# APPLICATION OF GREENSTONE DIGITAL LIBRARY SOFTWARE (GSDL) IN COLLECTION MANAGEMENT OF DIGITIZED RESEARCH ARTICLES

**Poonam V.Tapkir***

Librarian, K.T.S.P.Ms Hutatma Rajguru Mahavidyalaya, Rajgurunagar, Pune, India.

**A B S T R A C T**

Greenstone Digital Library Software (GSDL) is a complete suit for creating, building as well as distributing digital library collections. It specifically designed to be highly extensible and customizable. It adds new technology to the digital world to communicate and collaborate digital content of library to the end users. This paper interprets complete picture of GSDL software and shows how it is useful to create digital library collection of journals to make its more customized with the new Greenstone Librarian Interface. It has a great feature which allow user to create collection and metadata for documents, build it and put them in place for users to view. At the most advanced level, the Librarian Interface gives expert users interactive access to the full power of Greenstone, which could formerly be tapped only by running Perl scripts.

## INTRODUCTION

Nowadays we are living into a digital era. And more and more information is published into a digital format which has variety of aspect as well as formats like text, image, audio or video, that would permit easy access by information users. It's become tough job for librarian to organize and provide such a varied form of knowledge to its user. To develop and manage digital collection need of digital library were get come out. Digital libraries do replace the concept of a library as a repository of knowledge with the notion of it as a provider of information services to users. In this age of exponential growth of information materials in print as well as digital form, both physical and digital libraries are essential and can work in unison, supplementing each other.[1.] In this paper research scholar creates collection of journal article in Greenstone digital library software which will help to build collection of digital material and user can trace specific information of its need.

The Greenstone digital library software (GSDL) is an open-source system for the construction and presentation of information collections. It builds collections with effective full-text searching and metadata-based browsing facilities that are attractive and easy to use. Moreover, they are easily maintained and can be augmented and rebuilt entirely automatically. The system is extensible: software "plugins" accommodate different document and metadata types5. The aim of the Greenstone software is to empower users, particularly in universities, libraries, and other public service institutions, to build their own digital libraries. Interfaces available for the Greenstone digital library software: the four "core" languages English, French, Spanish, and Russian. The Greenstone "Collector" is an interactive subsystem for managing and accessing collections[2].

### Objective of this paper

The objective of this paper is to introduce library professionals with greenstone digital library software (GSDL) and shows most essential information about building up a digital library of collection containing journal articles in systematic style in windows operating system

### Literature review

Sonkar,Sharad Kumar (2005) discusses in detail all the issues related to the development of digital library of newspaper clippings and implementation of 'Greenstone Digital Library' software in developing such collection. Umesha Naik (2006) highlights the comparison, features, function and usability of OSS like Greenstone.Digital Library, DSpace and Ganesha Digital Library.Anuradha K.T.(2009) enable users for full text search feature in a widely used open source library automation package viz., Koha, by integrating it with an open source digital library software viz., Greenstone Digital Library Software (GSDL), by making use of Search and Retrieval by URL (SRU) feature available in both Koha and GSDL.K Rajasekharan

*✉ Corresponding author:*  **Poonam V.Tapkir**
Librarian, K.T.S.P.Ms Hutatma Rajguru Mahavidyalaya, Rajgurunagar, Pune, India

(2007) has created a Self-Instructional Guide for Beginner's to Build up a Digital Library with Greenstone.

### Digital Library management systems

Digital Libraries have greatly evolved during the last few years. They are no longer only the digital counter part 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. Digital Library management system evolved with the inception of Digital Library [3]. Digital Library management system provides the appropriate framework both for the production and administration of 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 major characteristics that distinguish them from each other are the class of functionality offered, the type of object model for information being supported, and the openness of their architecture's. The DLMS (Digital Library Management System) available are commercial as well as open source. But, Open Source DLMS's (Digital Library Management System) are the one that will be studied. Open source digital library management software's provide extensible features to administrators' and allows an organization to showcase their digital achieve to world audience. With full rights of software available under GPL and source code being provided with the software, Organization's can extend the functionality of the software as being required for the particular operation[4].

### Greenstone digital library software

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[6]. A configuration file determines the structure of a collection. Digital libraries are radically reforming how information is disseminated and acquired in UNESCO's partner communities and institutions in the fields of education, science and culture around the world, and particularly in developing countries. This software encourages the effective deployment of digital libraries to share information and place it in the public domain[4].

### Special Features of Greenstone digital library software

Listed below are some of special features possessed by the Greenstone:

**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 [7].

**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 [8] . 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[9] .

**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[5]. 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[10] .The Dublin Core metadata scheme is used for most electronic documents; however, provision is made for other schemes[11].

**Plug-in extends system's capabilities:** Plug Ins can be written to accommodatenew document types, the collection cancontain pictures, music, audio, video clips,etc. It also supports multilingual documents[12].

**Customization:** The Greenstone allows customization of presentation of collection that are based on EXtensible Stylesheet Language transformation (XSLT) and other agents that govern the definite functions of Digital library.

**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 in to 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

The production version, Greenstone2, is recommended for end-user librarians, while Greenstone3 is an emerging system currently intended for experimental use by computer scientists and information technologists. Greenstone3 is fully compatible with its predecessor, and can run old collections without any modifications whatsoever and make them indistinguishable from the original both visually and in terms of interaction. Librarian-level users can adopt Greenstone2 today, secure in the knowledge that though the software is developing their
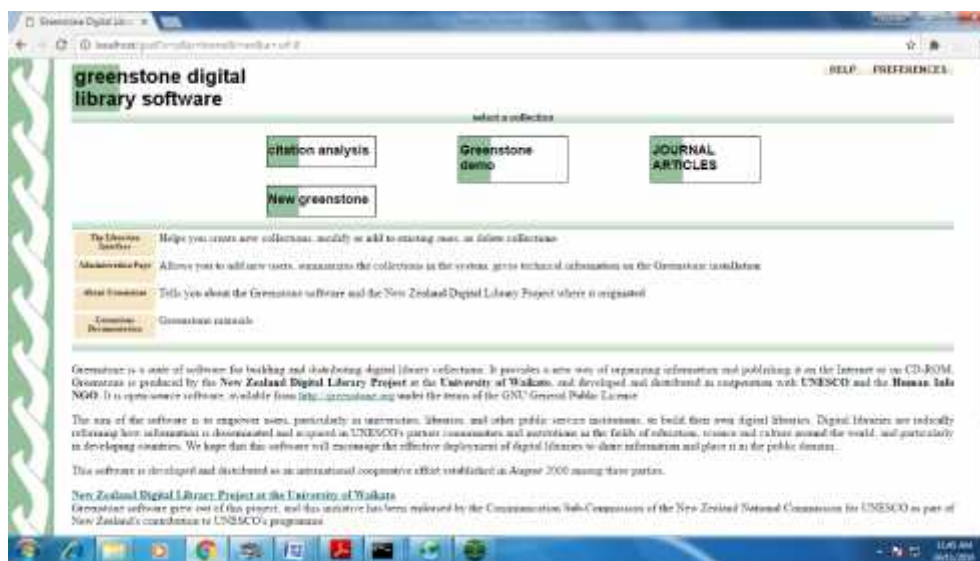


**Fig** welcome page of greenstone digital library software.

***Collections can be published on the Internet or on CD-ROM*:** The software can be used to serve collections over the World-Wide Web. Greenstone collections can be made available, in precisely the same form, on CD-ROM. The user interface is through a standard web browser (Netscape is provided on each disk), and the interaction is identical to accessing the collection on the web—except that response times are more predictable. The CD-ROMs run under all versions of the Windows operating system[4].

***Software distribution and development***

Greenstone is issued under the terms of the GNU General Public License. It originated in, and the current production version (Greenstone2) was designed about seven years ago, although it is continually being extended. A complete redesign and reimplementation, Greenstone3, has been described and released, informed by experience with the current system and the problems and challenges faced by users, international collection developers, and practicing librarians. Greenstone3 allows documents to be dynamically added to collections; provides more flexible ways to dynamically configure the run-time system by adding new services; lowers the overhead incurred by collection developers when accessing this flexibility to organize and present their content; and modularizes the internal structure. The design is based on widely accepted standards that were unavailable when Greenstone2 was designed[5].

collections will still run in exactly the same form tomorrow. The remainder of this article focuses exclusively on Greenstone2.

**Development of digital collection of 'Journal article**

***Installation of GSDL Software***

Installation of GSDL software required suitable version of JRE. User can get latest version of it through web also. Greenstone needs version 1.4.0 or higher. The current version as of initial install is 1.6_05. With the help of such requirements digital collection can be built by Greenstone Digital Library Software.

***Image magick*:** ImageMagick is a free and open-source software suite for displaying, converting, and editing raster image and vector image files. It can read and write over 200 image file formats[7]. Greenstone requires an image processor to process image collections, that is ImageMagick. Without ImageMagick processor system can build collections, but that they will not be able to contain images.

***Accessing Greenstone collection***

For viewing collections over the Internet Greenstone Welcome page act as the main interface. Top side of the page contains links to each collection available for viewing to the users. In the
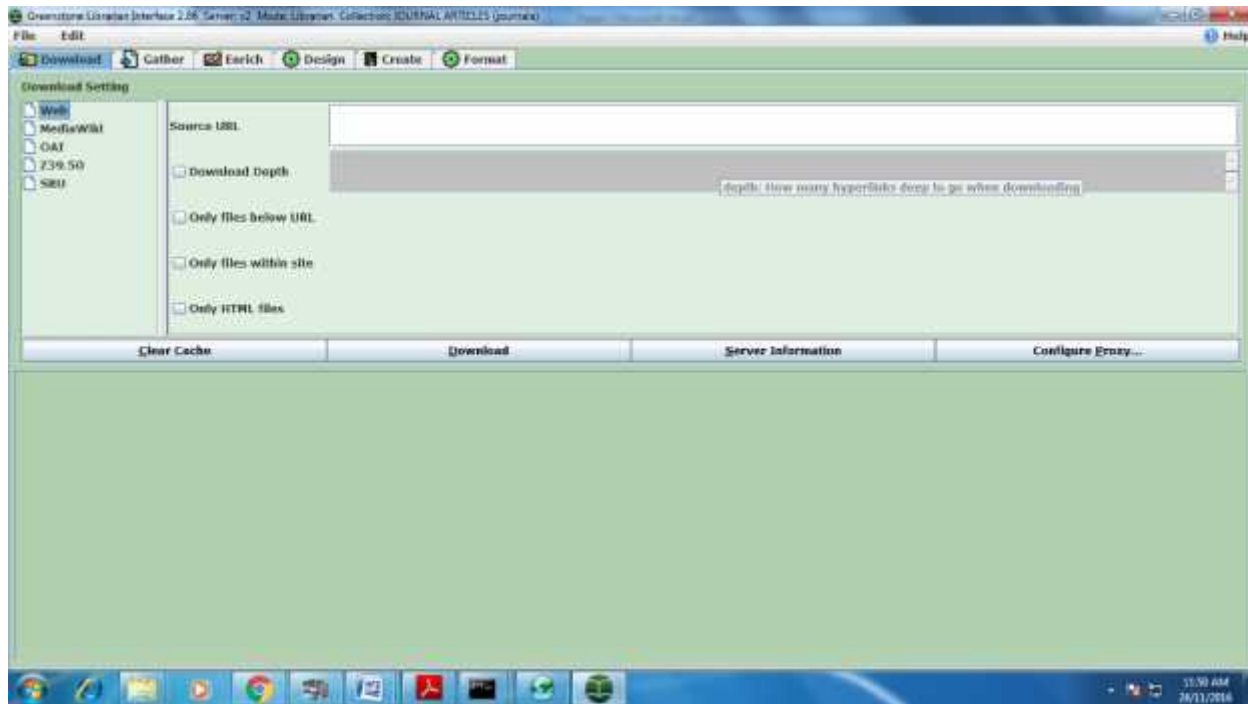
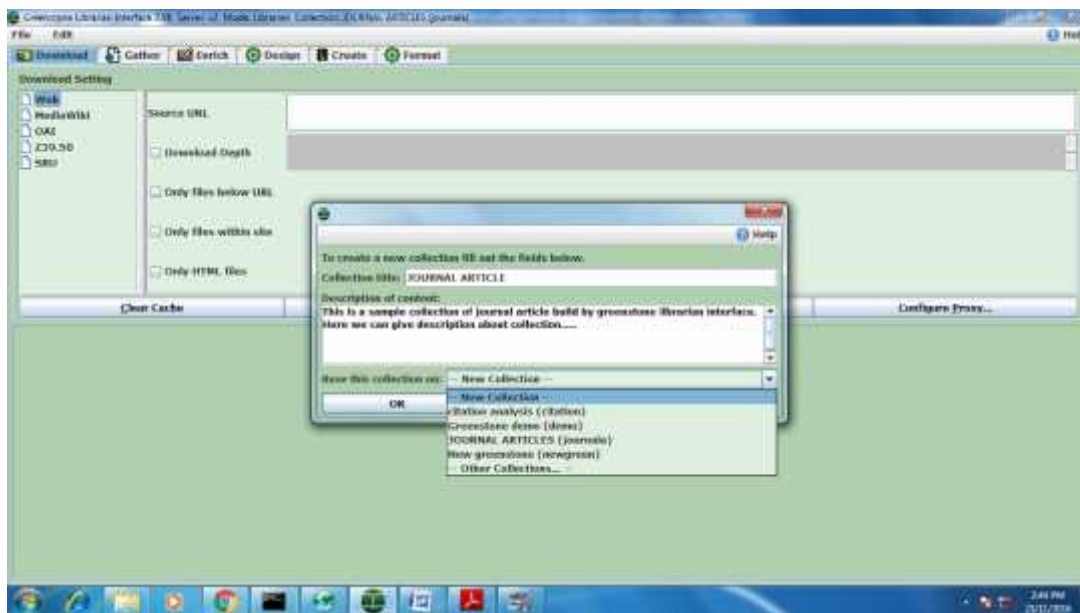**Fig** Greenstone digital library software Interface.



**Fig** Starting new collection and adding information about collection "Journal Article"

next section there has a links which providws information about Greenstone. The Administration allows administrators to manage users, view configuration files and error logs, and accesses other programming functions.

### Use Of The Greenstone Librarian Interface To Build And Edit Collections

The Greenstone Librarian Interface allows administrators to build and edit collections in Greenstone. The GLI opens in three steps in which user first command window  and it will

open for this user  will not use this window, but it must remain open while user are working in the Greenstone Librarian Interface , then "Enter Library" pop-up will open, and at  last the Greenstone Librarian Interface  itself will open. Six tabs across the top of the GLI allow user to work with different aspects of collections i.e. Download tab allows user to download resources from the Internet to add to user's collection, Gather tab allows user to add files to new or existing collections Enrich tab is where all of Greenstone's metadata is managed ,Design tab allows user to customize the search and browse
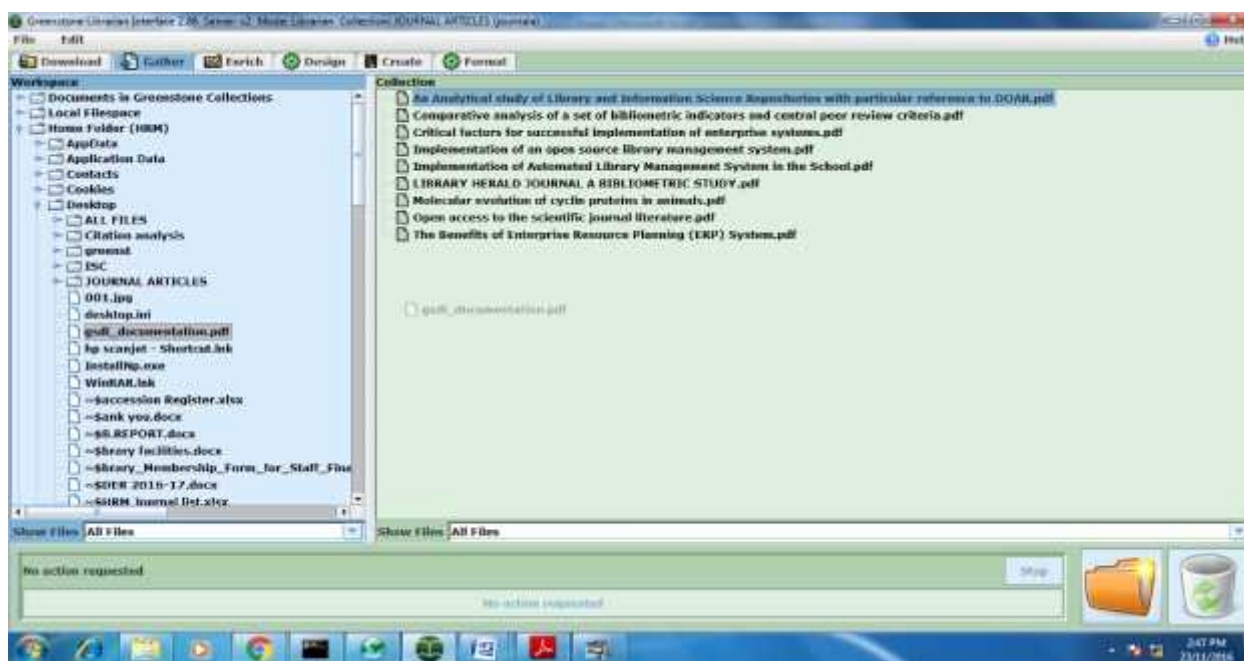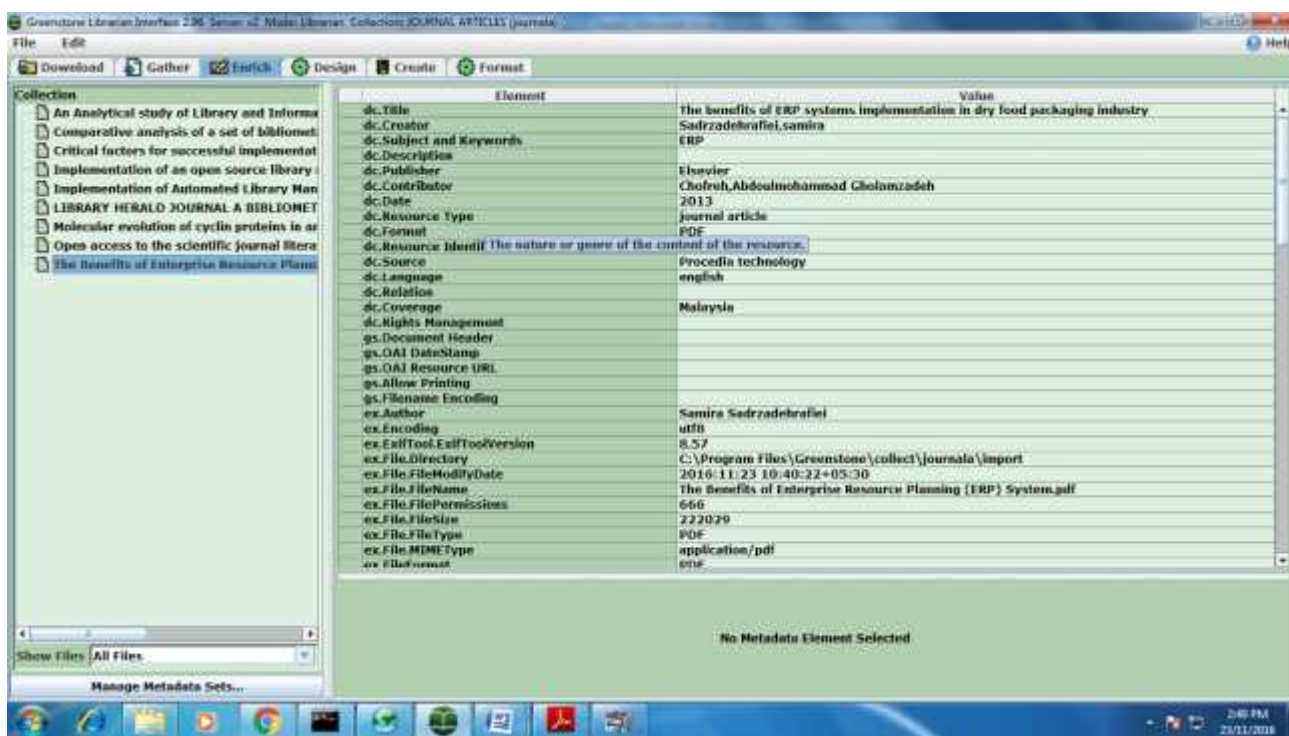
**Fig** Importing items into a collection.



**Fig** Assigning Metadata Using the Enrich View

indexes for the Greenstone User Interface ,Create tab use to build and preview collections, and format tab allows management of display macros and search and browse indexes. While building the collection named "Journal articles" functionality within each of these tabs is addressed in the following sections on building digital collection.

### Creating the collection

While creating the collection it must be built by user before any customizations can be made. This collection was built on a suite of files that made up 10 item records. For building the collection user must have to go to file and click on new collection then the box will open up for the title then enter the database name "Journal article" to the collection. Then in the Description of Content user has gives detailed information about collection .

## administrative settings

Some basic administrative settings must be defined for every collection created by GSDL and these administrative privileges is set through the Preferences. User finds preferences in Preferences tab. The window opens to the General tab in which User can set email address. With the help of mode tab user can decide which editing privileges will need to modify the collection. For the collection Journal article, the user mode should be set to Expert mode *and* that mode will appear in the title bar of the GLI just before the collection name. This ensures user about editing mode which user are using.

## Importing items into a collection

In this stage user add files to the collection with the help of Gather tab. To import items into a collection user opens Local File space, then navigate to the directory on the hard drive containing files for the library and finally select the files and drag them with the mouse into the Collection window.

## Enriching Documents with Metadata

The next stage is to enrich the documents by adding metadata. Metadata is the data about the documents, such as title, creator, subjects and so on. The metadata element electing each
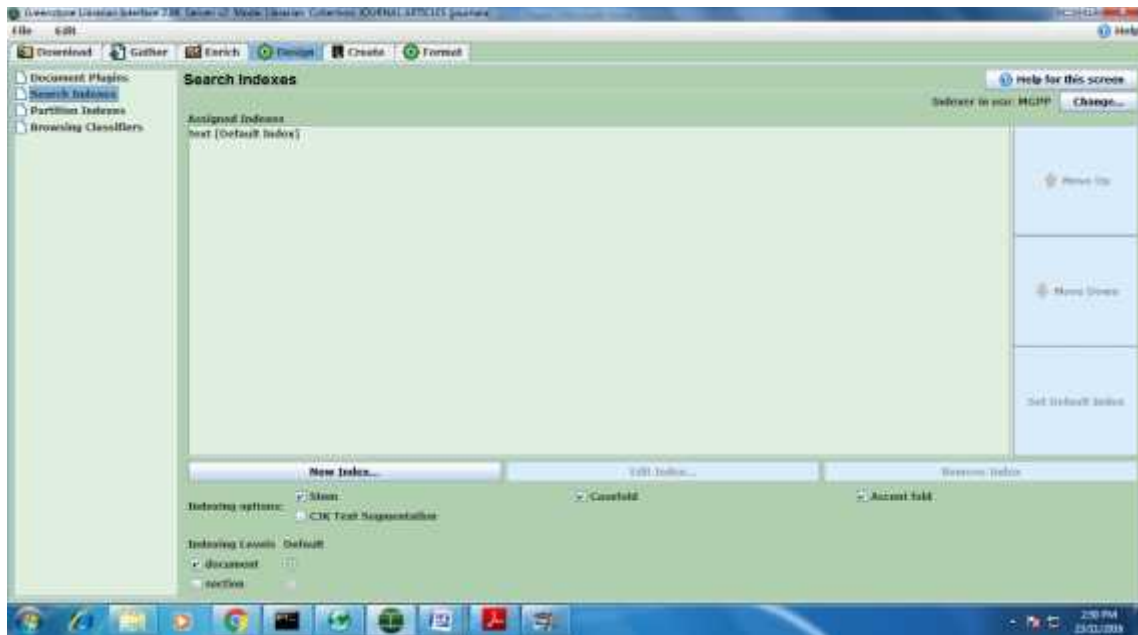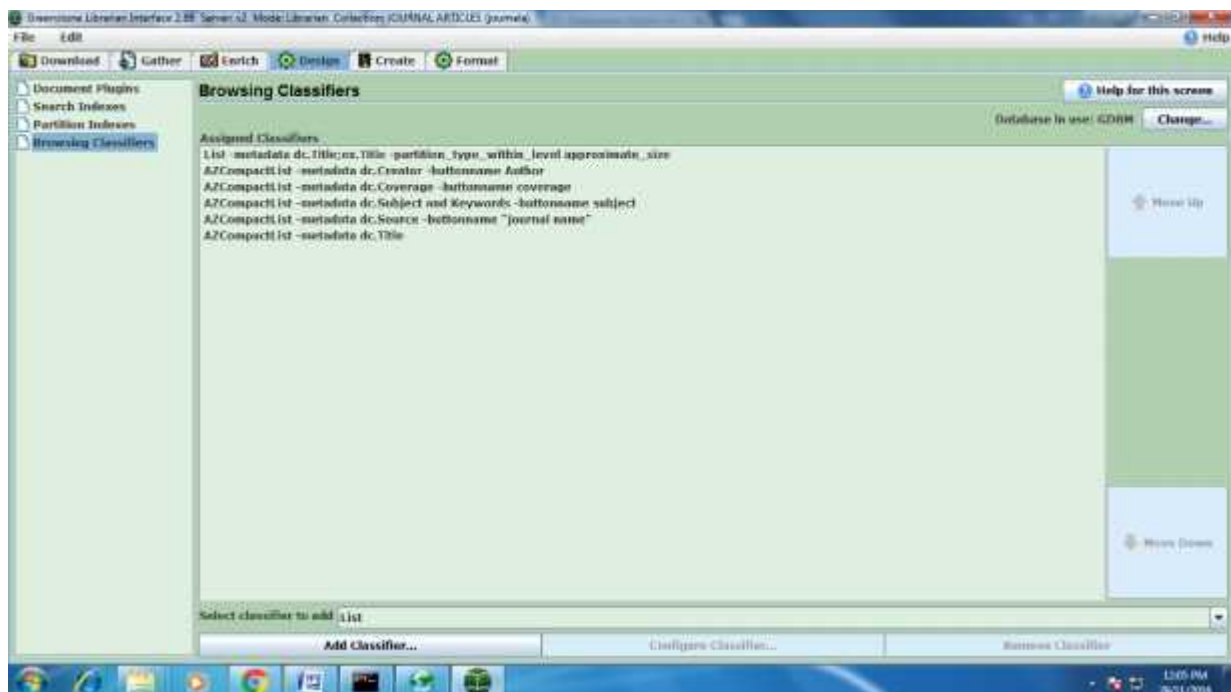
**Fig** Search indexes selected for collection

**Fig** creating browsing classifiers for collection.

document file name under collection. pre-fixed with dc. (eg. dc.Title) denotes dublin core metadata, ex. (eg. ex.Title) denotes extracted metadata, exp. (eg. exp.Title) denotes exploded metadata. Here we use Dublin Core metadata by giving value against the metadata element after You can Select the individual document and add metadata such as title,

metadata for each document on each metadata field in the Value box against the Element. Here research scholar use Dublin Core metadata and that is why dc. is prefixed with name of metadata elements such as Title, Publisher etc.
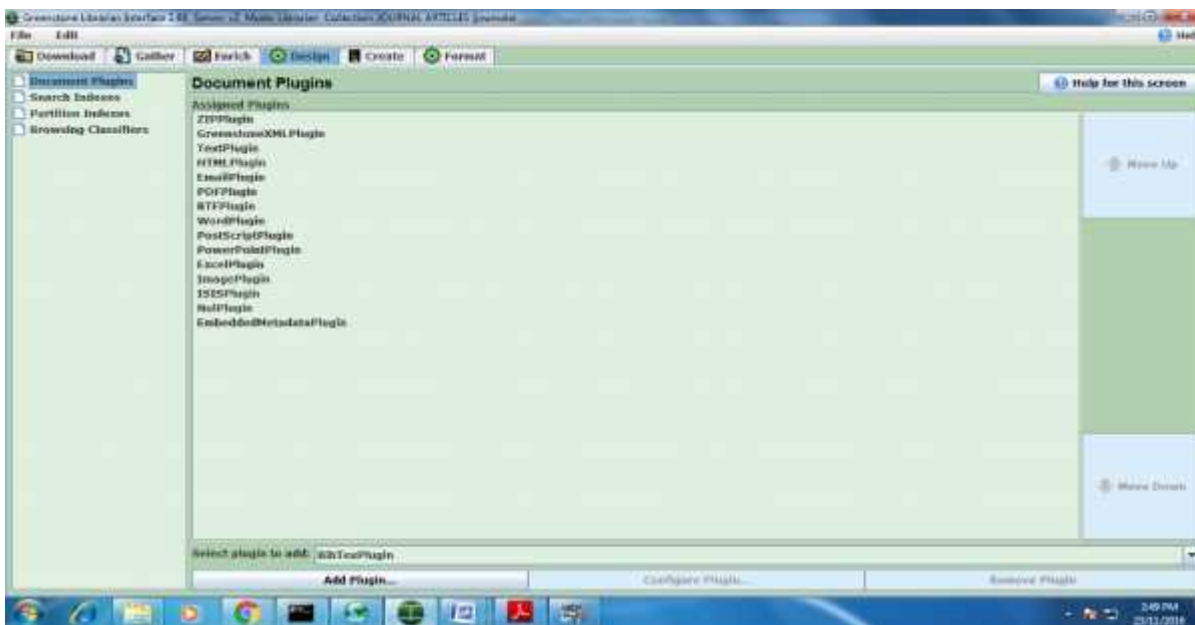
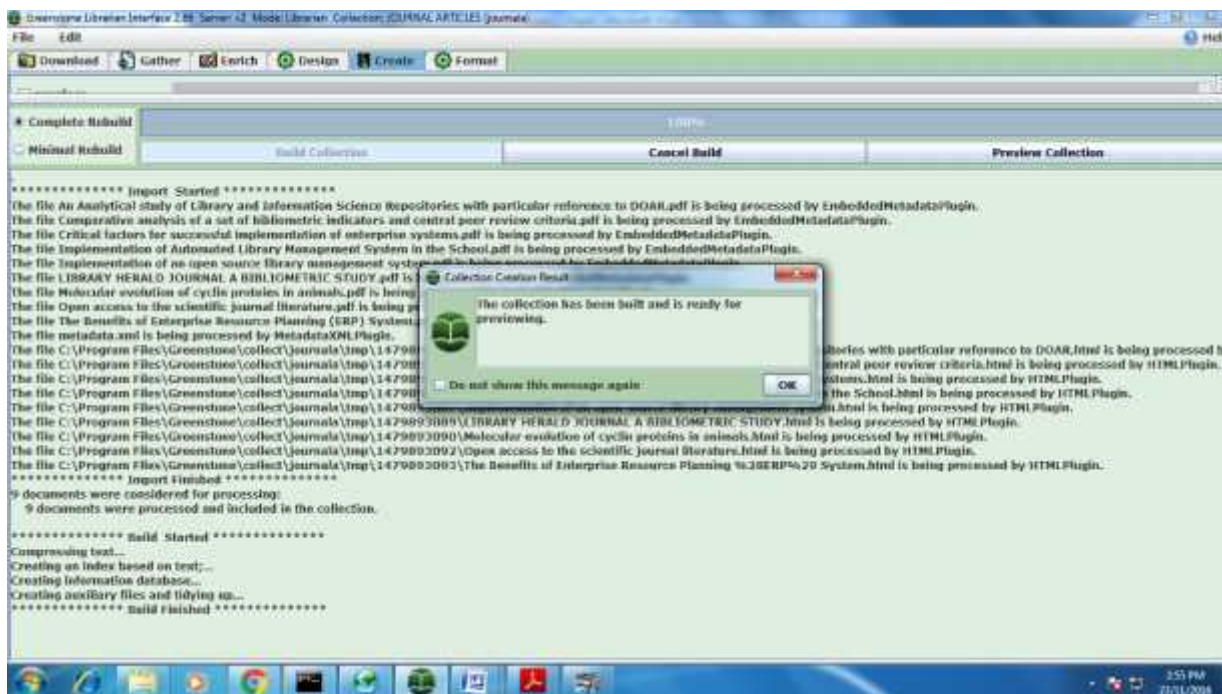

**Fig** List of document plug- ins for collection



**Fig** Building collection for "journal article"

creators or subjects manually. Click on Enrich tab and it will bring up a panel. Left side of the panel under Collection tab shows the files. The right side, on clicking, will allow adding

### Managing metadata in greenstone

In this stage user assigning metadata with the help of enrich view and Greenstone Digital Library Software has two default
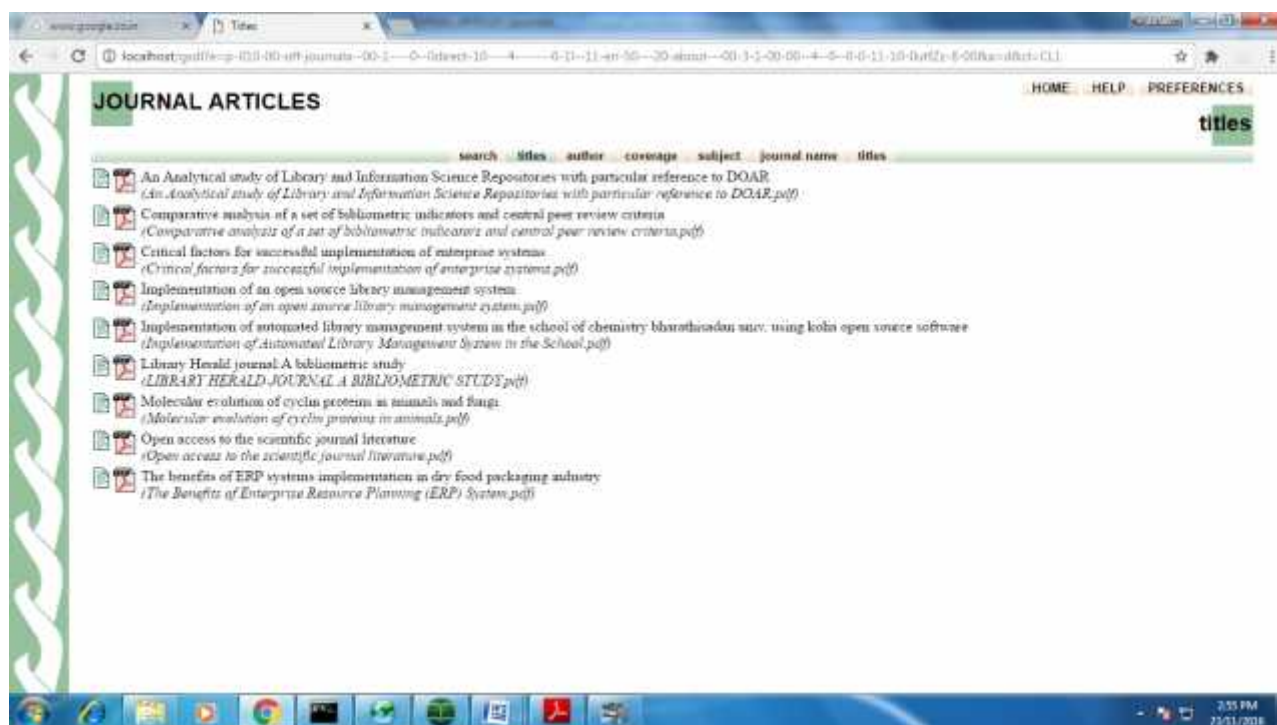
**Fig** Previewing the collection.

and Extracted Metadata Element Set (ex). Dublin Core currently allows for manually adding metadata associated with the fifteen core elements in the schema. The Extracted Metadata is created when documents are ingested and is dependent on the plug-in configurations that were completed in the previous section of this manual.

While assigning metadata user must have to type the Title of the document against dc.Title, type the Creator (Author) of the documents against dc.Creator, and type the Subjects (Keywords) against dc.Subject and Keywords for each selected document shown below Collection tab as in the above figure.

### Design the Collection

**T**hen design the collection by choosing the needed features given under the Design menu. Collection designs consist of many facets as given in the left side pane.

### Creating and formatting search and browse indexes

Greenstone facilitate user to customize its search and browse functionality to best serve the material in the collection and the users who will need to access it. User done this by associating metadata elements with different aspects of the search and browse functions.

### Search indexes

The search indexes in Greenstone Library software controls metadata elements which can search on through the interface. For example, if user want search on Title, it needs to be defined through the search indexes. With the help of design tab user operates search indexes. User can open up default Assigned Indexes for search indexes. Each Assigned Index controls a search option in the Greenstone User Interface.

### Browsing indexes

Browsing indexes allow users navigation the collection homepage and browse the collection by selected criteria. Browse indexes are built on specific metadata elements that are designated as Browsing Classifiers in Greenstone. Browsing Classifiers are also managed through the Design tab in the GLI. Browsing classifiers are found at the bottom of the Design tab menu

| Metadata Element | Button Name |
|---|---|
| dc.creator | Author |
| dc.coverage | coverage |
| dc.subject and keywords | subject |
| dc.source | Journal name |

### Document plug-ins

Through user selected plugins, Greenstone can import digital documents in formats including text, html, jpg, tiff, MP3, video, and Word, among others. The text, PDF, HTML and similar documents are converted into Greenstone Archive Format (GAF) which is an XML equivalent format. Plug-ins, written in perl language, will translate the source document into a common form, parse them and extract metadata from them. For example, the html plug-in converts the Web pages to greenstone archive format and extract metadata that is explicit in the original document from the html source tags[5]. Click on Document Plug-ins to add the required Plug-ins4 needed to convert the document into the document format (greenstone archive format) required for greenstone. All plug-ins, needed for handling common documents, will be loaded by default at the time of installation. Kindly note that if proper Plug-in is not loaded, the software cannot build the digital library collection.

### Viewing and editing of collection

With the help of Enrich tab user see the metadata associated with a collection. If user have not added any files and built a preliminary collection, it will see an empty collection window with no values selected on the right hand side of the panel. User completed further steps with the help of create tab and build the collection, but user will not be able to view metadata for files as. User were enter manually Dublin Core metadata in this section of the GLI .Because all of the metadata necessary for the collection has been extracted, user removed Dublin Core from the collection by Manage Metadata Sets and selects Dublin Core from the Assigned metadata sets then closed the window. While selecting file from the collection, user were only see extracted metadata associated with that file.

### Building the collection

During the building process, the text of the documents will be compressed and the indexes specified will be generated in order to ensure the availability of the collection for search and retrieval. GNU Data Base Manager, a program used within the Greenstone software, will store the metadata of each document. Managing Gigabytes is the program used by the Greenstone system for fulfilling indexing that incorporates compression techniques[1]. The Create panel is used to construct a collection based on the documents and assigned metadata. The brunt of this work is borne by the Greenstone code itself. The user controls this external process through a series of separate interaction screens, each dealing with the arguments provided to a certain stage of the creation process. The user observes the building process though a window that shows not only the text output generated by Greenstone's importing and index building scripts, but also progress bars that indicate the overall degree of completion of each script. When user satisfied with the arguments, then clicks on Build Collection. Greenstone continually prints text that indicates progress, and this is shown along with a more informative progress bar.

### Previewing the collection

The Preview Collection button is used to view the collection that has been built. Clicking this button launches a web browser showing the home page of the collection. In practice, previewing often shows up deficiencies in the collection design, or in the individual metadata values, and the user frequently returns to earlier stages to correct these. This button becomes active once the collection has been created. The newly created

collection will also have been installed on your Greenstone home page as one of the regular collections. This opens up a hierarchically structured file of help text, and account is taken of the user's current context to highlight the section that is appropriate to the present stage of the interaction. Furthermore, as noted above, whenever the mouse is held still over any interactive object a small window pops up to give a textual "tool tip," as illustrated near the bottom of Figure .

## CONCLUSIONS

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. This software provides a way of organizing information and publishing it on the Web in the form of a fully-searchable, metadata-driven digital resource. It can also be run in a non-networked environment (standalone), operating from removable media such as a USB Flash Drive and DVD. A general-purpose digital library like greenstone is a useful tool to provide information services in our libraries. Absence of knowledge on how to use it should not come in the way of exploiting the advantages it offers. This documentation may be used as a tool to bring in more people to the growing constituency of greenstone users.

## References

1. Rajasekharan,K., Nafala,K.M. (2007) . Building up a Digital Library with Greenstone:A Self-Instructional Guide for Beginner's, Kerala Institute of Local Administration.

2. *Naik,Umesha., Shivalingaiah,D.(February 2006). Digital Library Open Source Software :A Comparative Study. 4th International Convention CALIBER-2006,2-4*

3. Digital Libraries. 2012. Retrieved from ftp://ftp.cordis.europa.eu/pub/ist/docs/rn/castelli.pdf

4. Tramboo, Shahkar., Humma., Shafi,S.M., Gul, Sumeer. (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.

5. Greenstone Developers Guide. (n.d.). Retrieved from ttp://www.greenstone.org/developers-guide .

6. K.T. Anuradha., R. Sivakaminathan. (2011). 2009 International Symposium on Computing, Communication, and Control (ISCCC 2009), Singapore,pp.330-333.

7. Greenstone Overview. 2012. Retrieved from http://www. greenstone.org /manuals/gsdl2/en/ html/ Chapter_overview_of_greenstone.htm

8. Witten, I. H., Bainbridge, D., & Boddie, S. J. (2001). Greenstone: Open-source digital library software with end-user collection building. *Online information review*, *25*(5), 288-298.

9. Witten, I. H., Boddie, S. J., Bainbridge, D., & McNab, R. J. (2000, June). Greenstone: a comprehensive open-source digital library software system. In *Proceedings of the fifth ACM conference on Digital libraries* (pp. 113-121). ACM

10. Don, K. J., Bainbridge, D., & Witten, I. H. (2002). The design of Greenstone 3: An agent based dynamic digital library. *Initial design document of Greenstone3, Dec*.

11. Press, N. I. S. O. (2004). Understanding metadata. *National Information Standards*, 20. Retrieved from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

12. Sonkar,Sharad Kumar., Makhija, Veena., Kumar, Ashok., Dr.Singh, Mohinder.(2005). Application of Greenstone Digital Library (GSDL)Software in Newspapers Clippings. *DESIDOC Bulletin of Information Technology ,*25(3),9-16.