PERFORMANCE ANALYSIS OF WEB SEARCH ENGINE USING JMETER

1JYOTSANA KUMARI, 2HEMLATA, 3SHASHI SHARMA, 4MANISH MATHURIA, 5ANKIT KR. TIWARI

1,2Computer Engineering Department, Jaipur Institute of Technology, Jaipur
3Asst. Prof., Department of Computer Science, Jaipur Institute of Technology, Jaipur
4,5Asst. Prof. of CSE, Maharishi Arvind College of Eng. & Research Center, Jaipur
Email: 1jyotsnakumari093@gmail.com, 2chtanvi@gmail.com, 3shashi_sharma26@yahoo.com.com,
4manish4598@gmail.com, 5ankittiwari148@gmail.com

Abstract: The Performance of Web Applications is varying according to the locations of server as well as internet capability at client side. But for a common user which has normal Internet capability but still some Web Application took more time in loading when compare. The most popular Search Engines are Google, Yahoo, and Bing; these search engines always try to attract audience form the world with different services. This research tried to test performance as well as analyze the loading capability of the server when large amount of users request for the same page. This testing and analysis is done using Apache’s Load Testing Tool: JMeter, which give multiple options to judge performance using different parameters. The test plan is configured using HTTP Request Default whereas sampler is chosen named HTTP Request to handle client request for required service. The results obtained are aggregated to compile final result to value the performance of Web Search Engine.

Keywords: Traffic load, Web Search Engines, JMeter, Search Engine Optimization.

I. INTRODUCTION

A web server is a computer system that processes requests via HTTP, the basic network protocol used to distribute information on the World Wide Web. The term can refer to the entire system, or specifically to the software that accepts and supervises the HTTP requests to reach user’s requirement. The web Server is a keen entity of Web Oriented System which is whole sole responsible for the completion of Client-Server action. Instead, other devices are just tried to many communication channels between to ends.

1.1 Web Server Load

Web traffic load is the amount of data sent and received by visitors to a web site. This necessarily does not include the traffic generated by bots. Since the mid-1990s, web traffic has been the largest portion of Internet traffic.[1] This is determined by the number of visitors and the number of pages they visit. Sites monitor the incoming and outgoing traffic to see which parts or pages of their site are popular and if there are any apparent trends, such as one specific page being viewed mostly by people in a particular country. There are many ways to monitor this traffic and the gathered data is used to help structure sites, highlight security problems or indicate a potential lack of bandwidth.

Not all web traffic is welcomed. Some companies offer advertising schemes that, in return for increased web traffic (visitors), pay for screen space on the site. Sites also often aim to increase their web traffic through inclusion on search engines and through search engine optimization. Modern high-traffic websites must serve hundreds of thousands, if not millions, of concurrent requests from users or clients and return the correct text, images, video, or application data, all in a fast and reliable manner. To cost-effectively scale to meet these high volumes, modern computing best practice generally requires adding more servers.

1.2 Web Service Management by Web Server

The interaction between Web Servers and Web Services is simple: HTTP is a simple protocol browsers use to communicate with Web Servers. Web Servers, fulfill users’ requests and store the information users provide. Meanwhile, Web Services allow different Web Servers to communicate and interact with one another in order to process the request and/or commands of the user.

A Web Server is the central nervous system of Web Site. It is the Web Server that hosts both the components of a Web page such as the actual Web page HTML files, CSS files and templates and all other essential technologies that make a Web site function the way it does. Although all Web servers function similarly, the set up and the way a server could be set can vary drastically. The Client-Server network, on the other hand, is a highly centralized network system with one central computer as the server. This set up is easy to manage and secure [2]. Web Service is a very powerful tool that has greatly enhanced the efficiency and communication among businesses. According to the World Wide Web
Consortium, “a Web Service is a software system designed to support interoperable machine-to-machine interaction over a network. Web Services as a ’reusable software components based on XML and related protocols that enable near zero-cost interaction throughout the business ecosystem.

1.3 How Server Handle multiple client requests
Web server respond to the client request in either of the following two ways:
- Sending the file to the client associated with the requested URL.
- Generating response by invoking a script and communicating with database

2. How Server Handle multiple client requests

2.1. History of Search Engine
1. Google - Offering everything from image searches, map searches, news searches, etc. With impressive keyword relevancy and a continuously improving search algorithm, it's easy to see why Google is still the reigning champ.
2. Yandex - Yandex is most used search engine in Russia. In fact it is an Russian internet Company. It was launched in 1997. Yandex also provides services like Yandex Maps, Yandex Music, Online translator and many other services.
3. Yahoo - While Yahoo has been suffering as of late, it's still a classic and a popular search engine.
4. Bing - The Microsoft powered search engine prides itself on being a "decision engine" by offering search suggestions on the side column and providing extra search options.
5. Ask - Clean layout and handy results grouping.
6. AOL Search - AOL continues to be used, primarily by people who still use AOL. They're out there somewhere.
7. Excite - It was launched in 1995.
8. DogPile - The once alternative to Google is getting a comeback and is a great alternative to bigger search engines.
9. Duck Duck Go - Doesn't track your search history and is avoids spammy sites.
10. The Internet Archive - This search engine lets users travel back in time to see how web pages looked in years gone by. A very fun search engine to play around with [4,5].

II. WEB SEARCH ENGINE

Search Engine refers to a huge database of internet resources such as web pages, newsgroups, programs, images etc. It helps to locate information on World Wide Web. User can search for any information by passing query in form of keywords or phrase. It then searches for relevant information in its database and return to the user.

2.1. History of Search Engine
1. Google - Offering everything from image searches, map searches, news searches, etc. With impressive keyword relevancy and a continuously improving search algorithm, it's easy to see why Google is still the reigning champ.
2. Yandex - Yandex is most used search engine in Russia. In fact it is an Russian internet Company. It was launched in 1997. Yandex also provides services like Yandex Maps, Yandex Music, Online translator and many other services.
3. Yahoo - While Yahoo has been suffering as of late, it's still a classic and a popular search engine.
4. Bing - The Microsoft powered search engine prides itself on being a "decision engine" by offering search suggestions on the side column and providing extra search options.
5. Ask - Clean layout and handy results grouping.
6. AOL Search - AOL continues to be used, primarily by people who still use AOL. They're out there somewhere.
7. Excite - It was launched in 1995.
8. DogPile - The once alternative to Google is getting a comeback and is a great alternative to bigger search engines.
9. Duck Duck Go - Doesn't track your search history and is avoids spammy sites.
10. The Internet Archive - This search engine lets users travel back in time to see how web pages looked in years gone by. A very fun search engine to play around with [4,5].
search results list, the more visitors it will receive from the search engine's users, and these visitors can be converted into customers. SEO may target different kinds of search, including image search, local search, video search, academic search, news search and industry-specific vertical search engines. SEO is the process of improving the organic ranking of a website with leading search engines. Professional SEO firms like Submit Express help to improve search engine rankings for their clients by modifying their websites to better reflect what search engines are looking for.

IV. LOAD ON PUBLIC WEB SERVER

Literature Review

In this study our concern is Web search engines. In the literature there are two types of search engine evaluation approaches: testimonial and shootout. Testimonials are casual studies and state the general impression obtained after executing a few queries. Shootouts are rigorous studies and follow the information retrieval measures for evaluation purposes.

The Gordon and Pathak (1999) study measures the performance of eight search engines using information-needs. For measuring performance it calculates recall and precision at various document cut-off values (DCVs) and uses them for statistical comparisons. Intermediaries prepare the queries from the information-need descriptions of real users [6].

Desirable features of Web search evaluation according to Gordon and Pathak (1999); features 1–7 and Hawking et al. (2001): features 8–11

1. The searches should be motivated by genuine information-needs of Web users
2. If a search intermediary is employed, the primary searchers information-need should be captured as fully as and with as much context possible and transmitted to the intermediary.
3. A sufficiently large number of searches must be conducted to obtain meaningful evaluations of search engine effectiveness.
4. Most major search engines should be considered
5. The most effective combination of specific features of each search engine should be exploited (i.e. the queries submitted to the engines may be different)
6. The user who needs the information must make relevance judgments (Hawking et al. (2001) assumes that independent judges can do it)
7. Experiments should (a) prevent bias towards search engines (e.g., by blinding or randomizing search outputs), (b) use accepted information retrieval measures, (c) employ statistical tests to measure performance differences of search engines.
8. The search topics should represent the range of information-needs over which it is desired to draw conclusions
9. Result judging should be appropriate to the type of query submitted (e.g., some queries may need a one-line answer).
10. Document presentation should be like that of a Web browser (images should be viewable, if necessary it should be possible to follow links).
11. Dead links should count as useless answers.

The study reported in Hawking et al. (2001) evaluates the effectiveness of 20 search engines using TREC-inspired methods with 54 queries taken from real Web search logs. The performance measures used include precision at various DCVs, mean reciprocal rank of first relevant document, and TREC-style average precision. Recall has not been used. Statistical testing reveals high inter correlations between performance measures and significant differences between performances of search engines.

A different kind of work is the study of Mowshowitz and Kawaguchi (2002), which measures the performance of search engines using the overlap of URLs of the matching pages. This study uses the similarity between the response vector of the collection of search engines and the response vector of a particular search engine, which is defined as bias, to evaluate the performance of that search engine. Another study is the Chowdhury and Soboroff (2002) work, which presents a method for comparing search engine performance automatically based on how they rank the known item search result. In this study, initial query–document pairs are constructed randomly. Then, for each search engine, mean reciprocal rank is computed over all query–document pairs. If query–document pairs are reasonable and unbiased, then this method could be useful. However, construction of query–document pairs requires a given directory, which may not always be possible. There is also the Soboroff, Nicholas, and Cahan (2001) study, which involve ranking retrieval systems without relevance judgments. It proposes a new methodology in the TREC environment; however, it is related to our work because it replaces human relevance judgment with randomly selected documents from a pool [7, 8].

3. V. LOAD TESTING USING JMETER

1. JMeter offers following benefit in Performance testing
2. JMeter can be used to test performance of both static resources such as JavaScript and HTML, as well as dynamic resources, such as JSP, Servlets, and AJAX.
3. JMeter can discover maximum number of concurrent users that your website can handle.

Performace Analysis of Web Search Engine using JMeter
JMeter provides a variety of graphical analyses of performance reports [9].

5.1. Traffic Load Parameters
- **Number of Threads**: Number of users connects to target website
- **Loop Count**: Number of time to execute testing
- **Ramp Up Period**: The Ramp Up Period is the most important parameter. It represents the ability of the server to handle heavy load. The higher the Ramp Up Period, the better is the server performance.
- **Deviation**: The deviation is shown in red - it indicates the deviation from the average. The smaller the better.
- **Black**: The total number of current samples sent.
- **Blue**: The current average of all samples sent.
- **Red**: The current standard deviation.
- **Green**: Throughput rate that represents the number of requests per minute the server handled [10].

5.2. JMeter Configuration

**Step 1) Add Thread Group**
1. Start JMeter
2. Select Test Plan on the tree
3. Add Thread Group
   - Right click on the Test Plan and add a new thread group: **Add -> Threads (Users) -> Thread Group**

**Step 2) Adding JMeter elements**
Now we determine what JMeter elements in this test. The elements are
- **HTTP request Default**
  - Right-clicking on the Thread Group and selecting: **Add -> Configure Element -> HTTP Request Defaults**.
  - In the HTTP Request Defaults control panel, enter the Website name under test (http://www.google.com) [11].
- **HTTP Request**
  - Right-click on Thread Group and select: **Add -> Sampler -> HTTP Request**.
  - In HTTP Request Control Panel, the Path field indicates which **URL request** you want to send to Google server.
  - For example, if you enter "search" in Path field, JMeter will create the URL request http://www.google.com/search to Google server.

**Step 3) Adding Graph result**
- JMeter can show the test result in Graph format.
  - Right click Test Plan, **Add -> Listener -> Graph Results**

**Step 4) Run Test and get the test result**

VI. RESULT ANALYSIS

6.1. Graphs
In this test, the throughput of **Bing Server** is 212.92/minute. It means Bing Server can handle 212.92 requests per minute.

![Graph Result: Bing Search Engine (No. of Threads-50)](image-url)
In this test, the throughput of **Yahoo Server** is 219.701/minute. It means Yahoo Server can handle 219.701 requests per minute.

![Graph of Yahoo Search Engine](image)

In this test, the throughput of **Google Server** is 274.316/minute. It means Google Server can handle 274.316 requests per minute.

![Graph of Google Search Engine](image)

6.2 Comparison

The throughput of website under test [http://www.yahoo.com](http://www.yahoo.com) is 219.701/minute. It means this server handle 219.701 requests per minute, lower than Google and Yahoo.

The throughput of website under test [http://www.google.com](http://www.google.com) is 274.316/minute. It means this server handle 274.316 requests per minute, lower than Google.

In this test, the throughput of Google server is 274.316/minute. It means Google server can handle 274.316 requests per minute. This value is quiet high so we can conclude that Google server has good performance.

**NOTE:** The above values depend on several factors like current server load at Google, the internet speed, your CPU power etc.
CONCLUSION

In this study we present the method for the performance evaluation of Web search engines. Knowing the most effective Web search engines satisfying the current information-need is important both at a personal and business enterprise level. However, the definition of “the best” changes due to both the changing information-needs of users and the changing quality and nature of search engines.

REFERENCES