

Response Time Performance Testing of Greenstone and DSpace Digital Library Software in Handling Rich Text Data

Shaka A. Imavah¹ and Fasola P. Abifarin²,
Department of Library and Information Technology,
School of Information & Communication Technology
Federal University of Technology, Minna, Niger State Nigeria

Email: ¹sa.imavah@futminna.edu.ng

²fp.abifarin@futminna.edu.ng

ABSTRACT

This study carried out performance testing on response time of Greenstone and DSpace software in handling rich text data in Federal University of Technology, Minna, Nigeria. Rich text data are data that contains a combination of texts, tables, symbols and many other text structures. The study sought to achieve two objectives and two corresponding research questions were formulated. A web-experimental design was adopted for the study while a Participant Observation Template (POT) served as the instrument for recording users' observations for response time of relevant data retrieved. Out of the 54 students targeted, only 35 attended the training hence 35 copies of the POT were distributed to an intact class of trained 500 level students in the Department of Library and Information Technology, Federal University of Technology Minna, Nigeria. 34 copies were returned and only 33 were found usable representing 97.14% response rate. The data collected were presented in tables and graphs and analysed using mean and standard deviation, while the hypothesis was tested using t-test statistic at 0.05 level of significance. Findings of the study revealed that the mean response time of Greenstone was faster than DSpace software by 0.01s in handling rich text data but the difference was not statistically significant.

Keywords: Digital libraries, Performance Testing, Response time, Greenstone, DSpace, Web Experiment, Open source

African Journal of Computing & ICT Reference Format:

Shaka A. Imavah and Fasola P. Abifarin (2018), Response Time Performance Testing of Greenstone and DSpace Digital Library Software in Handling Rich Text Data, *Afr. J. Comp. & ICT*, Vol.11, No.1, pp. 28 - 35.

© Afr. J. Comp. ICT, March 2018; ISSN 2006-1781

1. INTRODUCTION

Digital libraries and its associated technologies have greatly transformed the activities of traditional libraries. It is largely believed that the foundation of digital libraries was laid by such visionaries as Vannevar Bush in 1945

and J.C.R Licklider in 1965[1]. The birth of this concept has been attributed to the advent and rapid growth of computers and other related technologies such as the Internet, database management software and information retrieval systems thus making it increasingly easy for the

development of tools that make the acquisition, processing, organisation, storage and retrieval of information more effective and efficient [2].

Digital libraries are organised collection of digital objects such as text, images, sounds, videos and other multimedia formats that can be accessed electronically over a network. Digital libraries enable users access library collections that have either been digitized or “born digital” irrespective of time and geographical barrier, insofar as an active network connection is available to a user. Newby [3] posited that digital libraries will eventually become the most dominant means through which library users of the next millennium will have access to materials that are housed in the traditional brick-and-mortar libraries. This assertion is already gradually taking shape as a result of the proliferation of Digital Library Management Systems (DLMS).

Digital Library Management Systems (DLMS) are software packages which make the development and management of digital libraries easy. These systems are typically expected to handle a variety of contents ranging from texts, images, sounds/audio and videos in such a manner that their capture, management, dissemination and archiving are done effectively [1]. DLMS could either be categorised as proprietary or open source. The dwindling resources allocated to most libraries have shifted the focus of librarians from expensive commercial digital library management software to their freely available open source counterparts.

Therefore, in making decisions on which open source DLMS to adopt, librarians are often handicapped due to the limited information on the actual performance of these systems.

Performance testing is a systematic process through which software are tested in order to determine the performance of a system within a particular period. According to Barber [4] the aim of every performance testing is usually to ascertain the performance of a system within a controlled environment in a timely manner. The writer further posited that the main purposes of performance testing include: to ascertain how much of the expected load the system can handle, the robustness of the application under expected and unforeseen load, to find out users’ acceptance of the system when deployed, and to determine the response time of the system to users’ request.

Response time is an aspect of performance evaluation criteria which specifies the time it takes a digital library to carry out tasks such as navigation, browsing, searching or

obtaining resources [5]. It is the “average time a digital library takes to process all requests including link response time and search response time.” Similarly, Fuhr *et al.* [6] defined response time as the time a system or functional unit takes to react to a given input. Response time has been identified as one of the most crucial measures for evaluating digital libraries.

The Greenstone Digital Library Software (GDLS) or Greenstone is a group of interdependent software for creation and dissemination of digital collections [7]. It is a very popular and widely deployed open source application for building digital libraries worldwide. According to Shree Kumar in Manhas [8] the Greenstone software is promoted by the New Zealand Digital Library Project (NZDL) at the University of Waikato and enjoys the sponsorship of the United Nations Educational Scientific and Cultural Organisation (UNESCO). The initial release of Greenstone software was in 2000 under the GNU public license [9].

DSpace is another very popular open source application for building digital libraries and institutional repositories. It is a software package that provides tools for the management of digital assets and is commonly used as the basis for building institutional repositories [10]. It enables organisations to capture and describe digital materials with the aid of a submission workflow module and other ingest options, distribute digital assets of organisations over the web through a search and retrieval system and preserve digital assets over the long term [11]. DSpace was jointly developed by Hewlett Packard (HP) Laboratories and Massachusetts Institute of Technology (MIT) libraries and the first version was released to the public in November 2002 [12]. Both Greenstone and DSpace have the capacity to ingest a wide range of data objects such as datasets, aural, visual, audio-visual, and textual information of various formats.

Textual information has emerged as the predominant format of information in the world. Textual information is any information that is presented using words and characters. The advancements in computing technology have made it possible for textual information to be produced, organized, stored and disseminated at an alarming rate. According to Rouse ^[13], text is a “human-readable sequence of characters and the words they form that can be encoded into computer-readable formats such as American Standard Code for Information Interchange (ASCII), UTF-8 among others. Texts are generally classified into two broad categories namely; plain text and rich text. Plain texts do not contain any kind of formatting; only white spaces and line breaks while rich

texts also known as formatted texts or styled text contain formatting beyond the basic semantic elements such as colours, italics, bold. Furthermore, they are able to incorporate special features such as tables, graphics, hyperlinks, diagrams and so on.

This study therefore conducted a performance test on the response time of Greenstone and DSpace in handling rich text data.

Statement of the Problem

The component which has made the development and deployment of digital libraries less cumbersome is the Digital Library Management System (DLMS). Most times, librarians use their intuition in selecting digital library management systems for building digital collections. In using intuition alone, parameters such as software features are used in arriving at decisions. Digital libraries, by virtue of their complex nature require librarians to test the potentialities of these DLMS in areas of performance because this is the only way that the effectiveness of these software can be determined. When these software are put to task and observed, the choice of any particular software can be easily justified and the objectives of creating the digital library *ab initio* can be achieved.

It has been observed through extensive literature review that most comparative studies on digital library management software in the past often evaded carrying out performance testing of any sort on these software [14], [15], [16]. They simply carry out their study through the comparison of commonalities such as software features, number of installations and operating system compatibility among others.

However, the presence of these features alone does not guarantee that the software would work efficiently, but when these systems are tested for performance, better selection choices could be made. This study therefore, conducted performance testing on Greenstone and DSpace with a view to determining the more efficient software in handling rich text data.

Aim and Objectives

The aim of this study is to conduct performance test on the response times of Greenstone and DSpace in handling rich text data. To achieve this aim, the specific objectives are to:

1. determine the response time of Greenstone and DSpace software in handling rich text data.
2. compare the response time of Greenstone and DSpace software in handling rich text data.

Research Questions

The following research questions guided the study:

1. What is the response time of Greenstone and DSpace software in handling rich text data?
2. Is there any difference between the response time of Greenstone and DSpace in handling rich text data?

Research Hypothesis

The following null hypothesis was tested at 0.05 level of significance:

HO₁: There is no significant difference between the response times of Greenstone and DSpace in handling rich text data.

2. REVIEW OF RELATED LITERATURE

According to Magnussen [17], the foundation of the concept of digital libraries is not totally new. In 1945, Dr. Vannevar Bush of the United States Office of Scientific Research and Development talked about a device called a “memex”. The device proposed by Bush had the capabilities of providing individuals with some kind of personal filing and library system that is mechanized. The “memex” would also have the capacity to store and index a vast range of information resources such as books, photographs, periodicals and so on for easy retrieval [18]. Biswas and Paul [15], reported that the term “digital library” appeared in print for the first time in a 1988 report of the Corporation for National Research Initiatives. The authors also averred that the Digital Libraries Initiative of 1994 which was constituted by the National Science Foundation (NSF), Defence Advance Research Projects Agency (DARPA) and National Aeronautic Space Agency (NASA) further popularised the term.

The word digital library conveys different meanings to different categories of individuals and groups, depending on the perspective from which the term is viewed. In the opinion of Witten *et al.* [19], the contents of a digital library are focused and include various formats such as visual data, audio, textual information, multimedia, maps among others. The authors further posited that digital libraries are usually integrated with functions that aid access and retrieval, acquisition, organization, and maintenance of the collections.

Kumar [20] conducted a comparative study of open source digital library software. The scope of this study was limited to DSpace, Eprints and Greenstone. Also, the

study focused on evaluating the most popular and widely deployed software for building digital collections with the aim of helping digital library administrators make decision on the available software. The data for the study was collected through the use of checklist containing various parameters which have been weighted according to their importance to the package. The study revealed that most of the software are in developing stage but are good at providing a good service and concluded that DSpace emerged the best option among the three software that were studied.

Andro *et al.* [21] studied ten (10) software comprising six (6) open source (Invenio, Greenstone, Omeka, EPrints, ORI-OAI, DSpace) and four (4) proprietary software (Mnesys, DigiTool, Yoolib, CONTENTdm). Questionnaire method was used for collecting data from the software companies. The study concluded that all the solutions that were surveyed are of good quality.

Tramboo *et al.* [22] studied three digital library software with special reference to Greenstone, Eprints and DSpace. Data was collected for the study by online survey and extensive study of related software documentation and associated technical manual. The conclusion of this study is that it is difficult to propose one specific DLMS system as the most suitable for all cases.

3. RESEARCH METHODOLOGY

This study adopted Web-based experimental design. This research design is a type of online research method which involves experiments that are carried out with the support of Internet technologies such as servers, web browsers, computer networks, and computer hardware [23]. This design is suitable for the evaluation of digital libraries because they are systems that run entirely on web technologies.

The population of this study was the intact class of 500 level students of the Department of Library and Information and Technology, Federal University of Technology, Minna. The entire class comprising 54 students was targeted for this study but only the 35 students who attended the pre-testing training were used as participants. This category of students was considered because the collections in the databases are well suited for their undergraduate project needs; they have already been exposed to and are familiar with the test parameters used in this study; and majority of the students possess the basic technical skills such as internet browsing and online searching skills required to conduct this study.

The instruments that were used for this study are the treatment instrument and the test instrument. The treatment instruments are the two digital libraries that were created using the Greenstone and DSpace digital library management software respectively. The test instrument was the Participant Observation Template (POT) which was designed for the purposes of recording observations for response time of the two digital library systems,

The Greenstone digital library software and DSpace source codes were downloaded from www.sourceforge.net and used to create two separate databases on two standalone computers of similar specifications. Retrieved students' undergraduate projects contained in CD-ROMs were processed and merged. A uniform identifier was assigned to uniquely identify each document.

The processed data were ingested into the databases and described using the Dublin Core metadata standard. Each document was assigned 33 keywords in order to guarantee exhaustivity and specificity for easy retrieval. These terms were used to query the databases while an embedded digital stopwatch was used to ascertain the time it takes for each digital library database to respond to user query.

The location for the performance evaluation was the computer laboratory of the Department of Library and Information Technology, Federal University of Technology, Minna, Nigeria. The participants were trained and their consent sought after which a copy of the POT was distributed to them to enable them record their observations for response time for both digital library systems. Each participant was asked to select three keywords each from the bespoke index developed for this study i.e. one simple term, one compound keyword and one inverted keyword. These keywords were used to query the two databases one after the other and their observations for response time for each search activity. A total of 99 iterations were carried out and the data generated were computed using the Microsoft Excel 2013 spreadsheet application while the results were presented in tables and graphs and analysed using mean, and standard deviation.

4. RESULTS AND DISCUSSION

Research Question 1: What is the response time of Greenstone and DSpace software in handling rich text data?

Table 1 shows the response time and standard deviation of Greenstone and DSpace in handling rich text data. The result reveals that for the number of iterations carried out (N=99), the mean response time and standard deviation of Greenstone are 0.013s and 0.007s respectively. Similarly, the mean response time and standard deviation of DSpace are 0.014s and 0.013s respectively. This implies that Greenstone is faster than DSpace in handling rich text data.

Research Question 2: Is there any difference between the response time of Greenstone and DSpace in handling rich text data?

The response times recorded for each iteration as seen in Table 1 were entered into a Microsoft Excel 2013 spreadsheet application and were used to compute the mean response time of Greenstone and DSpace Software using the formulae:

$$\text{Mean response time} = \frac{\sum t}{n}$$

where: t = Response time

n = Number of iterations

Table 2 shows that the mean response time recorded for Greenstone and DSpace Software for 99 iterations (N=99) are 0.013s and 0.014s respectively. This indicates a difference of 0.001s between the mean response time of Greenstone and DSpace in handling rich text data.

Figure 1 shows the response time (y-axis) of Greenstone, in the solid line and DSpace, in the round dots against users' query (x-axis). The graph showed small but steady random fluctuations in the response times of both software. However, an abnormal upward trend was observed for Greenstone at q2 and q36 while DSpace recorded similar patterns at q2, q24, q36 and peaked at q68. The abnormal upward trends observed could be as a result of the participants' error such as failure to refresh the homepage of the DSpace library system before supplying a fresh query.

Hypotheses Testing

HO₁: There is no significant difference between the response time of Greenstone and DSpace software in handling rich text data.

From Table 3 it can be deduced that the result is not significant at P > 0.05 level of significance (P = 0.341). This implies that there was no statistically significant difference in the mean response time of Greenstone (\bar{X} = 0.013, μ = 0.007) and DSpace (\bar{X} = 0.014, μ = 0.013) in

handling rich text data. Thus, the null hypothesis is retained.

The findings from this study revealed that the response times for Greenstone software and DSpace software were short. The digital library built with the Greenstone software recorded a mean response time of 0.013s while the digital library developed using DSpace software recorded a mean response time of 0.014s. This implies that the two digital libraries are very fast in handling rich text data.

The findings partly support the argument of Nielsen in Jin [5] which says that most digital libraries return a response time that is usually less than 1.0s.

These extremely short response times could be attributed to the environment in which the experiment was conducted. The performance test of these libraries was conducted directly on the servers hosting these libraries in lieu of accessing and assessing the libraries from a remote server over the Internet.

5. CONCLUSION AND FUTURE WORKS

The study conducted a performance testing of the response time of Greenstone and DSpace in handling rich text data. The results showed that Greenstone recorded a mean response time of 0.013s while the mean response time for DSpace was 0.014s. The study further revealed that the difference between the mean response time of Greenstone and DSpace in handling rich text data was 0.01s.

Consequently, the study concludes that the response time of the digital library system built with Greenstone software was slightly faster than the one built with DSpace in handling of rich text data. However, there was no significant difference in the mean response time of Greenstone and DSpace in handling rich text data when t-test was used to analyse the mean response times recorded for both digital library systems. This may be attributed to the homogeneity of the digital objects contained in these digital library systems.

Future researches could be conducted to test the response time performance of these software with other digital objects such as photographs, sounds, and multimedia formats either homogeneously or heterogeneously.

REFERENCES

- [1]. Candela, L., Athanasopoulos, G., Castelli, D., El Raheb, K., Innocenti, P., Ioannidis, Y., . . . Ross, S. (2011). *Digital Library Reference Model- In a Nutshell*. European Union: Osmotica.
- [2]. Sun, J., & Yuan, B. (2012). Development and Characteristic of Digital Library as a Library Branch. *IERI Procedia*, 2(1), 12-17.
- [3]. Newby, G. B. (1996). *Digital Library Models and Prospects*. Urbana-Champaign: University of Illinois.
- [4]. Barber, S. (2006). *Introduction to Performance Testing*. Washington: PerfTestPlus.Inc.
- [5]. Jin, S. (2014). *Evaluating Existing Digital Libraries as a Prototype with the Suggested Criteria: Content, Usability, and Performance Evaluation Criteria*. Illinois: Unpublished Student Proposal at University of Illinois at Urbana-Champaign.
- [6]. Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., . . . Sølvsberg, I. (2007). Evaluation of Digital Libraries. *International Journal of Digital Libraries*, 8(1), 21-38. doi:10.1007/s00799-007-0011-z
- [7]. www.greenstone.org. (2015, April 15). *About Greenstone*. Retrieved from Greenstone Digital Library Software Web site: <http://www.greenstone.org/>
- [8]. Manhas, R. (2008). Development of Health Sciences Digital Library Using Greenstone Digital Library Software at Baba Farid University of Health Sciences. *Planner*, 359-372.
- [9]. Witten, I. H., & Bainbridge, D. (2003). *How to build a digital library*. San Francisco: Morgan Kaufmann Publishers.
- [10]. DSpace.org. (2015, May 5). *Introducing DSpace*. Retrieved from DSpace Website: <http://www.dspace.org/index.php/Introducing-DSpace/>
- [11]. The DSpace Developer Team. (2015, February 26). *DSpace 5.X Documentation*. Retrieved from Duraspace Website: <https://wiki.duraspace.org/display/DSDOC5x>
- [12]. Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., . . . Harford, J. (2003, January 12). DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, ix(1). doi:10.1045/january2003-smith
- [13]. Rouse, M. (2005, April 10). *Computer Fundamentals Glossary*. Retrieved from Techtarget Web site: <http://whatis.techtarget.com/definition/text>
- [14]. Goh, D., Chua, A., Khoo, D., Khoo, E., Mak, E., & Ng, M. (2006). A Checklist for Evaluating Open Source Digital Library Software. *Online Information Review*, 30(4), 360-379.
- [15]. Biswas, G., & Paul, D. (2010). An evaluative study on the open source digital library softwares for institutional repository: Special reference to Dspace and greenstone digital library. *International Journal of Library and Information Science*, 001-010.
- [16]. Kumar, L. (2013). Use of Open Source Software in Digital Library: A Comparative Study. *e-Library Science Research Journal*, 1(4), 1-4. Retrieved December 10, 2015, from <http://www.lsrj.in/UploadedArticles/23.pdf>
- [17]. Magnussen, A. (2003). Creating Digital Libraries: a Model for Digital Library Development. *10th Asia Pacific Special Health and Law Librarians Conference* (pp. 1-15). Adelaide: Unpublished.
- [18]. Bush, V. (1945, July 1). As We May Think. *The Atlantic Monthly*, pp. 1-10.
- [19]. Witten, I., McNab, R., Boddie, S., & Bainbridge, D. (2000). Greenstone: A Comprehensive Open-Source digital library software system. *Fifth ACM Conference on Digital Libraries* (pp. 113-121). San Antonio, Texas: Association for Computing Machinery(ACM), Inc.
- [20]. Kumar, V. (2009). Comparative evaluation of open source digital library packages. *OSLS 2009: National Seminar on Open source library solutions* (pp. 1-14). Varanasi: Unpublished MLIS thesis, Department of Library and Information Science, Banaras Hindu University, India.
- [21]. Andro, M., Asselin, E., & Maisonneuve, M. (2012). Digital libraries: Comparison of 10 software. *Library Collections, Acquisitions, & Technical Services (Elsevier)*, 36(3-4), 79-83.
- [22]. Trambo, S., Shafi, S., & Gul, S. (2012). A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone. *International Journal of Computer Applications*, 59(16), 1-9.
- [23]. Reips, U. (2013). The Web Experiment. In M. (. Wirtz, *Dorch, Psychologisches Wortebuch* (16th ed., p. 1660). Bern: Huber.

Table 1: Comparison of response time (seconds) of Greenstone and DSpace software in handling rich text data

Query	Greenstone	DSpace	Query	Greenstone	DSpace	Query	Greenstone	DSpace
q1	0.014	0.013	q34	0.011	0.012	q67	0.03	0.076
q2	0.05	0.037	q35	0.012	0.014	q68	0.023	0.118
q3	0.011	0.013	q36	0.050	0.051	q69	0.024	0.020
q4	0.014	0.013	q37	0.012	0.024	q70	0.014	0.014
q5	0.012	0.013	q38	0.012	0.012	q71	0.021	0.013
q6	0.026	0.013	q39	0.013	0.014	q72	0.013	0.012
q7	0.011	0.012	q40	0.013	0.015	q73	0.015	0.014
q8	0.006	0.005	q41	0.013	0.013	q74	0.014	0.013
q9	0.025	0.012	q42	0.013	0.014	q75	0.014	0.012
q10	0.014	0.013	q43	0.013	0.012	q76	0.011	0.012
q11	0.015	0.012	q44	0.012	0.013	q77	0.011	0.014
q12	0.025	0.011	q45	0.014	0.02	q78	0.012	0.011
q13	0.011	0.012	q46	0.014	0.014	q79	0.012	0.013
q14	0.002	0.002	q47	0.025	0.012	q80	0.015	0.015
q15	0.003	0.002	q48	0.014	0.013	q81	0.014	0.012
q16	0.005	0.013	q49	0.013	0.013	q82	0.013	0.012
q17	0.011	0.012	q50	0.015	0.004	q83	0.012	0.010
q18	0.009	0.011	q51	0.012	0.013	q84	0.014	0.012
q19	0.005	0.012	q52	0.023	0.013	q85	0.012	0.012
q20	0.010	0.012	q53	0.015	0.009	q86	0.013	0.012
q21	0.005	0.002	q54	0.012	0.013	q87	0.014	0.013
q22	0.002	0.005	q55	0.012	0.012	q88	0.012	0.013
q23	0.010	0.013	q56	0.017	0.013	q89	0.014	0.012
q24	0.005	0.040	q57	0.012	0.011	q90	0.016	0.012
q25	0.003	0.010	q58	0.014	0.005	q91	0.011	0.012
q26	0.008	0.013	q59	0.014	0.012	q92	0.013	0.012
q27	0.010	0.009	q60	0.011	0.012	q93	0.013	0.011
q28	0.009	0.012	q61	0.016	0.013	q94	0.014	0.013
q29	0.010	0.011	q62	0.011	0.009	q95	0.011	0.012
q30	0.012	0.012	q63	0.011	0.012	q96	0.012	0.009
q31	0.009	0.011	q64	0.015	0.012	q97	0.013	0.011
q32	0.005	0.012	q65	0.009	0.013	q98	0.011	0.013
q33	0.002	0.012	q66	0.012	0.014	q99	0.012	0.011
				Mean response time \bar{X}			0.013	0.014
				Standard deviation μ :			0.007	0.013

Table 2: Comparison of the Mean Response time of Greenstone and DSpace Software

Variables	Greenstone	DSpace	Difference in Mean Response Time
Mean Response Time	0.013s	0.014s	0.001s
Standard Deviation	0.007	0.013	

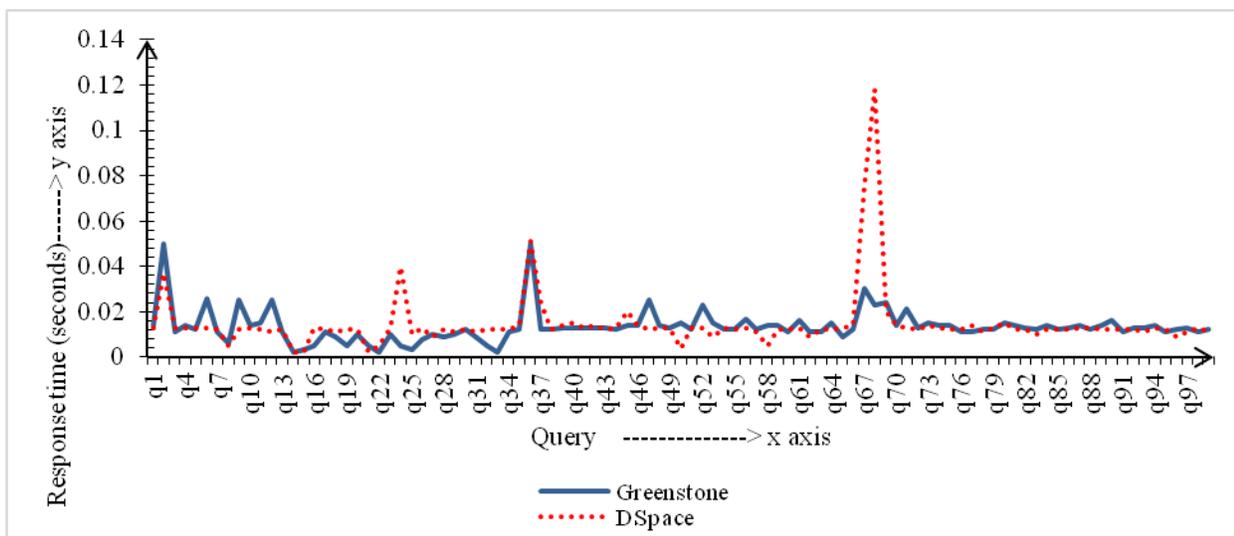


Figure 1: Response Time of Greenstone and DSpace software in handling rich text data

Table 3: t-test Comparison of the mean response time of Greenstone and DSpace in handling rich text data

Variables	N	Df	Mean (\bar{X})	SD (μ)	t-value cal.	t-value critical	Sig. (two tail)
Greenstone	99		0.013s	0.007			
		98			-0.958	1.984 ^{ns}	.341
DSpace	99		0.014s	0.013			

^{ns}Not significant at 0.05 level