ROLE OF OPEN SOURCE SOFTWARE TO MEASURE USAGE OF DIGITAL LIBRARY AND ITS IMPROVEMENT : A CASE STUDY

ACHINTYA K MANDAL **SUBODH GOPAL NANDI**

Abstract

The present study discusses on the digital library initiatives on the light of keystone principles. It includes open source software, and open source library software. This study analyses webliazer, an open source software as measuring tool of the digital library usage. It highlights on various methods of improvement, updation procedure of the digital library by using server log file analysis for proper maintenance of the library homepage by exploiting the user feedback at Visva-Bharati Library.

Keywords : Digital Library/ Open Source Software/ Keystone Principles/ Webalizer/ Visva-Bharati Library

1. Introduction

To build a digital library is one of the main issues developed to the library professionals of recent times. The term 'digital library' is often used synonymously as "electronic library" and "virtual library", but this term has some identity, which differentiate from other terms. These particulars or elements may be identified as under:

- The digital library is not a single entity;
- The digital library requires technology to link the resources of many;
- The linkages between the many digital libraries and information services are transparent to the end users;
- Universal access to digital libraries and information services is a goal [1].

2. Definition

Digital Library Foundation (DLF) has defined on digital libraries as:

"Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so

that they are readily and economically available for use by a defined community or set of communities [2]."

3. Digital library Components

Like conventional libraries, digital libraries have three main functions: a) Collection, b) Organisation and representations, and c) Access and retrieval. According to Johnson & Magusin the digital materials of a library includes the following:

- Digital surrogates, or records that represent physical items – for example, the bibliographic records in the library's online catalogue or citations in a research database.
- Digital resources that are derived from non-digital materials – for example, the electronic version of a journal for which there is a print publication.
- Born-digital materials that exist in digital form only and for which there are no print equivalents – for example, a journal that is electronic only.

The digital library collections are not limited to document surrogates: they extend to digital artifacts that cannot be represented or distributed in printed formats. Collection includes those techniques, which help library professionals to detect information resources useful for the library clientele. According to the need criteria relevant to library clientele, the information resources are to be collected and organized by classifying and indexing.

The key issue of Digital Library Project is an effective integration of tools for data capturing, data compression, storage, organization, semantic indexing, retrieval and navigation. The navigation should be user friendly and be based on worldwide available standard graphical user interfaces (GUIs) with multimedia technology.

Access and retrieval are to designing and organizing materials within a physical shape and thus effective retrieval of information items out of collected and organized resources by the library clientele successfully be made.

4. Keystone Principles and Digital library Initiatives

'Keystone Principles' has been formulated in 1999, as a set of principles and action items to guide specially the academic librarians for extending future action based traditional library values. "The Keystone Principles invoke and express the urgency of three areas requiring explicit action based on a vivid set of user-centered principles:

- Principle One: Scholarly and government information is a "public good" and must be available free of marketing bias, commercial motives, and cost to the individual user.
- Principle Two: Libraries are responsible for creating innovative information systems for the dissemination and preservation of information and new knowledge regardless of format.

- Principle Three: The academic library is the intellectual commons for the community where people and ideas interact in both the real and virtual environments to expand learning and facilitate the creation of new knowledge" [3].

Under these three principles some 'action items' has also been formulated; among these action items some are relevant for library resource development; library consortia; creation of digital libraries: from existing library documents, electronic reserves, online content, access tools, etc. To success this digital environmental programme librarian should have to apply his / her own efforts and collective endeavors. Although these Principles and Action Items has prescribed for Academic Libraries, those are equally applicable to all types of Libraries. These Keystone Principles gear up librarians for creation of library network, creation of library website, use of other library network system, use of other organisation's website, etc. Thus the Keystone Principles has encouraged librarians to find out the possible ways to satisfy the different queries of library clientele and supply required information to them.

5. Open Source Software (OSS)

Open source generally refers to a program where source code is available to the general public for their use and/or modification from its original design without paying any charge for that. [4] Open source code is typically created as a collaborative effort in which programmers improve upon the code and share the changes within the community. Open source software (OSS) helps programmers on the Internet read, redistribute and modify the source code, forcing an expedient evolution of the product. The process of eliminating bugs and improving the software happens at a much quicker rate than the traditional development process of commercial software as the information is shared throughout the open source community Thus uses of OSS in various applications have been increased at a very high rate. The application field of library and information science has no exception of it [5].

5.1 Open Source Library Software

Currently available Open Source library-related applications, which include not only integrated library management systems but also a range of innovative functionality, such as:

- DSpace - A digital library system to capture, store, index, preserve, and redistribute the intellectual output of a university's research faculty in digital formats.
- MOSST - Modular Online Software for Self-paced Tutorials, to create web-based tutorials.
- OSCR - Open Source Course Reserve, to manage electronic course reserve material, either in PDF or URLs.
- RAKIM - A web-based real-time virtual reference environment.
- Greenstone – Provides facilities of browser-based access, full-text and field-specific searching, use of Dublin Core and other metadata schema, advanced data

compression techniques that lower response times when searching large collections, a customisable interface based on a configuration file, multilingual interface, extensive use of plug-ins to convert documents in different formats such as MS Word, PDF, HTML, or email, administrative features that support access control and user activity logs etc.

- FEDORA – A software used for general-purpose repository system, gives organizations a flexible service oriented architecture for managing and developing their digital content.

Some of the other popular library Open Source Software has already been identified as: a) Koha, b) MyLibrary, c) Avanti, d) Eprints, etc.

6. Case Study: Visva-Bharati Library

Visva-Bharati (VB) Library has a good collection of nearly 8 lakhs documents in its Central Library along with 12 sectional libraries. It has taken the drive to built this library as a digital one by converting its own resources into digital form; by procuring born digitized documents, viz., e-journals, e-books, CD-ROM etc. and by sharing other institution's / organization's documents. Sharing is always reciprocal between libraries / institutions and for digital library, sharing depends on good access facilities through web pages and the regular maintenance of the existing web page / homepage. For the good maintenance of its homepage VB Library has decide to use OSS as it is using some OSS successfully in other library operations. The Redhat Linux 9.0 is serving as the operating system of the server machine for the proprietary software Unicode compliant Libsys version – 4.

For any digital library, web-based access facility is an essential way to provide any service to the end users. For doing so, an informative and easy-to-use library homepage is essential. Proper maintenance of that homepage is not an easy task. Here the term "proper" does not necessarily mean that the librarian or the webmaster will update pages periodically (a) on the basis of his choice and/or (b) the comments and the suggestions received from the end user, via e-mail or form posted in suggestion box if any. In our daily experience, we found that the frequency of the second option i.e., the suggestions or comments from the end user is not always surprising.

6.1 The Webalizer

Web server log file analysis plays an important role to get the user feedback indirectly. Page-hit counter, implemented in many web pages is an important metric, carries the information that how many times the web page is browsed by the users and can be generated from the log file analysis. Visva-Bharati is using the Webalizer, an open source web server log file analysis program. It can produce highly detailed, easily configurable usage reports in HTML format for viewing with any standard web browsers.

The important features of webalizer [6] is given below:

- It is written in C language, extremely fast and highly portable.
- It supports any standard common log file format server logs. In addition, several variations of the combined log file formats are supported, allowing statistics to be generated for referring sites and browser types as well.
- Generated reports can be configured from the command line, or by use of one or more configuration files.
- It supports multiple languages (almost in 34 different language) includes Chinese, Danish, Dutch, English, French, German, Greek and so on.
- Unlimited log file sizes and partial logs are also supported.
- Distributed under the GNU General Public License, complete source code is available, as well as in binary form for some popular platforms.

6.2 The Application

The Webalizer software has been strongly influenced the VB Library. The VB Library is using this software for usage analysis of the library's website, its pages. By analyzing the usage the library homepage or web pages are being updated regularly.

Fig. 1a & Fig 1b shows a sample screenshot of the Visva-Bharati library homepage. The diagram/graph/statistics are generated from VB Library server presently visible in university intranet by using the OSS, Webalizer. These are almost self-explanatory. Fig. 2a & Fig. 2b shows the summary of last 12 months usage report in graphical and in tabular form respectively.

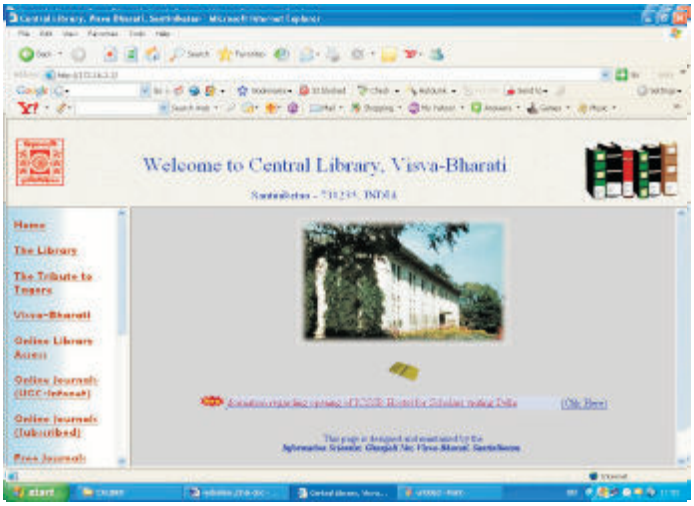

Fig. 1a A sample screenshot of VB Library homepage

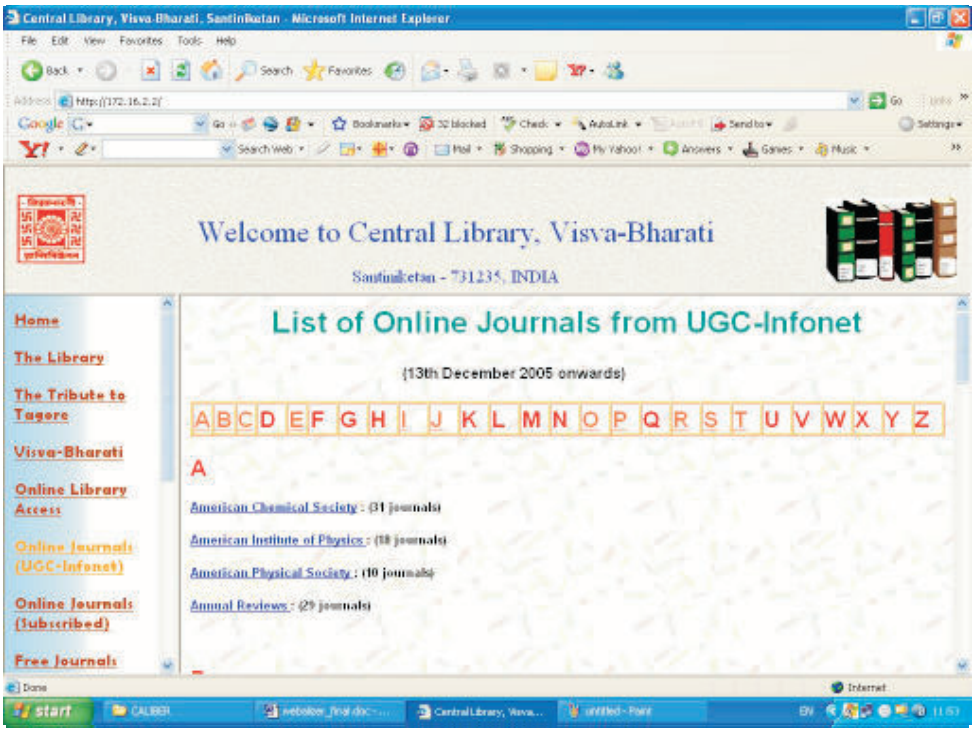

Fig. 1b A sample screenshot of VB Library homepage

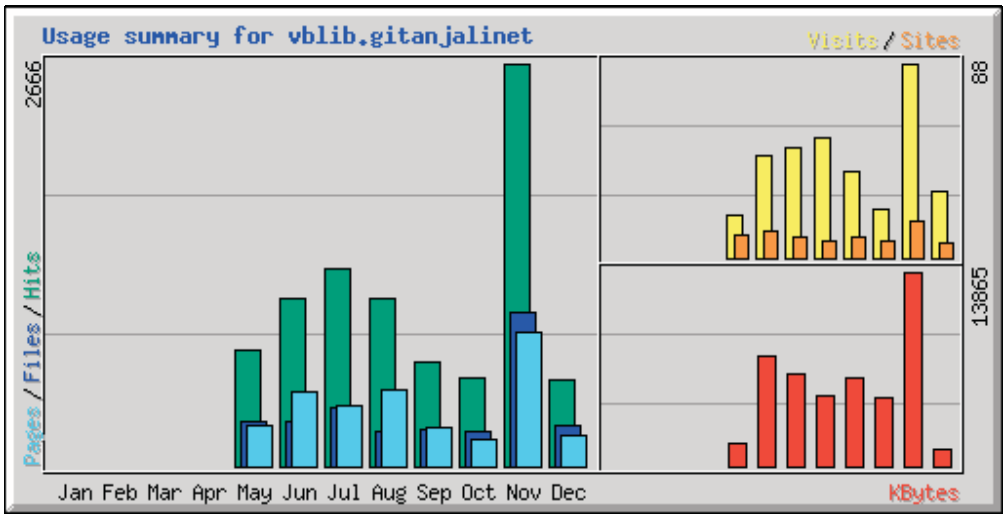

Fig. 2a Graphical Representation of summary of last 12 months

Summary of Month										
	Daily Avg					Monthly Totals				
Month	**Hits**	**Files**	**Pages**	**Visits**	**Sites**	**KBytes**	**Visits**	**Pages**	**Files**	**Hits**
Dec-06	21	10	7	1	7	1258	30	208	271	568
Nov-06	91	35	30	3	17	13865	88	890	1024	2666
Oct-06	32	13	9	1	8	4846	22	178	235	585
Sep-06	27	9	10	1	9	6360	39	261	246	692
Aug-06	36	7	16	1	8	5027	54	504	229	1116
Jul-06	42	12	13	1	9	6643	50	406	392	1309
Jun-06	37	9	16	1	12	7859	46	499	293	1113
May-06	84	33	29	2	10	1694	19	265	297	762
Totals						**47552**	**348**	**3211**	**2987**	**8811**

Fig. 2b Summary of last 12 months in tabular form

The detailed statistics of web use i.e., date-wise, hour-wise etc. can be generated for any particular month just by clicking that month in Fig. 2b. Here for an example we are showing the details for the month of Nov 2006. Fig. 3 shows the monthly statistics for Nov 2006. Fig. 4a & Fig. 4b shows the daily usage in graphical and in tabular form. If one wishes to have the statistics of hourly usage of VB Library's web use, it can also be possible with the help of the OSS, Webalizer as shown in Fig. 5a and 5b. Lastly Fig. 6 shows the URL statistics.

Monthly Statistics for November 2006		
Total Hits	2666	
Total Files	1024	
Total Pages	890	
Total Visits	88	
Total KBytes	13865	
Total Unique Sites	17	
Total Unique URLs	59	
Total Unique Referrers	31	
Total Unique User Agents	9	
	Avg	**Max**
Hits per Hour	3	150
Hits per Day	91	368
Files per Day	35	197
Pages per Day	30	107
Visits per Day	3	12
KBytes per Day	478	3388

Fig. 3 Monthly statistics for Nov 2006

Daily Usage Statistics:

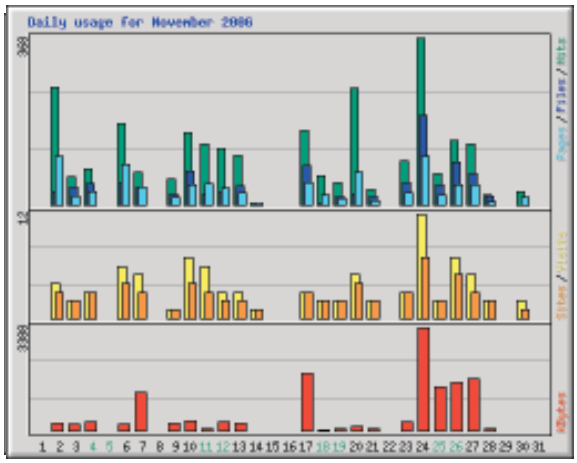

Fig. 4a Daily Usage Statistics in graphical form

Daily Statistics for November 2006

Day	Hits		Files		Pages		Visits		Sites		KBytes	
2	259	9.71%	29	2.83%	107	12.02%	4	4.55%	3	17.65%	237	1.71%
3	62	2.33%	40	3.91%	20	2.25%	2	2.27%	2	11.76%	186	1.34%
4	78	2.93%	46	4.49%	30	3.37%	3	3.41%	3	17.65%	267	1.92%
5	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	180	6.75%	50	4.88%	90	10.11%	6	6.82%	4	23.53%	209	1.51%
7	72	2.70%	39	3.81%	39	4.38%	5	5.68%	3	17.65%	1252	9.03%
8	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
9	57	2.14%	24	2.34%	17	1.91%	1	1.14%	1	5.88%	236	1.70%
10	159	5.96%	75	7.32%	43	4.83%	7	7.95%	4	23.53%	299	2.15%
11	134	5.03%	23	2.25%	47	5.28%	6	6.82%	3	17.65%	69	0.50%
12	125	4.69%	29	2.83%	37	4.16%	3	3.41%	2	11.76%	261	1.88%
13	110	4.13%	41	4.00%	27	3.03%	3	3.41%	2	11.76%	254	1.83%
14	5	0.19%	0	0.00%	4	0.45%	1	1.14%	1	5.88%	0	0.00%
15	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
16	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
17	165	6.19%	87	8.50%	48	5.39%	3	3.41%	3	17.65%	1865	13.45%
18	63	2.36%	5	0.49%	22	2.47%	2	2.27%	2	11.76%	26	0.19%
19	48	1.80%	20	1.95%	13	1.46%	2	2.27%	2	11.76%	58	0.42%

20	257	9.64%	24	2.34%	72	8.09%	5	5.68%	4	23.53%	119	0.86%
21	34	1.28%	18	1.76%	9	1.01%	2	2.27%	2	11.76%	62	0.45%
22	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
23	96	3.60%	48	4.69%	26	2.92%	3	3.41%	3	17.65%	270	1.95%
24	368	13.80%	197	19.24%	106	11.91%	12	13.64%	7	41.18%	3388	24.44%
25	69	2.59%	45	4.39%	22	2.47%	2	2.27%	2	11.76%	1444	10.42%
26	142	5.33%	95	9.28%	43	4.83%	7	7.95%	5	29.41%	1596	11.51%
27	133	4.99%	69	6.74%	45	5.06%	5	5.68%	3	17.65%	1705	12.30%
28	22	0.83%	20	1.95%	7	0.79%	2	2.27%	2	11.76%	60	0.44%
29	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
30	28	1.05%	0	0.00%	16	1.80%	2	2.27%	1	5.88%	0	0.00%

Fig. 4b Daily Usage Statistics in tabular form

Hourly usage statistics:

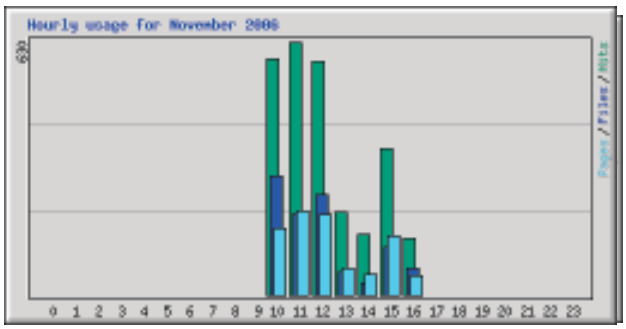

Fig. 5a Hourly usage statistics in graphical form

Hourly Statistics for November 2006												
Hour	Hits			Files			Pages			KBytes		
	Avg	Total		Avg	Total		Avg	Total		Avg	Total	
0	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
1	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
2	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
3	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
4	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
5	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
6	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
7	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
8	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%

9	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
10	20	589	22.09%	10	295	28.81%	5	165	18.54%	78	2260	16.30%
11	21	630	23.63%	7	204	19.92%	7	211	23.71%	120	3469	25.02%
12	20	581	21.79%	8	251	24.51%	6	200	22.47%	86	2505	18.07%
13	7	207	7.76%	1	57	5.57%	2	67	7.53%	17	485	3.50%
14	5	152	5.70%	1	30	2.93%	1	52	5.84%	5	141	1.02%
15	12	366	13.73%	4	119	11.62%	5	146	16.40%	120	3490	25.18%
16	4	141	5.29%	2	68	6.64%	1	49	5.51%	52	1514	10.92%
17	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
18	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
19	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
20	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
21	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
22	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%
23	0	0	0.00%	0	0	0.00%	0	0	0.00%	0	0	0.00%

Fig. 5b Hourly usage statistics in tabular form

URL Statistics:

Top 21 of 59 Total URLs					
#	**Hits**		**KBytes**		**URL**
1	162	6.08%	31	0.22%	/
2	147	5.51%	59	0.43%	/top_index.html
3	145	5.44%	74	0.53%	/left_index.html
4	137	5.14%	109	0.79%	/body_index.html
5	69	2.59%	114	0.82%	/announcement.html
6	22	0.83%	103	0.75%	/online_journals_UGCInfonet.html
7	20	0.75%	96	0.69%	/online_journals_VB.html
8	20	0.75%	20	0.14%	/open_archives.html
9	19	0.71%	12	0.08%	/imp_links.html
10	18	0.68%	1599	11.53%	/e_Books.html
11	18	0.68%	10	0.07%	/tributeToTagore.html
12	17	0.64%	25	0.18%	/library.html
13	16	0.60%	69	0.50%	/administration.html
14	15	0.56%	65	0.47%	/contacts.html
15	14	0.53%	7886	56.88%	/freeJournals.html
16	14	0.53%	10	0.07%	/other_library.html
17	13	0.49%	62	0.45%	/usage/
18	9	0.34%	270	1.95%	/ICSSRhostel.pdf

19	5	0.19%	137	0.99%	/usage/usage_200610.html
20	4	0.15%	278	2.00%	/usage/usage_200611.html
21	1	0.04%	76	0.55%	/usage/usage_200609.html

Fig. 6 URL Statistics

From the above detailed analysis webmaster is now able to get the user feedback for proper maintenance of the web pages. For example from Fig. 5a & 5b, it is seen that the server is accessing by the users almost from 9.30 to 16.30 hrs. So, for any reason if the webmaster needs to shutdown the server he can exclude that time period. From Fig. 6 It is also seen that except the index pages the three maximum accessed pages are the announcement page, page containing the complete list of online journals from UGC Infonet programme and the list of online journals purchased from Visva-Bharati Library respectively. So these pages should be updated on urgent basis rather than the other pages. In this way server log file analysis program plays an important role for proper maintenance of the home page by exploiting the users feedback indirectly.

7. Conclusions

In conclusion it may be said that the construction of Digital Library is a challenging process. Now-a-days library users of university library are very much familiar with Internet and its usage. Users of university library community access different types of information even from outside of the library and beyond library hours. Library homepage should be well linked for internal items of information or resources. Library web pages should be well linked to external websites for accessory external resources so that library clientele may get current information on required needs. For this purpose library web pages / homepages be updated on regular basis. The Webalizer software can help the librarian or Webmaster to cope-up this update on the basis of need of the library clientele.

References

1. http://www.arl.org/sunsite/definition.html
2. International Encyclopedia of Information and Library Science. 2nd. ed. Ed. By John Feather and Paul Sturges. London: Routledge, 2003. p.468.
3. The Keystone Principles: An action plan for values-based librarianship. C&RL News, February 2000. Vol. 61 No. 2
4. http://www.opensource.org
5. http://www.webopedia.com
6. http://www.mrunix.net/webalizer/

BIOGRAPHY OF AUTHORS

Mr. Achintya K. Mandal received the B.Sc. degree in physics and the Post Graduate Diploma in computer science and application from the University of Burdwan, in 1996 and 1999, respectively, and the Master of Computer Application from Visva-Bharati University, Santiniketan, in 2002. From September 2004 he is working at Visva-Bharati University, as Information Scientist. He has published various article in national and International Publications.

Email : achitanya99@yahoo.co.in

Dr. Subodh Gopal Nandi, M.A., MLIS, Ph.D. is now working as Librarian at Visva-Bharati, Santiniketan, West Bengal, India since August 2004. Has more than 27 years of Professional experience in University Libraries at Burdwan University and Visva-Bharati. He has published 30 articles in different LIS Journals and Conference Proceedings.

Email : sgnandi2@yahoo.co.in