IASET: International Journal of Library &
Educational Science (IASET: IJLES)
ISSN(P):Applied; ISSN(E): Applied
Vol. 2, Issue 1, Jan - Jun 2016, 33-44

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

OPEN SOURCE DIGITAL LIBRARY SOFTWARE'S: EXCEPTIONAL INDICATION TO DSPACE, EPRINTS AND GREENSTONE – AN OVERVIEW

G. SIVASUBRAMANIAN[1] & P. GOMATHI[2]

[1]Research Scholar, & Assistant Librarian, SNR Son's College, Coimbatore, India

[2]Assistant Professor, Depertment of Library And Information Science, Periyar University, Salem, India

ABSTRACT

Digital Library made a remarkable change in the field of Library and Information Science the user interface. In the present scenario the open source software awareness is highly increased in the world, especially among the librarians. This paper presents a study of three open source digital library management software used to assimilate and disseminate information to world audience. The methodology followed involves online survey and study of related software documentation and associated technical manuals. Open source digital library packages are gaining popularity nowadays. To build a digital library under economical conditions open source software is preferable. Examples of Open Source Softwares are: Apache, FreeBSD, GIMP, GNOME, KDE, LINUX, Mozilla, My SQL, D-Space. This paper briefly discussed in D-Space, Greenstone and E-Print.

KEYWORDS: Open Source, Digital Library, Digital Library Management Software

INTRODUCTION

Open source defines a method of software development, that harnesses the power of distributed peer review and transparency of progress. This technique helps to provide better quality software's having higher reliability, flexibility with lower cost, and an end to the traditional vendor lock-in. The source code and rights that were normally reserved for copyright holders are now being provided under a free software license that permits developers / users to study, change, improve and at times also to distribute the software.

Digital library refers to a collection that constitutes electronic resources, accessible through the World Wide Web. It often contains electronic versions of books, photographs, videos that are owned by a "physical" library. Open source digital library software presents a system for the construction and presentation of information collections.

What is Open Source Software

The basic ideas behind open source is very simple. When a program can read, redistribute, and modify the source code for a piece of software. The software evolves. People improve it, people adapt it, people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing.

Why Open Source Software for Digital Library

Open source software solutions standard dominating the information industry, particularly the academic libraries. Linux is a freeware and powerful operating system packed with lots of utilities and tools that are used full for developing information Retrieval system.

- It is a multi user or multiprocessing support.
- High performance and better networking support.
- Compiling source code and tune the required system.

The term digital library began to be held in the early 1990s, as universities and other institutions in most of the developed countries began to build digital collections.

The Open Source Software applications for the library and Information management that will be discussed in this paper are:

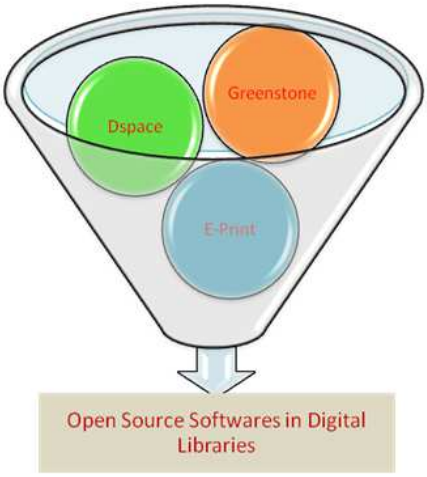

Figure 1

DIGITAL LIBRARY MANAGEMENT SYSTEMS

Digital Libraries have greatly evolved during the last few years. They are no longer only the digital counterpart of physical libraries (or physical museums, video achieves, etc.) rather they are intricate networked systems capable of supporting communication and collaboration among different, worldwide distributed user communities. The digital Library management system evolved with the inception of Digital Libraries**.** The digital Library management system provides the appropriate framework both for the production and administration of the Digital Library System by incorporating functionality essentially fundamental to Digital Libraries, and also provides provision for integration of additional software that provides more refined and advanced functionality.

Digital Library can thus be established by setting up and deploying a Digital Library Management System and then loading or harvesting content. This approach largely simplifies and reduces the effort required to set up a Digital Library that promises a guaranteed better quality of service. These generic systems have started to appear from the second half of 1990's even though implementing the devised DLMS features only to some extent.

THE DLMS' (DIGITAL LIBRARY MANAGEMENT SYSTEM)

I - Dspace

The DSpace is a joint project of the MIT Libraries and HP labs. It is a digital asset management system that allows institutions, such as libraries to collect, archive, index, and disseminate the scholarly and intellectual efforts of a community. Written with a combination of technologies by MIT, it is primarily used to capture bibliographic information

describing articles, papers, theses, and dissertations. DSpace is adaptable to different community needs. Interoperability between systems is built-in and it adheres to international standards for metadata format. Being an open source technology platform, DSpace can be customized to extend its capabilities. Some of its characteristics as shown in the DSpace documentation are as:

- It is a service model for open access and/or digital archiving for perennial access.
- Provides a platform to frame an Institutional Repository and the collections are searchable and retrievable by the Web.
- Helps to make available institution-based scholarly material in digital formats. The collections will be open and interoperable.

The organization of data model in DSpace is intended to mirror the structure of the organization using the DSpace. Each DSpace site is divided into communities, which can be further divided into sub-communities reflecting the typical university structure of college, department, research centre, or laboratory. Communities contain collections, which are groupings of related content. A collection may appear in more than one community. Each collection is composed of items, which are the basic archival elements of the archive. Each item is owned by one collection. Additionally, an item may appear in additional collections; however every item has one and only one owning collection. Items are further subdivided into named bundles of bitstreams. Bitstreams are, as the name suggests, streams of bits, usually ordinary computer files. Bitstreams that are somehow closely related (for example HTML files and images that compose a single HTML document) are organized into bundles.

As specified by **Robert Tansley, Mick Bass, Margret Branschofsky, Grace Carpenter, Greg McClellan, David Stuve (05-Oct-2005)** the bundles most items tend to included the following:

- ***ORIGINAL:*** The bundle contains the original, deposited bitstreams.
- ***THUMBNAILS:*** Thumbnails of any image bitstreams.
- ***TEXT :*** It includes extracts full-text from bitstreams in ORIGINAL, for indexing.
- ***LICENSE:*** It contains the deposit license that the submitter granted the host organization; putting it

differently it specifies the rights that the hosting organizations have.

- ***CC_LICENSE:*** It contains the distribution license, if any (a Creative Commons license) associated with the item. This license specifies what end users can do with the downloaded content.

Reason to Use DSPACE

- Largest community of users and developers worldwide
- Free Open Source Software
- Completely customizable to fit your needs
- Used by educational, government, private and commercial institutions
- It can be installed out of the box

- It can manage and preserve all types of digital content.

The Features of Dspace as Digital Management Software are as Follows

Figure 2

- **Authentication:** DSpace allows contributors to limit access to items in DSpace, at both the collection and the individual item level. The mechanism whereby the system securely identifies its users.
- **Authorization:** The mechanism by which a DSpace determines what level of access a particular authenticated user should have to secure resources controlled by the system is done by keeping access control policies that allow it to understand what credentials are required (if any) to undertake particular actions upon particular resources.
- **Non-dynamic HTML document Support:** As mentioned by **Tansley R, et al (2005)** in the documentation, DSpace simply supports uploading and downloading of bitstreams as-is. This mechanism is good for majority of file –formats like PDF, Word Document and so on. As far as HTML documents are concerned, they are complicated in the sense they consist of several files and are cross-linked with each other. This has important ramifications when it comes to digital preservation. Web pages also link to or include content from other sites, often imperceptible to the end-user. Thus, in a few year's time, when someone views the preserved Web site, they will probably find that many links are now broken or refer to other sites than are now out of context.

However, later on, when another user tries to view that HTML, their browser might not be able to retrieve the included image since it may have been removed from the external server. Hence the HTML will seem broken. The links preserved for images, videos, etc. are preserved as relative links. Any absolute link is stored as is and will continue to link the source as long as it is live, and will eventually change or disappear.

- **OAI-PMH Support:** The OAI_PMH is a protocol for metadata harvesting. This allows sites to programmatically retrieve or 'harvest' the metadata from several sources, and offer services using that metadata, such as indexing or linking services. DSpace exposes the Dublin Core metadata for items that are public (anonymously) accessible. Additionally, the collection structure is also exposed via the OAI protocol's 'sets' mechanism. OCLC's open source OAICat framework is used to provide this functionality.

- **Object Management:** The process of item ingestion in DSpace is via a web interface or the batch item importer. In workflow process for item submission will initiate, depending on the configuration of the collection. The workflow process may contain one or more steps as per the user need. The collection and communities in DSpace are created via web interface.
- **Import & Export:** Import & Export for Communities, Collections and Items is supported by DSpace. It also includes batch tools to import and export items in a simple directory structure, where the Dublin Core metadata is stored in an XML file. This may be used as the basis for moving content between DSpace and other systems.
- **Statistics:** Statistics are provided for administrative usage. Statistical reports/summary can be used for performing analysis on repository, providing information like number of items uploaded, searched, number of e-people registered with the system etc.
- **Handle System:** To help in creation of persistent identifier for every item DSpace makes use of Handle system's global resolution feature. DSpace requires a storage and location independent mechanism for creating and maintaining identifiers. DSpace uses the CNRI Handle System for creating these identifiers. A Handle server runs as a separate process that receives TCP requests from other Handle servers, and issue resolution requests to a global server or servers if a Handle entered locally does not correspond to some local content.
- **Customization & types of document supported:** DSpace allows customization to accommodate the multidisciplinary and organizational needs of a large institution. Albeit DSpace provides a flexible data object model. It does not allow construction of very different objects with independent metadata sets due to its The system can function with many file types, including: PDF, HTML, JPEG, TIFF, MP3, and AVI etc.
- **Standards Compliance:** The default configuration permits DSpace to store the metadata of an item in the Dublin Core Metadata Schema. This ensures that data can be exchanged with other standards compliant system, such as MARC21. MARC is an acronym for Machine-Readable Cataloguing.
- **Optimized Search & Browse:** As per **Bass, M J et al (n.d)**, the system allows end-users to discover content in a number of ways, including:

 By default indexing of basic metadata set qualified DC is provided by DSpace. While as indexing of other metadata sets is provided by the Jakarta Lucene search engine. Apache Lucene is written in Java and provides high-performance, full-featured text search engine library. It provides technology for any application that requires full-text search, especially cross-platform **(Lucene, 2012)**. Lucene supports fielded search, stemming & stop words removal. By default Browsing in DSpace is by title, author, and date field.

 Via external reference, such as a CNRI Handle. A persistent identifier used for every bitstreams of every item.

II - GREENSTONE

The Greenstone Digital Library Software is a project from New Zealand that provides a new way of organizing information and making it available over the Internet. Collections of information comprise large numbers of documents (typically several thousand to several million), and a uniform interface is provided to them. Libraries include many collections, individually organized, though bearing a strong family resemblance.

A configuration file determines the structure of a collection. Existing collections range from newspaper articles to technical documents, from educational journals to oral history, from visual art to videos, from MIDI pop music collections of ethnic folk songs.

A typical digital library built with Greenstone will contain many collections, individually organized. Easily maintained, collections can be augmented and rebuilt automatically. There are several ways to find information in most Greenstone collections. For example, you can search for particular words that appear in the text, or within a section of a document. The word search is provided as Greenstone constructs full-text indexes from the document text, that is, indexes that enable searching for any words in the full text of the document.

First, documents, shown at the bottom of the figure, are imported into the XML-compliant Greenstone Archive Format. Then the archive files are built into various searchable indexes and a collection information database that includes the hierarchical structures that support browsing. When this is done, the collection is ready to go online and respond to requests for information.

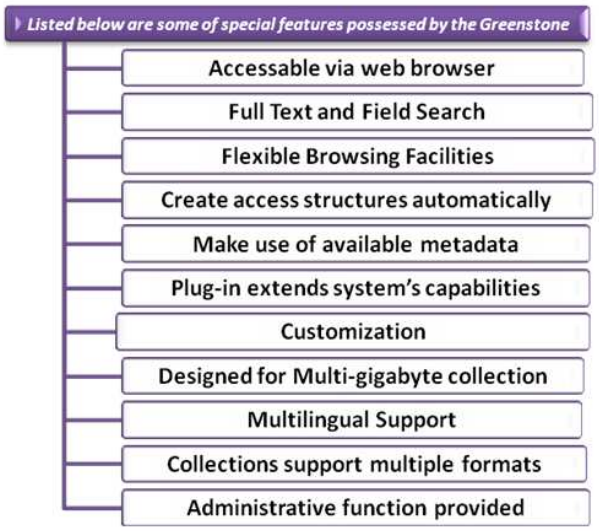

Figure 3

- **Accessible via web browser:** Collections are accessed through a standard web browser (Netscape or Internet Explorer) and combine easy-to-use browsing with powerful search facilities.
- **Full Text and Field Search:** The user can search the full text of the documents, or choose between indexes built from different parts of the documents. For example, some collections have an index of full documents, an index of sections, an index of titles, and an index of authors, each of which can be searched for particular words or phrases. Results can be ranked by relevance or sorted by a metadata element.
- **Flexible browsing facilities:** The user can browse lists of authors, lists of titles, lists of dates, classification structures, and so on. Different collections may offer different browsing facilities and even within a collection, a broad variety of browsing interfaces are available. Browsing and searching interfaces are constructed during the building process, according to collection configuration information.
- **Create access structures automatically:** The Greenstone software creates information collections that are very

easy to maintain. All searching and browsing structures are built directly from the documents themselves. No links are inserted by hand, but existing links in originals are maintained. This means that if new documents in the same format become available, they can be merged into the collection automatically. Indeed, for some collections this is done by processes that wake up regularly, scout for new material, and rebuild the indexes—all without manual intervention.

- **Make use of available metadata:** Metadata, which is descriptive information such as author, title, date, keywords, and so on, may be associated with each document, or with individual sections within documents. Metadata is used as the raw material for browsing indexes. It must be either provided explicitly or derivable automatically from the source documents. The Dublin Core metadata scheme is used for most electronic documents; however, provision is made for other schemes.
- **Plug-in extends the system's capabilities:** In order to accommodate different kinds of source document, the software is organized in such a way that "plug-in" can be

Written for new document types. Plug-in currently exist in plain text, html, Word, PDF, PostScript, E-mail, some proprietary formats, and for recursively traversing directory structures and compressed archives containing such documents.

- **Customization:** The Greenstone allows customization of presentation of the collection that are based on EXtensible Stylesheet Language transformation (XSLT) and other agents that govern the definite functions of Digital library. The architecture of Greenstone purvey:
  - A back end that provides services to manage documents and collections.
  - A front end that provides a web based interface for searching and presentation of documents,

collections.

- **Designed for Multi-gigabyte collection:** Collections can contain millions of documents, making the Greenstone system suitable for collections up to several gigabytes.
- **Multilingual Support:** Unicode is used throughout the software, allowing any language to be processed in a consistent manner. To date, collections have been built containing French, Spanish, Maori, Chinese, Arabic and English. On-the-fly conversion is used to convert from Unicode to an alphabet supported by the user's web browser.
- **Collections support multiple formats:** Greenstone collections can contain text, pictures, audio and video clips. Most non-textual material is either linked into the textual documents or accompanied by textual descriptions (such as figure captions) to allow full-text searching and browsing.
- **Administrative function, provided:** An "administrative" function enables specified users to authorize new users to build collections, protect documents so that they can only be accessed by registered users on presentation of a password, examine the composition of all collections, and so on. Logs of user activity can record all queries made to every Greenstone collection.

III - EPRINTS

EPrints is free software developed by the “University of Southampton, England”. EPrints repository collects, preserves and disseminates in digital format the research output created by a research community. It enables the community to deposit their preprints; poster prints and other scholarly publications using a web interface, and organizes these publications for easy retrieval. It is the world's first, most widely used, and by far the most functional of all the available OA IR software’s.

It is created for and specifically focused on OA functionality. EPrints is an extensible content management system. It has been extensively configured to accommodate the needs of academics and researchers amid at dissemination and reporting, but it could be easily used for other things such as images, research data, audio archives - anything that can be stored digitally, but you made more changes to the configuration. EPrints is OAI-compliant. It is highly configurable to achieve diverse needs, built on a coding platform that is amendable to rapid development.

The real strength of EPrints lies in its ease of use for both end-users and administrators. Submitting documents in EPrints are very straightforward. Users are taken through the submission process one step at a time and asked to provide metadata information along with an electronic copy of the document. Users can simply enter metadata such as document type, title, author name, date, etc. via a web form, no knowledge of HTML or XML is required.

The documents in an EPrints archive can be indexed to allow retrieval by online search engines like Google, which helps to ensure greater access to, and greater dissemination of any items uploaded to the archive. Searching is fairly limited in EPrints. As mentioned earlier, Boolean searching is not supported. It is also quite easy to run a search that yields no results. For end users accustomed to modern search engines and databases it might be discouraging to get an unsuccessful search with no suggestions for alternative search strategies.

In EPrints there is no such strict structural division into sections and collections that are still playing an important role, for example, to narrow the search to the repository. The idea is that all records are equivalent and do not form a hierarchy. Such objects, as an element, a collection of files, file, are similar to the analogous into DSpace. The element is also the fundamental unit of storage and contains all metadata, allowed for the external use. The hierarchical structure of elements is significantly different. DSpace uses a more rigid system, although it covers most of the needs of the repository. The EPrints allow you to create a more complex hierarchy based on different external representations.

Listed Below Are Some Of Special Features Possessed By The Eprints

- **Accessibility via web browser:** EPrints provides a web based interface that makes it easy to use and administer.
- **Full Text and Field Search:** Searching is based on metadata not full text based search is supported by EPrints **[27]**. Searching in EPrints allows scanning each of the metadata field types in the database by using simple or advanced search. Any metadata field can be searched with fine granularity by SQL querying the database.
- **Administrative function, provided:** EPrints archive can use any metadata schema as being provided by the administrator. The administrator decides what metadata fields are held about each EPrints item **[28]**. This is specified in three or four stages:
    - The Definition of a maximal set of metadata fields that should be stored (e.g. author, title, journal, journal volume, etc.).

- Definition of different types of EPrints (e.g. refereed journal article, thesis, technical report, unpublished preprint, etc.).
- Specification for each type which metadata fields should be stored, and which of those fields are mandatory.
- Decide how these metadata fields should be projected into the Open Archives world. (If necessary, interoperability can be switched off, but this is strongly discouraged.)

- **Open Source Software:** EPrints uses traditional technologies and runs on pure Open Source systems. It uses MySQL, Apache database and web server. MySQL is the world's most popular open source database, recognized for its speed and reliability and Apache has been the most popular web server on the Internet since April of 1996.Eprints is programmed by using the script language "Perl", that is low level but powerful.
- Three **user roles:** administrator, editor and author.
- Administrator role controls all back-end options such as organization of records, web interface appearance and functionality, and all other server-side settings.
- Editor role reviews submissions before they are published online and may edit metadata on submissions to maintain consistency or correct errors.
- Author role allows submission of documents and management of previously submitted documents.
- **OAI-PMH Support:** EAS is fully interoperable with OAI (Open Archives Initiative) Protocol for Metadata Harvesting **[29]**. The Open Archives Protocol allows sites to programmatically retrieve or 'harvest' the metadata from several sources, and offer services using that metadata, such as indexing or linking services. Such a service allows e-prints servers create the potential for a global network of cross-searchable research information, by allowing the contents of servers around the world searched simultaneously by using the OAI (Open Archives Initiative) protocol.
- **Multilingual Support:** Unicode is used throughout the software, allowing any language to be processed in a consistent manner **[30]**.
- **File formats supported:** Functions with many file types, including: PDF, HTML, JPEG, TIFF, MP3, and AVI etc. Metadata schema can be tailored to meet the requirements **[27]**.
- **Statistics:** Statistics are provided for administrative usage. The Statistical reports/summary can be used for performing analysis on repository **[30]**.
- **Customization:** The EPrints data modal consist of user defined metadata. In order to export data in other format plug-ins can be written. For developers who wish to access the core Digital Library functionality Core API in Perl language is provided **[30]**.
- **Item previews in EPrints:** Thumbnail preview of documents and images is generated automatically upon file upload **[30]**.

Table 1: Based on Above Discussion a Product Comparison Table for DSpace, EPrints & Greenstone is Drafted Below

Feature	DSpace	EPrints	Greenstone
Year of Creation	**2002**	**2000**	**1997**
License cost	Free	Free	Free
Product Type	Software	Software	Software
Update cost	Free	Free	Free
Resource Identifier	CNRI Handles	No	OAI Identifier
OAI-PMH	Yes	Yes	Yes
Supported Item Types (Storage *and* rendition)	Can store and manage all types of content	Can store and manage all types of content	Can store and manage all types of content
Metadata formats	Dublin Core, Qualified DC, METS	Dublin Core, METS	Dublin Core, Qualified DC , METS, NZGLS (New Zealand Government Locator Service), AGLS (Australian Government Locator Service)
User interface functions	End user depositions, Multilingual support.	End user depositions, Multilingual support.	End user deposition, Multilingual support.
Thumbnail Preview	Images	Images, Audio, Video	Images, Audio, Video
Searching Capabilities	Field Specific, Boolean Logic, Sorting options	Field Specific, Sorting options	Field Specific, Boolean Logic
Syndication	RSS, ATOM	RSS, ATOM	---
User Authentication	LDAP Authentication, *Shiboleth Authentication*	LDAP Authentication	User Groups
Statistical reporting	Count of Full Records	Count of Full records	Count of Full records
Software Platforms	Linux or Unix, Solaris, Windows	Linux, Unix, Windows,	Linux, Unix, Windows, Mac-OS
Databases	Oracle, PostgreSQL	MySQL, Oracle, PostgreSQL, Cloud.	Its Own
Programming Language	Java and JSP	Perl	C++, Perl, Java
Web Server	Apache and Tomcat	Apache	Apache/IIS
Associated Software	Jave, Apache, PostgreSQL, or Oracle	Linux or Unix, Apache, Perl	Apache, PERL, GNU C++ Compiler, JAVA, GNU Database manager
Machine-to-Machine Interoperability.	OAI-MHP,OAI-ORE, SWORD, SWAP	OAI-MHP,OAI-ORE, SWORD, SWAP,RDF	Z39.50, OAI-MHP
License	GNU	BSD	GNU

CONCLUSIONS

The Digital Library Management softwares (DLMS) present an easy to use, customizable architecture to create online digital libraries. It is difficult to propose one specific DLMS system as the most suitable for all cases. The study can be used as a reference guide by any organization or institution to decide which one will be ideal for creating and showcasing their digital collection. The choice usually depends on type/format of material, distribution of material, software platform and time frame etc for setting up a Digital Library.

REFERENCES

1. Chudnov Daniel (1999), “Open Source Software: the features of library systems”, *Library Journals*, Vol 124 (13), pp 40-43.
2. Arora jagadish (2006), “Building Digital Libraries: an overview”, *DESIDOC Bulletin of Information Technology*, Vol 21(6), pp 3-24.

3. O'Mahony, S. (2003), "Guarding the commons: how community managed software projects protect their work", *Research Policy*, Vol *32*(7), pp 1179-1198.
4. Witten, I. H., Bainbridge, D., & Boddie, S. J. (2001), "Greenstone: Open-source digital library software with end-user collection building", *Online information review*, Vol *25* (5), pp 288-298.
5. Arora, Jagadish (2006), "Building Digital Libraries: an overview", *DESIDOC Bulletin of Information Technology*, Vol 21 (6), pp 3-24.
6. Morgan, Eric Lease (2002), "Possibilities for Open Source Software in Libraries", *Information Technology and Libraries*, Vol 21 (1), pp 8-28.
7. Gullik, Robert (2002), "Digital Library in New Era", *Digital Libraries*, Vol 26 (2), pp 35-41.
8. Meitei LS, Devi P (2009), "Open Sources Initiative in Digital Preservation: The Need for an Open Source Digital Repository and Preservation System", 7th International CALIBER -2009, Pondicherry University, Puducherry.
9. Shahkar Tramboo. et. al (2012), "A Study on the Open Source Digital Library Software's: Special Reference to DSpace, EPrints and Greenstone", *International Journal of Computer Applications*,Vol (59)16, pp 1-9.
10. Goutam Biswas and Dibyendu Paul (2010), "An evaluative study on the open source digital library softwares for institutional repositories: Special reference to Dspace and greenstone digital library", *International Journal of Library and Information Science*, Vol 2 (1) pp 1-10.
11. Biswas G, Paul D (2008). "NewGenLib, The First Indian Open Source Software: a Study of Its Features And Comparison With Other Software", *23rd National Seminar of IASLIC held at Bose Institute on Library Profession in Search of a New Paradigm*, Kolkata pp. 10-13.
12. Sharad Kumar Sonkar, Veena Makhija, Ashok Kumar and Mohinder Singh (2005), "Application of Greenstone Digital Library (GSDL) Software in Newspapers Clippings", *DESIDOC Bulletin of Information Technology*, Vol 25(3) pp 9-17.