

# Statistical Analyzer of Webpage Contents

Sharada Vakkund<sup>[1]</sup>, Dr. Manu T M<sup>[2]</sup>

Department of Computer Engineering  
Department of Electronics and Communication Engineering  
, KLEIT, and Hubballi

## ABSTRACT

The web page components effects on the load time of the webpage and how it can be modified to reduce the load time. There are many performance measuring software automated tools for website, online website monitoring tools and a mathematical models which are used in context with web page contents to make quantitative and qualitative analysis of load time. The average time for an application to load must be of 8 seconds across the internet which tells the quality of the website or the services provided by the developer for the user's satisfaction. This study reveals there are different effects on web page content characteristics, and also there is increase in response time, if the total size of contents of the web page components is more. The evidence in this study shows that the web page response time depends on the contents of the web page in relationship to object type and size. Many efforts have been made to improve the performance on the load of webpage by reducing the computing time, by modifying the contents of the page or by making the objects to load faster. So, a detailed analysis of webpage characteristics has to be done by developers before deploying a website.

**Keywords** :— website, performance, url, analyser

## I. INTRODUCTION

Website is a collection of webpages and multimedia content, it is identified by a domain name and it should published on one webserver. Website plays a vital role on Internet. Website can be accessed through a public Internet Protocol (IP), such as Internet or a private Local Area Network (LAN) by referencing Uniform Resource Locator (URL) that identifies the website [2]. A unique web address is assigned to each website which can be accessed through Internet.

Webpages are the building blocks of website which are documents and it is composed in a plain text mixed with formatting languages Hyper Text Markup Language (HTML) and Extended Hyper Text Markup Language (XHTML).

### 1.1 Web Application and testing:

Web Application is a program which is stored on remote server and reached to the user through web browser using Internet. Web application Testing is checking whether the web application has any potential error before it is made available to the end user or before the code is sent to production team.

#### Website testing elements

1. JavaScript.
2. CSS (Cascaded Style Sheets).
3. Cookies.
4. Security.
5. Accessibility.

### 1.2 Performance Testing

Performance Testing is the process for deciding the performance of the device, computer, network or software. By performing performance testing we can check whether the system meets the specifications given by the vendor or manufacturer. Different Types of Performance Testing

1. Load testing.

2. Spike testing.
3. Stress testing.
4. Scalability testing.
5. Volume testing.

### 1.2.1 Common Performance Problems

The following is a list of common problems that exists while doing performance testing:

- a) **Long load time:** Load Time is a time that the application takes to start and it should be kept minimum.
- b) **Poor Scalability:** is the when the application is not capable to handle expected number of users.
- c) **Poor Response Time:** Response Time is the time taken for an application to response back when the user inputs the data.

### 1.3 Automation Testing

**Automation Testing** is testing an application using software testing tool instead of doing it manually by the tester.

#### Goals and Objectives of Test Automation

1. Saves time and money.
2. Find more bug and improves the quality.
3. Replace manual testers.
4. Increases the test coverage.
5. Improves quality.
6. Speeds the testing process.
7. Reduces Redundancy.

Many testing tools are available, some of them are commercial and some are open source.

Commercial Testing Tools- QTP, HP Mercury, WinRunner, IBM Rational Functional Tester, MicroFocus, SilkTest.

Open Source Testing Tools- Selenium, Sikuli, FitNesse

### Top 3 Automation Tools

1. Selenium.
2. QTP/UFT.
3. TestComplete

## II. RELATED WORK

Zhou Munyaradzi and others suggested that webpage load time is mainly affected by its page content rather than client and the webserver network conditions as in [1]. According to them the webpage has different components and each of these have different effects on the load time of a webpage. JavaScript's increases the number of HTTP requests which increases the load time. So the JavaScript's should be put at the bottom of the page to reduce the latency which increases in webpage load time. Images have width and height missing attributes which causes poor response time .So by specifying the width and height attribute of the images makes the webpage to load the entire objects ,hence it reduces responses time.HTML images have total size of 4\*305 bytes which approximately 3.24% of a total webpage size. He also says that increase in the distance between the client and webserver increases the response time of the webpage which affects the average load time of webpage. So there is need to CDN's (Content Delivery Network) to reduce the response time..

## III. OBJECTIVES

1. To see whether the requested webpage is loaded.
2. To extract the response time and size of objects that are present in the webpage.

## IV. PROPOSED SYSTEM

To load the web page and get the load time and size of the components (images, video, css, javascript's) present in a webpage that user has requested for and also get the number of requests. Here in the "website performance analyzer" tool, the user has to register then login and the he/she can enter the url and ask the tool to load the webpage and once the page has been loaded completely, it will then display the load time and size of other objects present in the webpage. The system design of the proposed system is shown in figure 1.

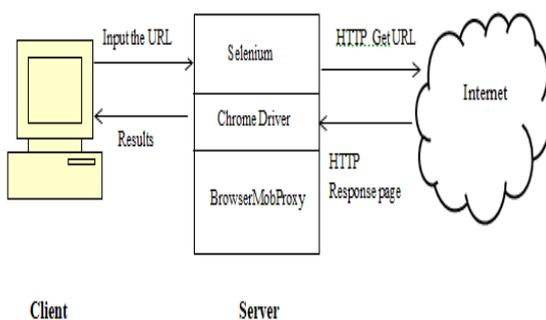


Fig 1: System Design

## V. METHODOLOGY

### A. Chrome Driver

Chrome Driver is a stand-alone server which is used to control chrome.Chromium team maintains chrome driver with the help from web driver contributor. Web Driver is an automated tool used for testing web applications with different browsers. Web Driver is an open source tool. It provides power for navigating to user inputs, JavaScript execution, webpages etc. [5].Selenium WebDriver is most popular tools for Web User Interface Automation.

Web User Interface Automation means the automatic execution of the events in a web browser. Selenium WebDriver can interact with different types of web browser.

### B. BrowserMobProxy

BrowserMobProxy is an open source tool capture performance data for web applications in the HAR format. It collects the data from client side. BrowserMobProxy is simple service to capture data and typically written using automation tool kit such as Selenium and Waitr. BrowserMobProxy is useful tool which is closely integrated with Selenium and can work independently. [8]

### C. HAR Format

HTTP Archive Format or HAR is a JSON formatted archive file format for logging of web browser interaction with a site. The extension of these files is .hr [9].HAR is an outline tool visualizing and can be used for analysis of page load performance. By using any web browser like Google Chrome, Firefox and Internet Explorer; we can capture the data of HTTP session

### D. Selenium

Selenium was developed by Jason Huggins in 2004. Selenium can be managed by many programming languages and testing frameworks [3, 4, 6, and 7].

Selenium is not only one tool but it is set of software's each provides set of different testing needs of an organization.

To start with selenium, the following requirements are:

- a) **Python:** Selenium supports python.
- b) **pip:** pip is a package management system which is used to install and manage software package written in Python.
- c) **Virtualenv:** provides an isolated terminal of python which allows the user to work on project without affecting about other projects.

### Implementation

The process flow of the website performance analyser is shown in figure 2.Here the user has to register first by providing his user credentials like Name, Email-Id and password. After he registers to the tool, he/she can login using email-id and password. If the password or email-id is wrong then it displays "Email or password" that time the user has to login again with a valid email and password. Once

he/she logs in to the tool, the homepage of the tool appears where he/she can enter the URL starting from “http://”, if he does not start with http:// then the server doesn’t identify the webpage which we have requested for. Once the URL is entered in the URL space, the server checks whether the requested URL it is a valid URL or not. If the requested webpage URL is correct, then the web page is loaded. Once the webpage is loaded completely, it displays a message saying “Page Loaded” and also displays the size and page load time of each component of the webpage. Later, if he/she wants to continue to test other website then he/she can enter the URL once again on which he wants to perform testing. Later once he/she has finished analysing the result. Then the users can logout of the page.

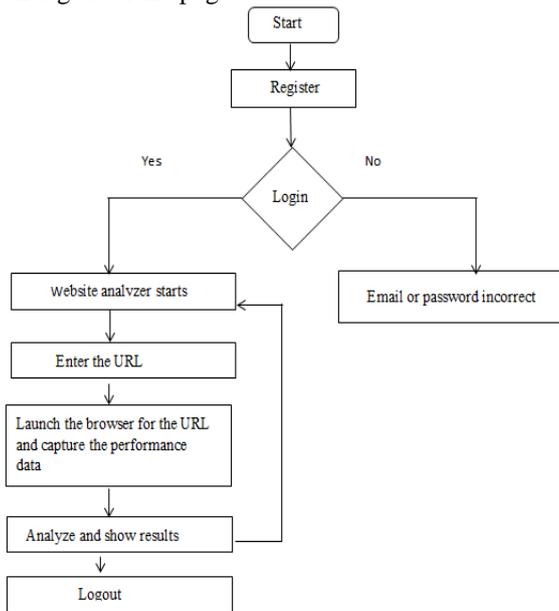
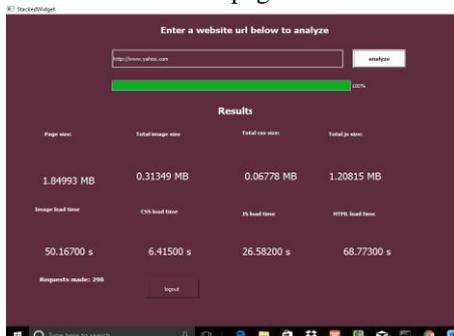


Fig 2: Process Flow the website performance analyser

## V. RESULTS

Once the user enters his/her user credentials ,he/she registers then he/she can login, once the user logs in he/she has to enter an URL on which automation testing has to be done starting with http:// and once the URL is entered and the response time of the web page content is displayed. Figure 3 shows the size and response time of the web page



Page Size: 1.84 MB

Fig 3: Snapshot of output

Request Made: 298

Table 1: Object Size and Load time

Object Type	Size(bytes)	Load Time
Total Image	0.31 MB	50.16 sec
CSS	0.67 MB	6.41 sec
JavaScript	1.20 MB	26.58 sec
HTML	0.65 MB	68.7 sec

## VI. CONCLUSIONS

The web page response time depends on the web page contents in relationship to object type and size. So a detailed analysis of webpage characteristics has to be carried out by developers before deploying a website to optimize the contents of the webpage to get faster load time of a page. Many efforts are made to improve the performance of the webpage load by doing the computations like by modifying the contents of the page or by making the objects to load faster. So the developers should try to improve the webpage contents to reduce response time when using other browsers.

## REFERENCES

- [1] “Effects of Web Page Contents on Load Time over the Internet” Zhou Munyaradzi1, Giyane Maxmillan2, Mutembedza Nyasha Amanda3. International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064
- [2] <https://en.wikipedia.org/wiki/Website>
- [3] <https://www.slideshare.net/krkstar/selenium-testing-project-report>
- [4] <http://qeworks.com/selenium-webdriver-architecture/>
- [5] <https://sites.google.com/a/chromium.org/chromedriver/>
- [6] [https://en.wikipedia.org/wiki/Selenium\\_\(software\)](https://en.wikipedia.org/wiki/Selenium_(software))
- [7] <http://www.guru99.com/introduction-to-selenium.html>
- [8] <http://www.tothenew.com/blog/browsermob-proxy-for-webpage-load-testing-using-selenium/>
- [9] <https://en.wikipedia.org/wiki/.har>