinneryandrajan.com/feed-comments-rss 7430

Output
JavaApplication12 (run) running... 143 | 4 | INS
```

```
JavaApplication12 - NetBeans IDE 6.9.1
File Edit View Navigate Source Refactor Run Debug Profile Team Tools Window Help
<default config> Search (Ctrl+I)

Output - JavaApplication12 (run)
2770 216.129.119.44 19/May/2009:23:52:47 www.twibuzz.com/robots.tx 1621
2771 74.6.22.103 19/May/2009:23:36:51 www.twibuzz.com/cgi/search.py?query=dvt 76150
2772 74.6.22.103 19/May/2009:23:36:52 www.twibuzz.com/style.cs 639
2773 74.6.22.103 19/May/2009:23:36:54 www.twibuzz.com/cgi/tweets.py?q=dvt 7809
2774 74.6.22.103 19/May/2009:23:36:54 www.twibuzz.com/style.cs 639
2775 154.8.75.149 19/May/2009:23:39:18 www.kinneryandrajan.com/?page_id= 14125
2776 154.8.75.149 19/May/2009:23:39:20 www.kinneryandrajan.com/wp-comments-post.ph 0
2777 216.129.119.44 19/May/2009:23:52:12 www.kinneryandrajan.com/wp-trackback.php?p=1 0
[1]-USER ACTIVITY LOG
[2]-ACCESSED URL LOG
[3]-USER AGENT LOG
[4]-BACK
ENTER YOUR CHOICE:
2

0 www.twibuzz.com
1 www.twibuzz.com/style.cs
2 www.twibuzz.com/logo.png
3 www.twibuzz.com/favicon.ic
4 www.twibuzz.com/cgi/tweets.py?q=denre
5 www.twibuzz.com/cgi/tweets.py?q=nb
6 www.twibuzz.com/cgi/tweets.py?q=may120da
7 www.kinneryandrajan.com/robots.tx
8 www.kinneryandrajan.com/?feed=comments-rss
9 www.kinneryandrajan.com/?page_id=
10 www.kinneryandrajan.com/wp-comments-post.ph
11 www.kinneryandrajan.com/?page_id=
12 www.kinneryandrajan.com/wp-comments-post.ph
13 www.kinneryandrajan.com/robots.tx
14 www.kinneryandrajan.com
15 www.kinneryandrajan.com/?page_id=
16 www.kinneryandrajan.com/wp-comments-post.ph

Output
JavaApplication12 (run) running... 143 | 4 | INS
```

