DESIDOC Journal of Library & Information Technology, Vol. 30, No. 6, November 2010, pp. 25-30

Design and Development of Prototype Astronomical Digital Image Library using Greenstone Digital Library Software

Md Nurul Alam and Pragya Pandey

Physical Research Laboratory, Navrangpura, Ahemdabad-380 009
E-mail: alam@prl.res.in; praggyap@prl.res.in

ABSTRACT

This paper provides a mechanism for the development of a digital library—Greenstone Astronomical Digital Image Library (GADIL)—for astronomical images using open source software Greenstone digital library, and its use for astronomy community and educational users in India. The system is based on open source software like GSDL, Apache, PERL, GRE and Imagemagik, and can serve as an ideal access tool of research-ready scientific images so as to enhance astronomers' productivity and increase public awareness of astronomy as a subject.

Keywords: GADIL, Greenstone Digital Library, astronomy, digital images library, digital library

1. INTRODUCTION

Digital libraries have collection of digital objects, including text, images, videos, and audios along with the methods for their access and retrieval. Robust and flexible digital library collection management and presentation software is essential for creating and delivering digital collection. But digital library technologies and contents are not static; continuous evaluation and investment are required to maintain the digital library as the needs for managing digital collections have increased greatly in the recent years. Some library software vendors such as DiMeMa, Luna Imaging Inc., Endeavour Information System Inc., etc. have released commercial digital library management systems. To avoid expensive license fees, a good option is to use open source alternatives. Open source application allows developers and users to modify and tailor it to their own particular needs. Like commercial software, open source software will not be a perfect solution, but gives developers and users the opportunity to modify functionality and create interfaces for integration with other software[1]. Some major open source software projects, for building and maintaining digital collections, are Dspace, Fedora, Eprints, and Greenstone. Greenstone is a suite of software for building and distributing digital library collections that provides a way of organising information and publishing it on Internet or on removable media (e.g., CD ROM/ DVD). Greenstone aims to empower users, particularly universities and research libraries, to build their own digital library collection in the fields of education science and culture.

2. FUNCTIONS OF A CORE DIGITAL LIBRARY

Digital library systems comprise a family of automated systems that together provide a comprehensive capability to manage the digital content of an enterprise. Capabilities of a digital library system can be divided into the following areas:

- Capture or creation of content.
- Indexing and cataloging (metadata).
- Storage.
- Search and query.
- Asset and property rights protection.
- Retrieval and distribution.

In addition, a digital library has the following salient features:

- It enables greater access to digital contents, which can be managed from remote locations and provide

Received on: 23 February 2010; Revised on: 15 May 2010

a way to enrich the teaching and learning environment.

- It enhances the power and capacity of library services and also provides workable solutions to problems of the print-based collections in traditional libraries.
- It provides access to multiple users simultaneously with continuous availability of documents.
- It reduces floor space requirements compared to conventional library to store the documents. The fifth law of Ranganathan, i.e., "the library is a growing organism" has always been putting pressure on traditional libraries for more and more physical space to accommodate new arrivals in the library. This problem is completely solved in a digital environment[2].
- It facilitates improved access to information by providing various sophisticated search and retrieval facilities.
- It facilitates information sharing among users through notification, file sharing and co-operative document preparation.
- It breaks the barriers of time, space, language, and culture.
- It facilitates improved collaboration among users which has profound impact on the scholarly information lifecycle—the process by which the researchers and scholars create, use and disseminate information.
- It reduces the gap between nation and people in terms of infrastructure, facilities and resources. Thus it reduces the digital divide[2].

3. OBJECTIVES

Astronomy is a subject with its roots in the study of images. These images are part of the presentation of scientific results; therefore, their archival for present and future use is necessary. There is a vast range of images of astronomical objects available on Internet. However, finding images of a specific object can sometimes be fairly difficult.

The Greenstone Astronomical Digital Image Library (GADIL) helps in collection of astronomical, research-ready images and make these available to the astronomical community and the general public. The other objectives of building GADIL were:

- To increase the astronomers' productivity through easy access to data.
- To provide an easy access to the scientific-quality images for the purpose of scientific and educational use.
- To encourage the further use of images that respects the scientific integrity of the data.
- To help astronomers in preparing figures for talks or papers.
- To create an interest in Astronomy among students.
- To increase public awareness and understanding of Astronomy as a subject.

4. SCOPE

GADIL is a prototype astronomical digital image library confined to a small sample of astronomical images available freely from the World Wide Web.

5. RELATED WORKS

Some of the existing digital libraries are:

(i) Bus and Coach Photo.com (http://www.busandcoach.com/simpleSearch.aspx?mode=all&list=1) image library is a comprehensive image collection developed by TATA Group. It is a commercial website for display of numerous buses and coaches manufactured by the company.

(ii) Gandhi serve (http://www.gandhiserve.org/cgi-in/if2/imageFolio.cgi?direct=Art/Drawings), a charitable foundation has created an online image archive of art, cartoon, documents, correspondence, etc. related to Mahatma Gandhi.

(iii) Headlines India.Com (http://www.headlinesindia.com/Archive/image_archive.jsp?j=1), an electronic news media has archived images related to India's news headlines.

(iv) National Archive of India (http://nationalarchives.gov.in/photo_gallery.html) has created a repository of images for the use of administrator and scholars.

(v) Freelance artist have created image libraries of own work for demonstration and publicity.

6. METHODOLOGY

As astronomy is a vast subject, a detailed analysis is required to understand the intension and extension ideas of the subject. This has been done by defining the subject and identifying the divisions and sub-divisions of astronomy using thesaurus and subject heading lists. Sample data was collected through keywords search from Internet.

6.1 Software Requirements

The following software were used for installation:

(i) Apache Web server.

(ii) JRE.

(iii) GSDL.

(iv) Imagemagik.

7. GREENSTONE LIBRARIAN'S INTERFACE

The simplest way of building a new collection is to use Greenstone Librarian's Interface. This allows collecting a set of documents, import or assigning metadata and building these into a Greenstone collection. .The librarian's interface can be run in one or four modes: Librarian assistant, librarian, librarian system specialists, and experts. The modes control the level of details within the interface and can be changed through preference in the file menu.

8. CREATION OF COLLECTIONS

8.1 Creation of New Collection and Selection of Metadata

To create a new collection, open the file menu and choose new (Fig. 1). Software provides various kinds of metadata, which can be selected on the basis of requirements. A new metadata set was created, which interprets the basic image data like object type and name, description, format (jpg, bmp), height, width, and resource identifier.

8.2 Gathering

Gathering implies selection of the files required for the collection. Dragging and dropping the files solve this purpose (Fig. 2).

8.3 Adding Metadata to the Document

While adding metadata to the document, the enrich panel comes into play (Fig. 3). The enrich tab brings a panel of information. On the left is the document tree representing the collection, while on the right, metadata can be added to individual documents. To view the document to which metadata is being assigned, double clicking on the document in left panel will open the document using the appropriate program.

8.4 Designing the Collection

Collection can be designed in the following steps:

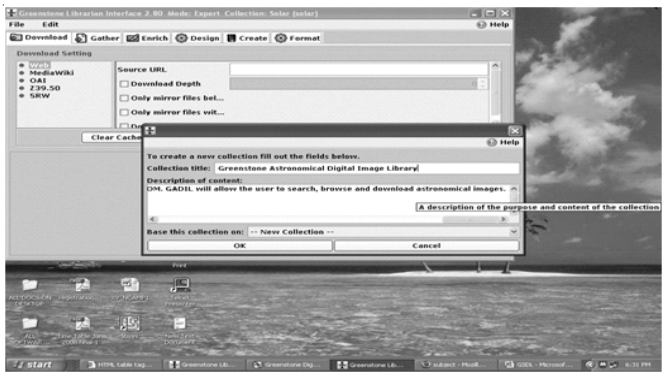

Figure 1. Creating a new collection.

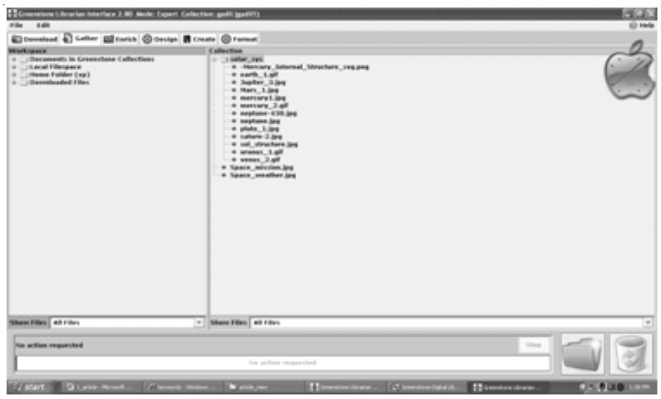

Figure 2. Gathering.

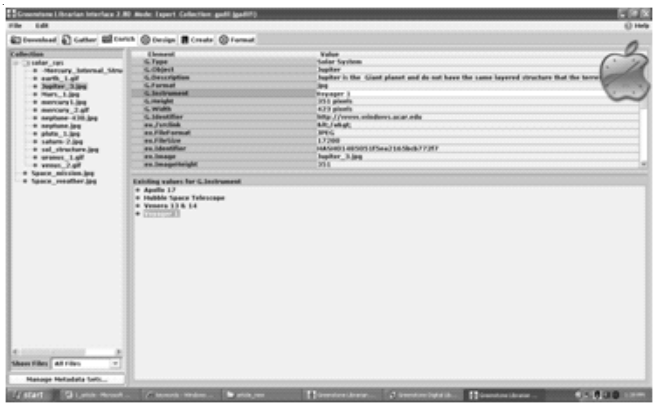

Figure 3. Adding metadata.

8.4.1 Document Plug-in

In Greenstone, documents and metadata specifications are imported by software modules called plug-ins (Fig. 4). Plug-ins enables Greenstone to support different document formats. It explains order of their appearance, types of plug-ins to be used, and what parameters to pass to.

8.4.2 Indexing

Indexes (Fig. 5) specify which parts of the collection are searchable. To manipulate an index command, the search index is chosen in the design panel. Browsing classifiers allows the user to browse the documents in a

collection; the browsing structures are generated automatically from the metadata that are associated with each document in the collection. Browsing classifiers are set up in the design panel. Classifiers generate a hierarchical structure that is used to display a browsing index.

8.4.3 Format Feature

Greenstone library webpages are generated dynamically when requested. Format commands are used to change the appearance of these pages—particularly how documents are shown in browsing and search result list. To manipulate a format command, the format feature (Fig. 6) section is chosen in the design panel, where use of html tags, metadata values, some customised format string items, and conditional expressions can be made.

9. FUNCTIONALITY/WORKABILITY OF GADIL

GADIL is a digital image library restricted astronomy and its related disciplines. It can be regarded as the first indigenous prototype digital image library developed using open source software Greenstone. GADIL can serve as an ideal access tool for astronomical community in India. It is available via the web allowing access to astronomical research-ready images with a user friendly interface at http://127.0.0.1/cgi-bin/library.exe. It is also available through self-installable CD-ROM. Figure 7 shows the home page of GADIL.

10. SERVICES PROVIDED BY GADIL

GADIL provides the following services to the user:

10.1 Single User Interface

GADIL provides different browsing classifiers (Fig. 8) for finding information such as search facility, type of images, subjects, and instruments used, etc.

10.2 Simple and Advanced Searches

GADIL provides both simple and advanced searches (Fig. 9) interface via the web, which in turn allows more targeted searching. For advanced searches (Fig. 10), different search fields like type of images, image objects, instruments used, and image formats, etc.) can

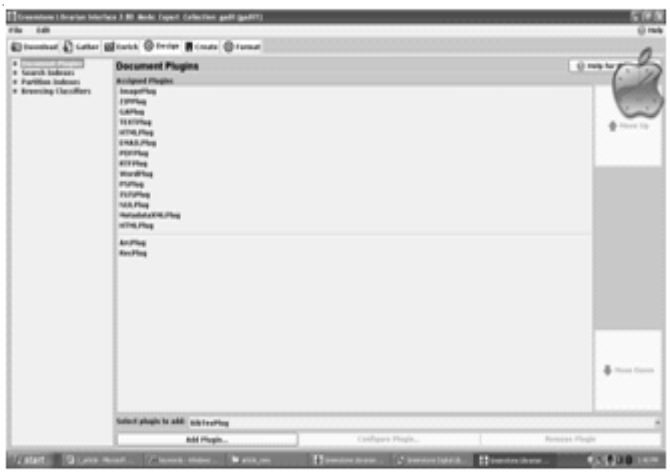

Figure 4. Document plug-in.

Figure 5. Index.

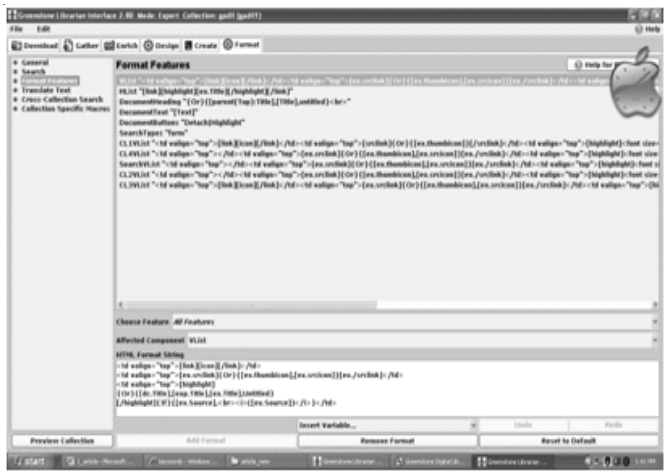

Figure 6. Format feature.

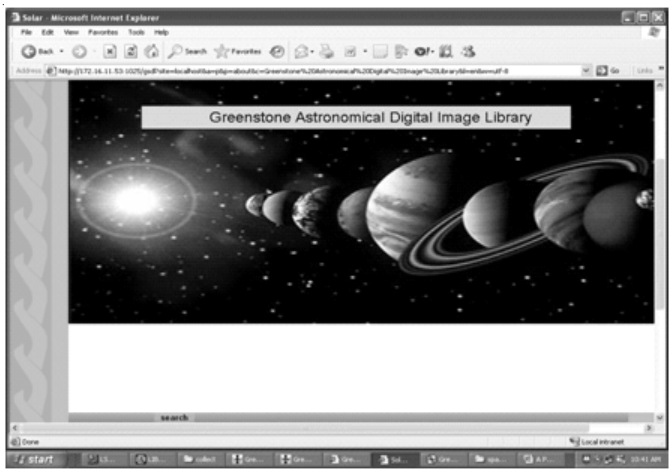

Figure 7. Home page.

be combined with each other and linked with logical operators AND,OR and NOT.

10.3 Browsing Facility

GADIL provides browsing facilities (Figs 11 to 16) by type of images, instruments used, subjects, and keywords. A list of images appear in an ascending order.

Figure 8. About page.

Figure 9. Simple search.

Figure 10. Advanced search.

Figure 11. Browsing by type of images in the collection.

Figure 12. Image results with description.

Figure 13. Browsing by instruments.

11. CONCLUSIONS

The preliminary results presented above suggest that the use of Greenstone as GADIL can be very useful for the astronomical community. GADIL not only aims to serve as a reference tool with research ready images but also as a didactic one with futuristic educational applications. During the course of development of the

Figure 14. Image results of browsing by Hubble space.

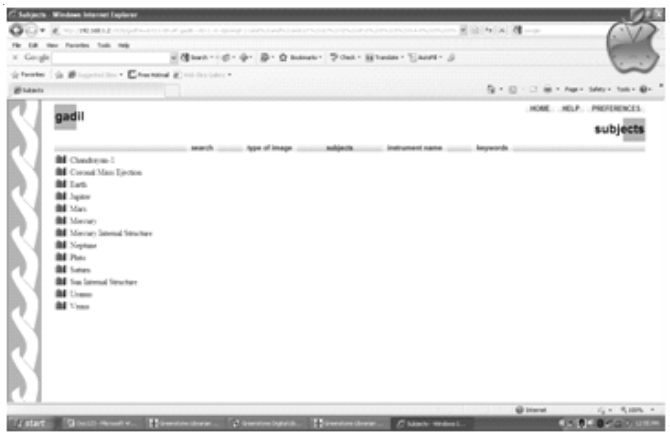

Figure 15. Browsing by subjects.

Figure 16. Image result for Mars.

GADIL, several issues related to the Greenstone software and otherwise, like metadata selection and editing as per needs, have been met. The metadata are the ancillary data needed to properly interpret the basic image information. Since there is a wide variety of ways in which a digital library may be customised, selecting metadata, editing and customisation keeping in view user's idiosyncratic requirements and possible approaches is a fairly tough task. Further, resolving problems related to the scientific data, i.e., images have been quite an issue. Format of the images, size and resolution, pixel examination, and color manipulation have been major challenges while developing the system.

GADIL, being an indigenous experimental prototype model, provides a beginning or standing platform for development of a comprehensive digital image library and is expected to serve as a top quality resource for scientists, researchers and students. It is important to mention that in future, if GADIL is made OAIPMH and Unicode compliant, the system would also be able to operate as a harvester thus facilitating interoperability and would eventually be successful in crossing the barriers of language, time, and location.

REFERENCES

1. Zhang, A. & Gourley, D.A digital collection management system based on open source software. *In* Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries, 27-31 May 2003, Texas. The IEEE Computer Society, USA, 2004. pp. 381. http://www.wrlc.org/dcpc/dcmspaper/ (accessed on 05.10.2009).

2. Bhatt, R.K. & Singh, K.P. Digital libraries: Emergence, features, challenges and opportunities. *In* International Conference on Digital Libraries, 24-27 February 2004, New Delhi. The Energy and Resource Institute, New Delhi, 2004. pp. 49-55.

About the Authors

Shri Md Nurul Alam graduated with Electronic Science and has a postgraduate degree in LIS from Jadavpur University, Kolkata. Professionally, he had worked with Electronic Media Library and Indian Institute of Technology, Kharagpur, and is presently working at Physical Research Laboratory under the aegis of Department of Space. He has keen interest in open source software (OSS) technologies and is constantly developing various application-based products using OSS technologies.

Ms Pragya Pandey graduated with Chemistry and Zoology and has a postgraduate degree in LIS and a diploma in Library Automation and Networking. Presently, she is working at Physical Research Laboratory.