

METADATA ENGINEERING IN GREENSTONE DIGITAL LIBRARY SOFTWARE

Muzamil Mushtaq

The present study deals with the incorporation and defining new metadata set and building a model digital library collection using Greenstone Digital Library Software (GSDL), which is an outstanding open source software for building a digital library. A case study was undertaken to modify and devise a set of metadata based on the approaches of research scholars and the B.Lib.I.Sc and M.Lib.I.Sc students of the Department of Library and Information Science, AMU in searching the documents available in the seminar library of the department as well as in the main central library of the university. The study tried to create some of the most usable approaches/search options of the users and enrich the resources gathered in the model digital library according to these elements. It defined the seminar library metadata with twelve bibliographical elements as the fields for searching and browsing the digital collection. The study serves as a model of devising and developing a schema of metadata based on the needs and requirements of a particular library and may help in building digital collection or group of collections using Greenstone Digital Library Software.

Keywords: Greenstone Digital Library Software; Metadata Element Set; Open Source Software; Aligarh Muslim University; Seminar Library; Enrichment of E-resources; User Approaches; Digital Library; Design and development of digital collection.

INTRODUCTION

Open access is ideally the openness of mind and the word has now been known more than a decade. This framework is basically a principle of making the scientific and scholarly knowledge freely accessible and unimpeded of any legal constraints. It is based on the premise that 'knowledge is for all' and authors believe that their ideas should be read by maximum and hence the scientific research is contemplated to be communicated freely and fairly without any restriction and control. Sometimes its philosophy is described in the sense that the research is funded by tax payers and should be available free of charge to them and that the research is a public good and hence should be available to all irrespective of their paying capacity [1]. The open access movement is augmented by many theories, scientists, research institutions, professional bodies and libraries, but the significant 1940 Statement of Principles on Academic Freedom and Tenure which was agreed upon by the American Association of University Professors (AAUP) and the Association of American Colleges and Universities, that the "*common good depends upon the free search for truth and its free exposition,*" and "*[a]cademic freedom is essential to these purposes and applies to both teaching and research.*" is worthwhile. The statement provides university faculty, "full freedom in research and in the publication of the results" without subjection to institutional censorship [2]. One of the important aspects of Open access is the open source software which revolves around sharing and collaboration of software programmes to freely download and redistribute for further use. Generally speaking, open source software is said to be a computer software whose source

Muzamil Mushtaq
Department of Library
and Information Science
Aligarh Muslim
University, Aligarh
naikoomuzamil@gmail.
com

code is available under a license that permits users to study, change, and improve the software, and to redistribute it in modified or unmodified form [3]. The basic idea behind open source software is simple in the sense that the software evolves when programmers can read, redistribute, and modify the source code for a piece of software. People improve it, adapt it, fix bugs in the software” [4]. Open source software is very often developed in a public and collaborative manner. It is the most prominent example of open source development and often compared to user-generated content or open-content movements [5]. Some of the applications of Open source software is in the form of open source digital library software which presents a system for the construction and presentation of information resources and their collection. It helps in building collections with searching, retrieval and metadata-based browsing facilities. Moreover, they are easily maintained and can be augmented and rebuilt automatically.

The present study has used Greenstone Digital Library software which was developed and distributed as an international cooperative effort by New Zealand Digital Library Project at the University of Waikato, United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Human Info NGO, based in Antwerp, Belgium in August, 2000 (idea formulated in 1997). The software provides different ways of organizing information with collections comprising of large numbers of documents ranging from books to articles, newspaper articles to technical documents, educational journals to oral history, visual art to videos and music collections to ethnic folksongs, etc. and more importantly the interface of searching them. It is basically a suite of software for building and distributing digital library collections. It helps to design fully-searchable, metadata-driven digital libraries [6]. The software possesses a number of significant features for building a digital library with multiple collections of electronic resources. The main features of the software are like full text and field search, flexible browsing facilities, create access structures automatically, make use of available metadata, plug-in extends system’s capabilities, customization, designed for multi-gigabyte collection, multilingual support, collections support multiple formats, administrative functions provided, etc. [7]. The main processes provided by

the software for building a digital collection is by using GLI interface, which consists of options like gather documents for building up the collection, enrich the documents by adding metadata, design the collection, its appearance and the access facilities, format the appearance of the digital library, create the collection and convert the digital library into a CD-ROM library.

The present study was undertaken to define a new metadata based on the approaches of the users in searching the online or digital content at the seminar library of the department of library and information science. It is a well-equipped library with a document collection consisting of about 5341 books and 403 bound volumes of periodicals. The library subscribes about 20 current journals of national and international repute. The seminar library of the department has been automated in 2003 with Alice for Windows software, and was designated as the first fully automated seminar library of Aligarh Muslim University. All the books and member cards are bar-coded with online issue and return facilities. The library also provides access to online public access catalogue (OPAC), online and CD-ROM databases, journals, reference books, etc. [8].

REVIEW OF RELATED LITERATURE

There are many aspects of designing and developing digital libraries using Greenstone Digital Library software with multiple options of customization regarding the building, formatting, browsing and searching the digital collection and hence many studies have been undertaken in these areas of research. The GSDL is the basically an outcome of the open access movement and number of open source software emerged and often described as computer software whose source code is available under a license that permits users to study, change, and improve the software, and to redistribute it in modified or unmodified form. Such programs can be read, redistribute, and their source code can be modified. People improve it, adapt it, fix bugs in the software [3] & [4]. So far as the Greenstone Digital Library software is concerned, it incorporates an interface that makes it easy for people to create their own library collection. The collections may be built and served locally from the user’s own web server, or remotely on a shared digital library host [9]. The main features of the

software consist of full text and field search, flexible browsing facilities, create access structures automatically, make use of available metadata, plug-in extends system's capabilities, customization, designed for multi-gigabyte collection, multilingual support, collections support multiple formats, administrative functions provided, etc [7]. It offers easy ways to personalize the display of the collection pages. By customizing the images and macros, it provides options to change the appearance and look of the home and about pages. The changes like the header image, colour of swirly side-bar and background color, appearance of browsing classifier buttons in the about page, default browser buttons, modifying the macros, changing the text appearing in the home in page, etc [10]. The feasibility of the Greenstone Digital Library software is that it helps to build digital libraries as cost-effective information retrieval systems, which can offer materials to the larger audience, and can simplify the process of information search [11]. Such digital libraries can assist human development by providing a mechanism for distributing information which can addresses specific problems along with the benefits from different methods of information distribution systems [12]. The Greenstone Digital Library software has also been used to establish Digital Library System for the University Library in order to solve the limitation of information access of electronic versions of books, research articles and papers, lecture notes, etc. The established digital library was very useful in organizing and accessing the resources [13].

OBJECTIVES OF THE STUDY

The study is aimed at analyzing the user approaches in identifying and searching the online resources and build a model digital library and customize the metadata elements on the basis of user searches by using Greenstone digital library software. The study is formulated with following objectives:

- To determine the searching options used by the research scholars and the students of the seminar library of the department of library and information science, AMU, Aligarh.
- To build a model digital library of e-resources on different areas of library science using greenstone digital library software.

- To engineer the metadata element set and customizing the digital library based on the approaches of the users in the seminar library.

RESEARCH METHODOLOGY

The methodology was divided into three stages on the basis of three objectives of the present study which can be discussed as follows:

- **Opinion from the seminar library users:** A structured questionnaire consisting of queries related to the search techniques used by the research scholars and the B.Lib.I.Sc and M.Lib.I.Sc students of the department and the approaches used by the majority of the users were selected for designing the bibliographical fields in the new metadata set.
- **Building a model digital library using Greenstone digital library software:** A list of articles, e-books, conference papers and other allied documents were downloaded and exported into GSDL using 'gather' option. These documents were enriched with the new metadata, followed by adding browsing and searching options to retrieve the relevant contents of the digital collection. The collection of the documents was formatted using various format options provided by the software and finally the model digital library was built.
- **Adding new metadata schema and customizing its bibliographical elements:** The main aspect of the present study was enriching the digital library with new metadata consisting of all those fields/elements proposed or widely used by the users of the seminar library. For this, first the existing metadata were removed from the list of default metadata sets and a new schema of metadata were selected and new bibliographical elements were added. Finally, the collection was enriched with the new metadata to build a model digital library with customized approaches of searching and browsing the content.

COLLECTION BUILDING

In Greenstone, the process of building any digital library starts with the creation of a collection with options of name and short description. Here the name 'DL Seminar' was used for the model

collection along with the description that the collection belongs to the seminar library of the department of library and information science. The downloaded files were exported into the collection using drag and drop option. Before using the second step of enriching the seminar collection, all the existing/predefined metadata sets were removed so

that a new metadata set can be added with new elements. These predefined metadata sets are Dublin Core (qualified and unqualified), RFC 1807, NZGLS (New Zealand Government Locator Service), AGLS (Australian Government Locator Service), etc. The outline of the steps can be visualized from the screenshots below (fig 1, 2 & 3):

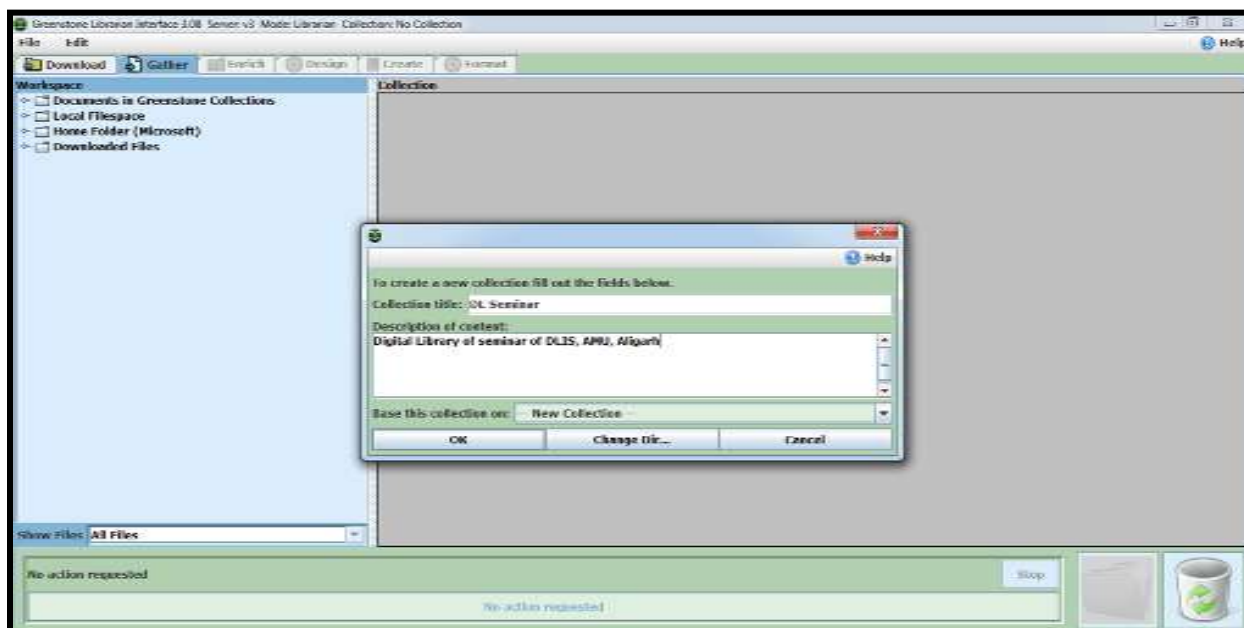


Fig 1: Collection Name = DL Seminar and Collection Description

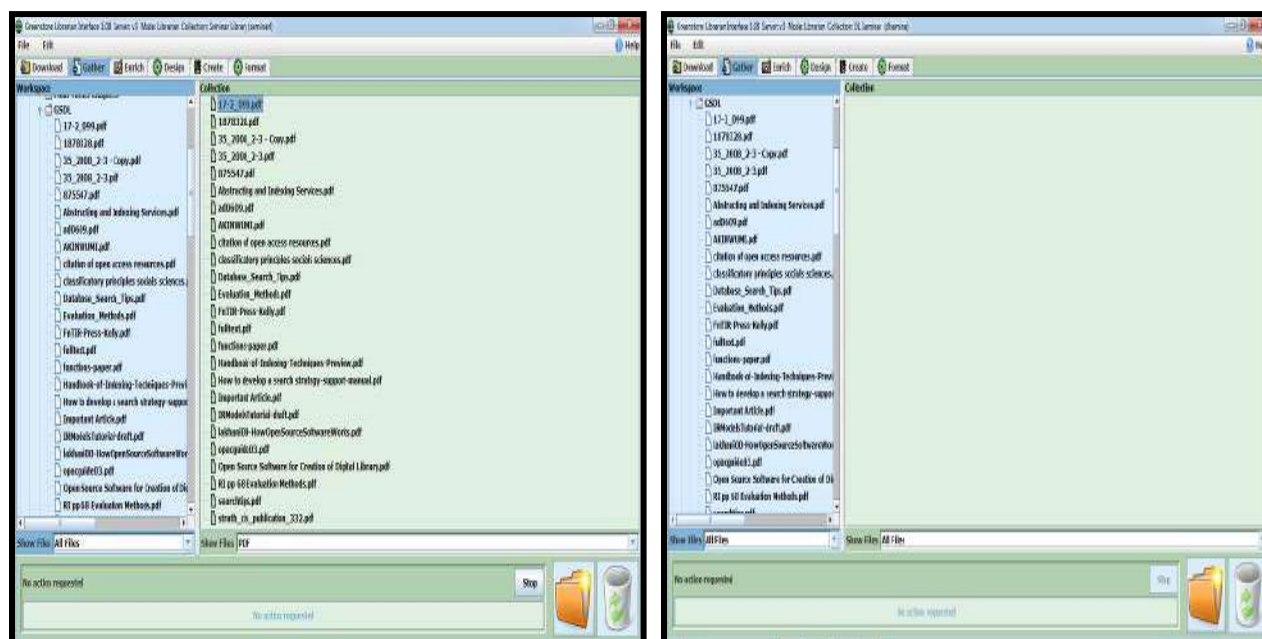


Fig 2: Exporting and Adding files by Drag and Drop

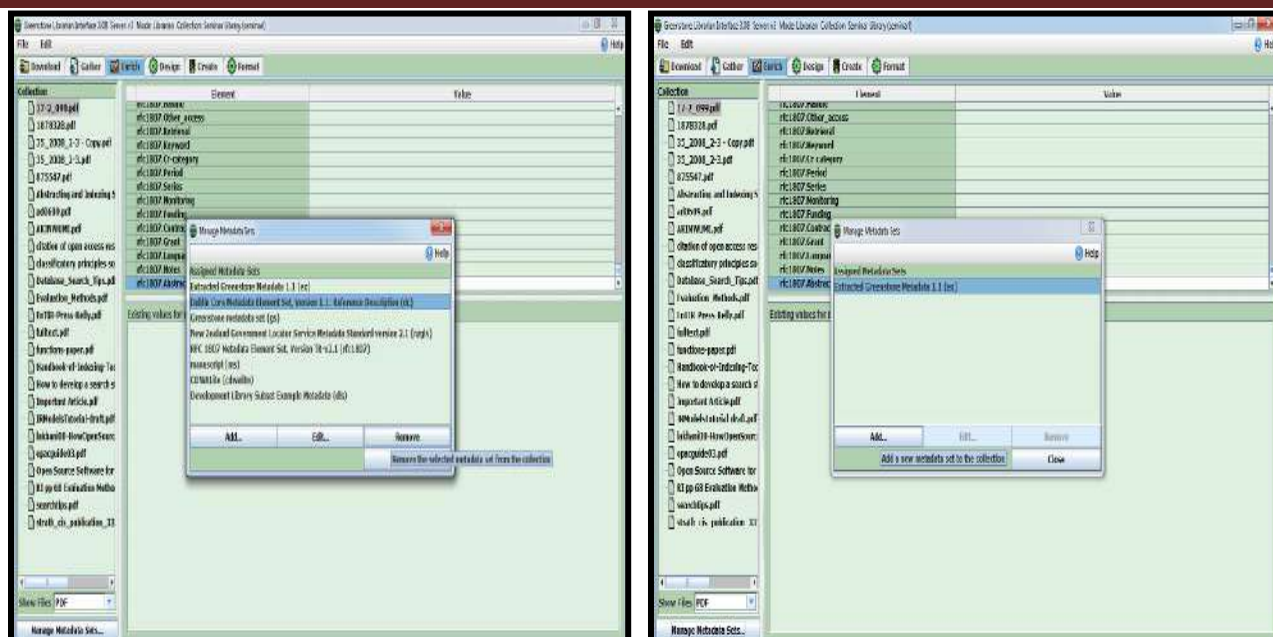


Fig 3: Removing of Existing Metadata Sets

Defining New Metadata

New metadata sets can be defined using Greenstone's Metadata Set Editor along with the 'Plug-ins' that are used to ingest externally-prepared metadata in different forms, and plug-ins exist for CSV, XML, EXIF, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace and METS files to process in the GSDL collection building [14]. While

enriching the collection with new metadata, the 'manage metadata sets' and add options help opening a window consisting of new metadata set information like metadata set title, metadata set namespace, metadata set description as shown in the fig 4. The metadata was named as seminar library metadata with sem.mds as file name.

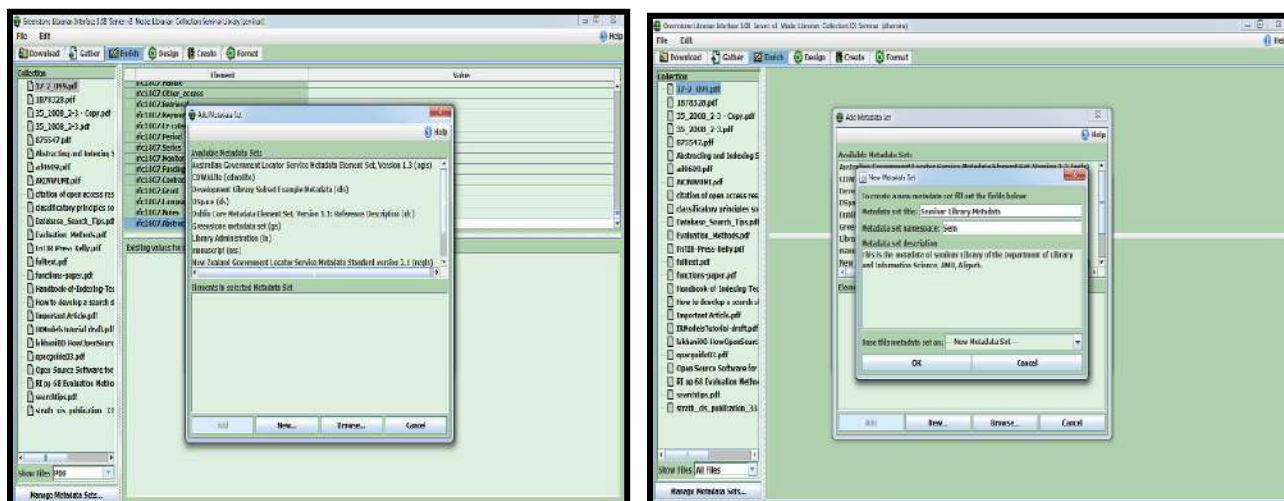


Fig 4: New Metadata (i)

Title, namespace and metadata description (ii)

Defining Metadata Elements

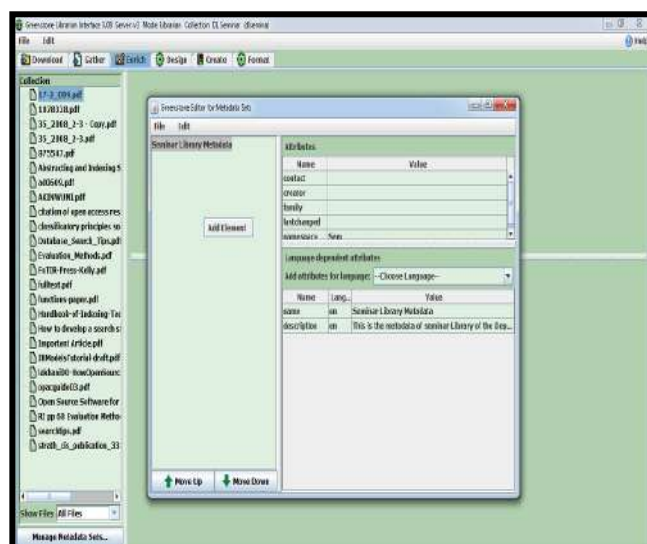
Once the new metadata schema is defined, the second step is to define the elements of metadata set. As initiated, the current study has used the approaches as proposed by the users as the most suitable bibliographical elements of the new

metadata. The elements added in the new metadata are as follows:

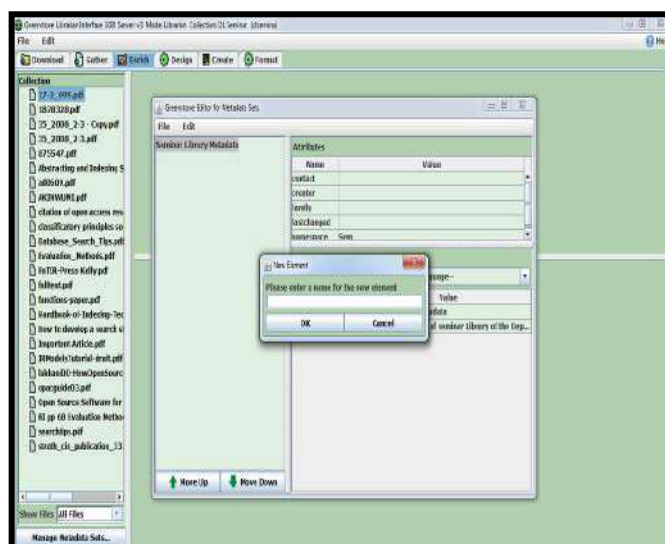
- Author
- Subject
- Title

- Name of the publisher
- Place of the publication
- Year of the publication
- Accession number
- ISBN/ISSN
- Call No.
- Document type
- Keyword
- URL/DOI

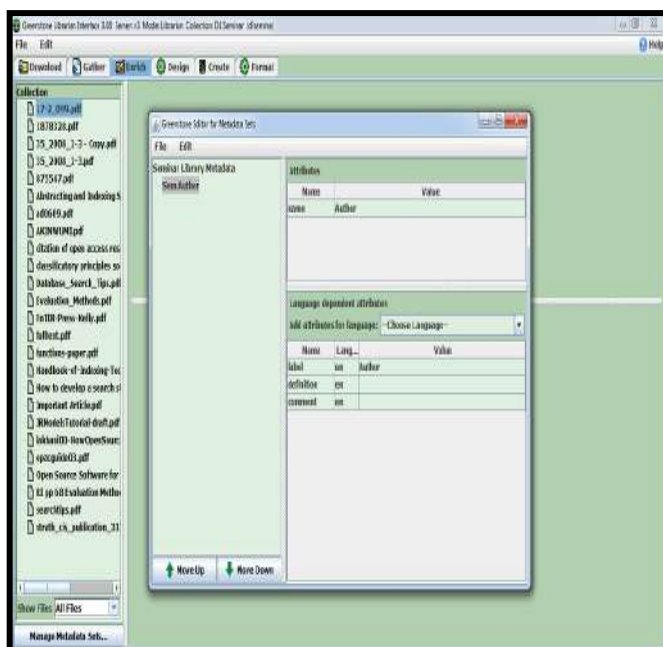
There are three parts of adding the metadata elements: label, definition and comment. While as the label indicates the name of the element, the definition gives some explanation of the metadata element and the comments refer to some additional remarks regarding the element. It is worthwhile to mention here that the definitions given to the bibliographical elements can be visualized at the time of enriching the collection. The various steps of defining the metadata can be visualized from fig 5 (i-iv), 6 and 7 below:



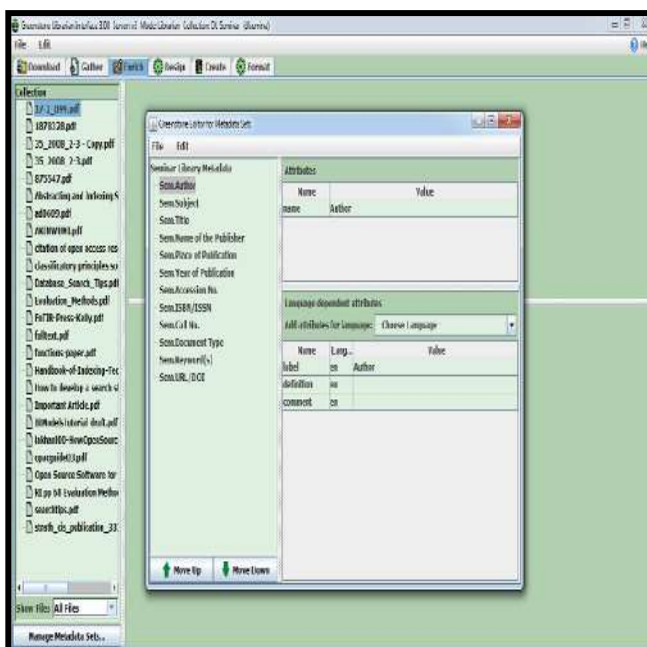
(i)



(ii)



(iii)



(iv)

Fig 5(i-iv): Page for Adding Metadata Elements

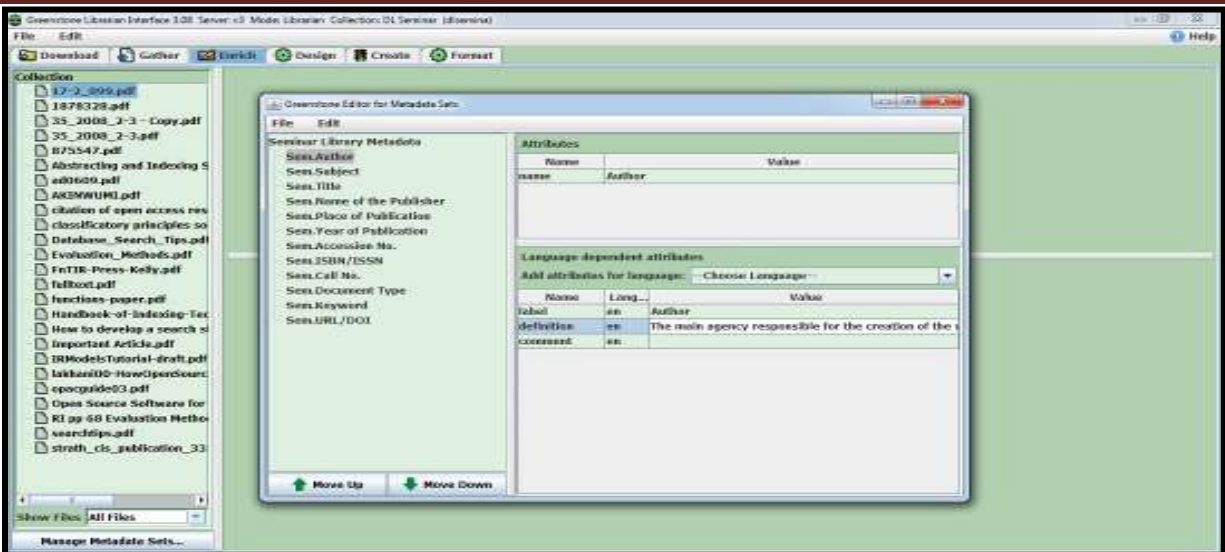


Fig 6: Adding Definition of the Elements

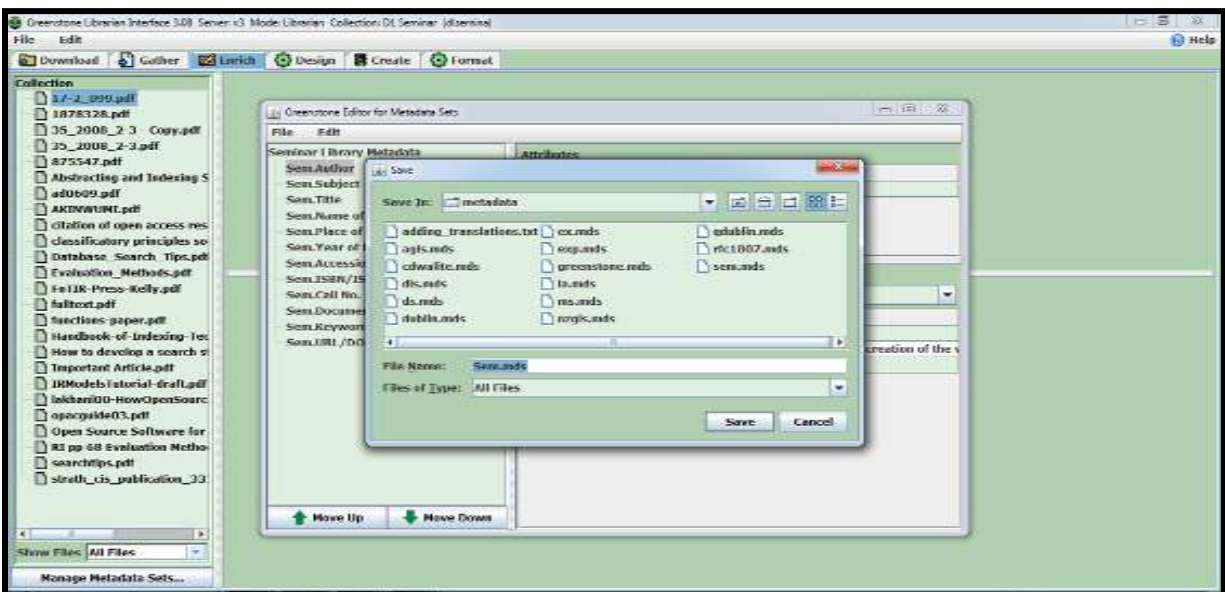


Fig 7: Saving the Metadata

Enriching the Collection

Once the files are gathered into the collection, the documents in the collection are assigned metadata using the 'Enrich' option of the window. This is the most time-consuming process because it enhances searching and retrieval of individual documents in the digital library. On the left side of the enrich tab, there is the document tree representing the collection, and on the right is the metadata which can be added to individual documents, or group of documents. According to Witten, Stefan Boddie and John Thompson [15], the operations at enriching the documents in the collection consists of:

- Assigning new and existing metadata values to documents.
- Assigning metadata to an individual document.
- Assigning metadata to a folder
- Assigning hierarchical metadata
- Editing or updating assigned metadata
- Reviewing the metadata assigned to a selection of files and directories.

In the present study, the enrichment of the documents was done by adding all the bibliographical details of the documents collected for the model digital library as represented in the fig 8.

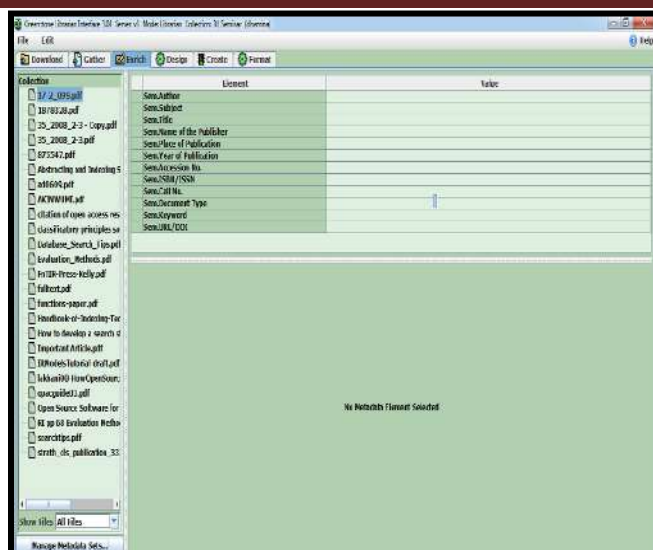
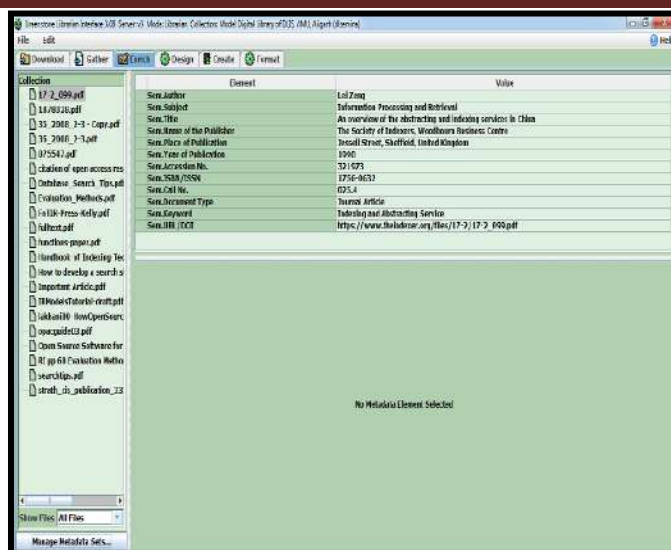


Fig 8: Enriching the collection

Designing the Collection

After the documents are enriched with the bibliographical details, the process of designing of the collection is done. In design view, document plugins, search indexes, partition indexes and browsing classifiers are provided for designing and developing the digital library collection. A long list of plugins help to process different kinds of files gathered in the digital library and there was no need to change or add any new Plugin for the model digital library collection (figure-9i). In case of search indexes, all the search options related to all the bibliographical elements used in metadata were selected and added for searching the contents of the



(new bibliographical elements will appear)

digital library with 'sem.keyword' as default index. Using all the elements as search options can help users to search more easily and effectively (figure-9ii) the contents of the collection by multiple approaches. The present study does not use any partition index as there was only one collection related to seminar library and only one type of documents were used to build the model digital library.

In case of browsing classifiers, a similar approach was used by including all the bibliographical elements as browsing options for the users with AZ Compact list as classifier (fig 10).

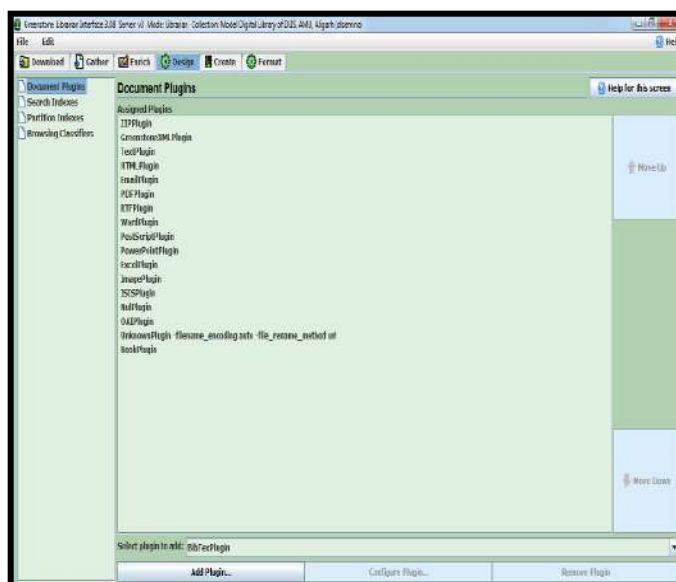
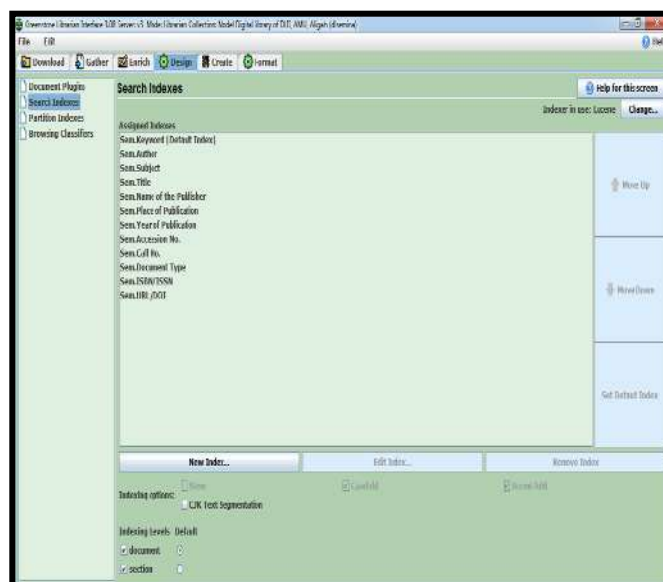


Fig 9: Document Plugins (i)



Search Indexes (ii)



© 2006 The Authors

(c) (d) (e)



0



1

C O N T E N T S

1 91 1

C



Fig 12: Model Digital Library

The fig 13 represents the browsing by author, subject and title respectively.

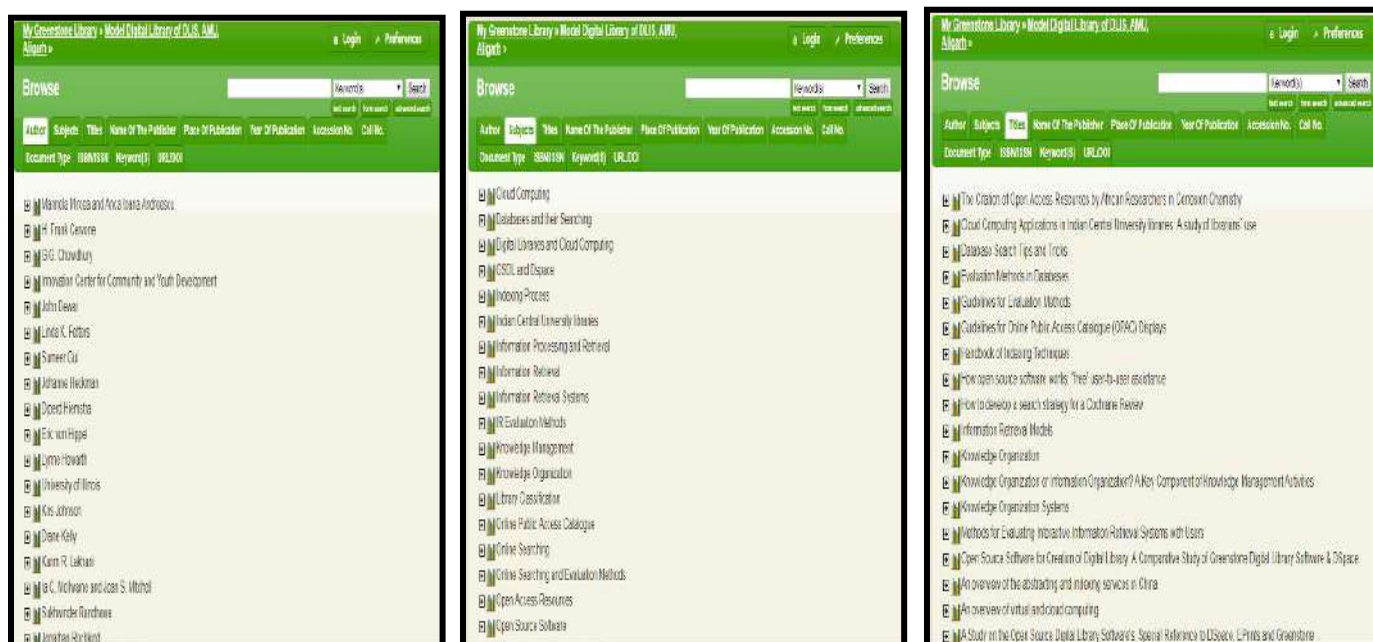


Fig 13: Browsing by Author, Subject, Title

Searching Options: Basic and Advanced

One of the main features of any information retrieval is the search options and the exhaustiveness of searching and identifying the documents available in the collection. The searching interface of the digital library provides text, form and advanced search techniques. In advanced search, Boolean operators, Granularity, ranking of results, hits per

page and search history is provided to fulfill the user approaches of searching a large chunk of information from the digital library. The preference options provide the user centric searching options and can be used in continuum based on the needs and requirements of the users.

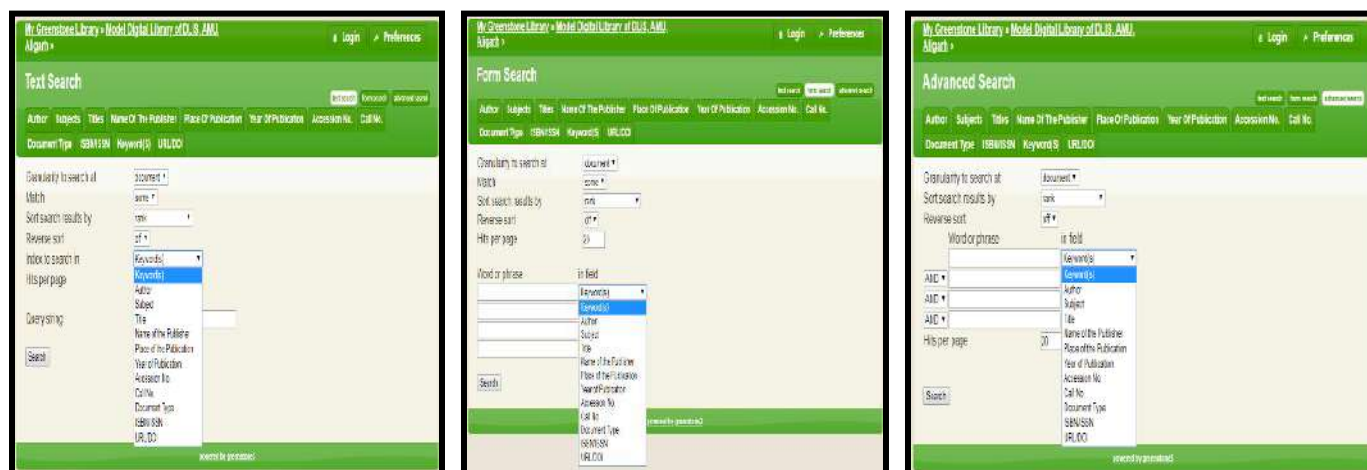


Fig 14: Text Search (i)

Form Search (ii)

Advanced Search (iii)

CONCLUSION

Design and development of digital content in the contemporary information and knowledge society is imperative and so emerges the need for a suitable platform to put the content in a form more realistic and suitable to the ultimate user. Open source software are boon in this direction which not only provide a wide array of designing, enriching, formatting, searching and more importantly those features which are based on user needs and approaches but also customization of modifying and remodeling the software to fit with the changing environment. These helps providing accessibility and retrieval platforms for identifying and locating relevant documents from a large volume of information. The greenstone digital library software is amongst the doable interfaces which help to create and build a quintessential collection of digital documents successfully and easily. It is sophisticated software to download, gather, enrich, format, design and build an outstanding digital library. Notably, the software is more suitable for any kind of digital collection with multi-lingual options available. The present study was undertaken to build a model digital library for seminar library in the Department of Library and Information Science based on the usability of OPAC and its viable search options coupled with potential approaches of users. Keeping the aspects of change in searching and retrieval of documents, a new metadata was added with all possible search approaches as bibliographical elements in the metadata. At least, twelve elements were introduced and adopted in the schema of metadata and a list of digital documents were added and enriched with the metadata. The study used the

different features of the software to manage metadata, designing, formatting, searching and browsing along with the appearance of the home and about page of the digital library. A well-developed model digital library was built with different basic and advanced search options, browsing, preferences, and the customized search techniques. The library provides the easy access to its collection with suitable text, form and expert searching. The present model can serve as the way out to design and develop a suitable collection or group of collections of digital documents using greenstone digital library software. It can help in understanding the various facets of modifying and adding a new metadata and designing a digital collection with those elements added in the metadata. The model can serve as the antecedent to design any digital collection with an ample number of features of efficient search and retrieval facilities using greenstone digital library software.

REFERENCES

- [1] United Nations Educational, Scientific and Cultural Organization UNESCO. *Introduction to Open Access, open access for library schools, Module 1*, 2015. Retrieved 12, September, 2018 from: <http://unesdoc.unesco.org/images/0023/002319/231920E.pdf>.
- [2] American Association of University Professors. 1940 Statement of Principles on Academic Freedom and Tenure, with 1970 Interpretive Comments, 2018. Retrieved 12, October, 2018 from <https://www.aaup.org/file/1940%20Statement.pdf>.

-
- [3] Herteliu, Eduard Emanuel. Open Source and Free Software Concepts Implemented in Publication of Academic Research. *Journal of Applied Quantitative Methods*, 5 (4), 2010, 692-705.
Retrieved 18, October, 2018 from: http://www.jaqm.ro/issues/volume-5,issue4/pdfs/14_ivan_herteliu.pdf.
- [4] Opensource.com. *What is open source?* Retrieved 22, October, 2018 from: <https://opensource.com/resources/what-open-source>
- [5] N. Pankaja & Mukund Raj P K. Proprietary software versus Open Source Software for Education. *American Journal of Engineering Research (AJER)*, 2(7), 2013, 124-130.
Retrieved 05, October, 2018 from: [http://ajer.org/papers/v2\(7\)/O027124130.pdf](http://ajer.org/papers/v2(7)/O027124130.pdf).
- [6] Greenstone.org. *About Greenstone*, 2018.
Retrieved 20, September, 2018 from: <http://www.greenstone.org/factsheet>
- [7] Witten, I.H. & Bainbridge, D. Creating digital library collections with Greenstone. *Library Hi Tech*, 23(4), 2005, 541-560.
DOI: 10.1108/07378830510636337
- [8] Aligarh Muslim University. *About the department*, 2018.
Retrieved 05, September, 2018 from: <https://www.amu.ac.in/departmentspage.jsp?did=80>
- [9] Witten, Ian H., Bainbridge, David & Boddie, Stefan J. Greenstone: open-source digital library software with end-user collection building. *Online Information Review*, 25(5), 2001, 288-298.
Retrieved from: <https://doi.org/10.1108/14684520110410490>
- [10] Rajasekharan, K & Nafala, K M. *Easy Ways to Customize the Display of Greenstone Pages*. 1-12.
Retrieved from: <http://greenstonesupport.iimk.ac.in/Documents/Customizing%20gsdl.pdf>.
- [11] Malkawi, M. *Digital library services in special libraries*, 2007.
Retrieved from: <http://www.unesco.org/new/en/communication-and-information/>
- [12] Witten, I. Digital libraries for the developing world. *Interactions*, 2(4), 2006, 47-49.
DOI:10.1145/1142169.1142187.
- [13] Enanu Tilahun. *Practical Application of Greenstone for managing a University Library System*, 2013, 3-19.
- [14] Greenstone. *Greenstone digital library user's guide, Making Greenstone Collections*, 2017.
Retrieved 03, October, 2018 from: http://www.greenstone.org/manuals/gsd12/en/html/Chapter_making_greenstone_collections.html.
- [15] Witten, Ian H., Stefan Boddie & John Thompson. Greenstone Digital Library. *User's Guide*, 2006.
Retrieved 03, October, 2018 from: http://www.greenstone.org/manuals/gsd12/en/html/User_en_all.html
-