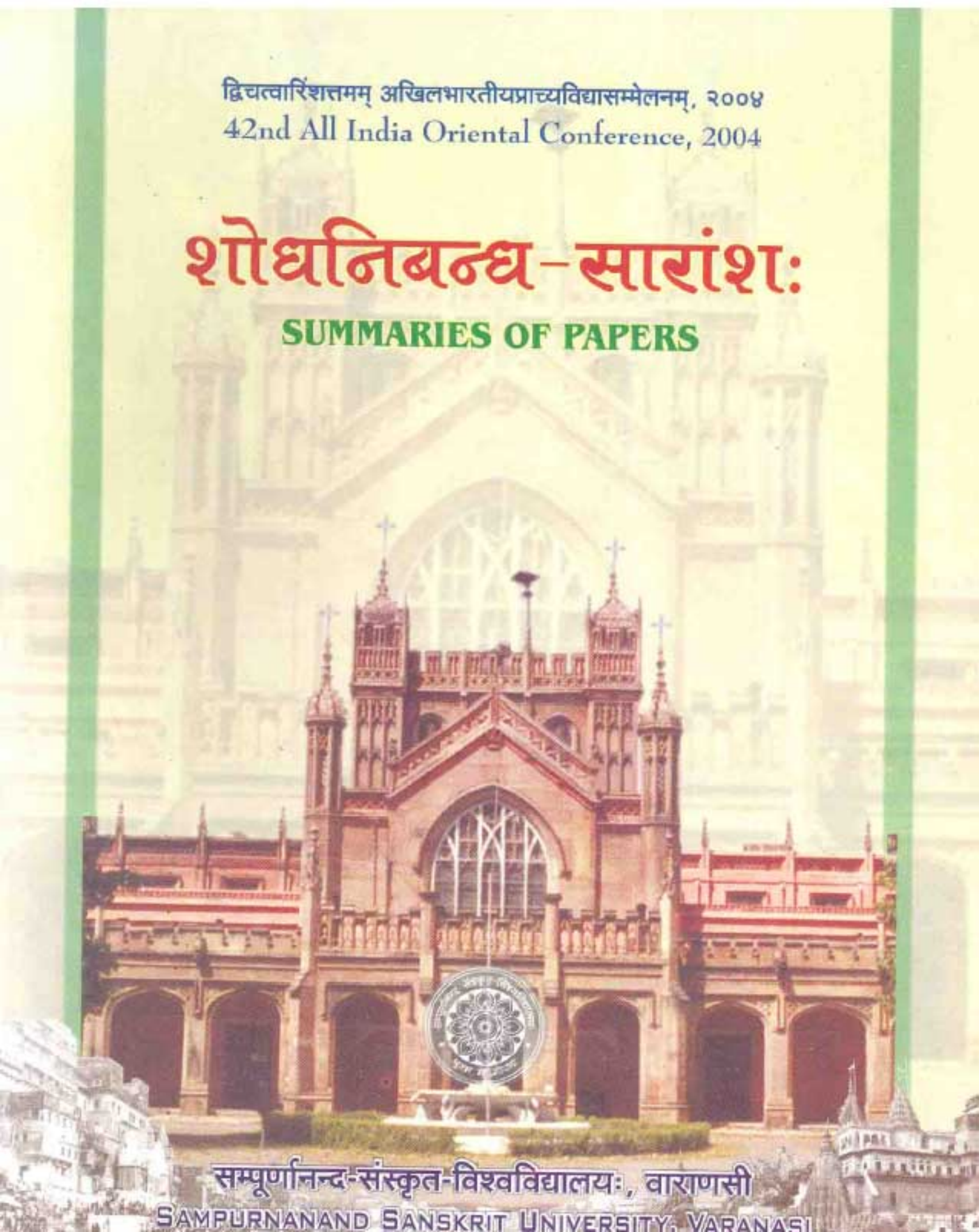
द्विचत्वारिंशत्तमम् अखिलभारतीयप्राच्यविद्यासम्मेलनम्, २००४

42nd All India Oriental Conference, 2004

शोधनिबन्ध-सारांशः

SUMMARIES OF PAPERS



सम्पूर्णानन्द-संस्कृत-विश्वविद्यालयः, वाराणसी

SAMPURNANAND SANSKRIT UNIVERSITY, VARANASI

All India oriental Conference (42nd Session), 4th to 6th November 2004, Sampurnanand Sanskrit University, Varanasi

DIGITAL ARCHIVING OF MANUSCRIPTS AND OTHER HERITAGE ITEMS FOR CONSERVATION AND INFORMATION RETRIEVAL

R. Raman Nair
ramannair_r@yahoo.com

0. Introduction

Information and Communication Technologies can help to save our heritage. Like environmental movements for saving endangered species of flora and fauna, there is an urgent need for saving the endangered heritage items too. Of this in the text sub-domain, a number of books, microfilms, microfiche, paper manuscripts, palm leaves, stone inscriptions and other such materials stand the risk of loss. Old paintings and sculptures are getting degraded every day. Traditional folklore items are vanishing along with the artists without any one taking interest in sustaining the art form after them. Music is also meeting a similar fate. Even ancient systems of life styles are getting wiped out. Life support heritage systems like Ayurveda and Siddha have already lost a considerable treasure of knowledge. All these have one thing in common - along with the loss of physical material heritage, the knowledge content in these also are lost. In the case of the material heritage, preservation technology has been applied for a long time. But efforts concentrating on the preservation of content are relatively new. It is here that digital technology will become increasingly relevant.

1. Expression of Cultural Heritage

Looking from the informatics angle, expressions of cultural heritage fall into texts, images (still and video) and sound categories. Multimedia technology can be used to conserve all these items. The text information may be palm leaf manuscripts, texts inscribed in stone

tablets, handwritten paper manuscripts, old printed records, books, microfilms, fiche etc. Ever since man started recording information and knowledge for communication, he has been perfecting or inventing techniques to make the process more flexible and transparent. The milestones in this direction are the invention of alphabets, paper, printing press, micro photography, and digital recording technique.

When Johannes Gutenberg of Germany invented the art of printing during the 1440s, the trend was to convert every manuscript into printed form. Microfilming technique helped to convert at least archival materials like ancient manuscripts, printed documents, journals and other records into microform. What is prevalent at present is conversion of ancient manuscripts and printed documents in to digital form. Now we have sculptures, paintings, monuments, books, journals, newspapers, and even museums and libraries in the digital format.

The print medium will be with us for a long time to come. But as a medium to retrieve required information from a large mass and to transmit information regardless of space and time, the traditional media and method, paper and print has got several limitations. The major limitations are:

- Printed documents occupies large space
- Finding a bit of information from a huge print collection is very difficult and time consuming;
- Related information, as graphics and sound cannot be integrated at one point, because print present information in linear format;
- The print media cannot provide interactivity;
- The quality of printed or written documents deteriorates when preserved for long period;
- Transfer of printed documents from place to place can be achieved only by physical means.

The emergence of microform helped to conserve manuscripts, paper documents and images without consuming much storage space. Transfer across space became easy due to their low weight and size. A number of materials were archived in microform which included palm leaf manuscripts, ancient records, maps, back volumes of journals, old and

rare books, etc. Although it ensured conservation and reduced storage space to some extent, other problems like selective dissemination, multiplication and transfer, cross-relating audio-video information, interactivity, etc could not achieved.

2. The Emergence of Digital Technology

These problems were solved when it was discovered in the 1930s that the 'ON' and 'OFF' states of an electric circuit could be interpreted as '1' and '0' and they could be used as the alphabet of a flexible media. Information recorded using this binary language is appropriately known as digital information. Although, the binary system or digital communication was invented in the 1930s, its large-scale application had to wait until the 1960s when computers acquired enough memory to store large quantum of information. The progress achieved in the digital technology and secondary storage media like high capacity hard disks and compact discs facilitated compact and durable storage of information of all kinds. The hypertext and hypermedia features enabled integrating text with graphics, sound, video and animation, which can be used in an interactive manner. In a networked environment the digitized information can be transferred regardless of space and time. Also, a bit of information can be retrieved from any part of the world in seconds and the digitized information lasts as it is for a long period. Thus digitization has brought in flexibility and durability in information storage and retrieval. Any information stored using the binary language permits fast retrieval integrating various components of text, graphics, sound, video etc.

3. Definition of Digital Archives

Digital archives, library or museum is actually a library or museum and not a mere mechanical store of digitized information of antique materials of historical importance. It is more than a mere storage of digitized content. A digital or virtual museum or archives is a fusion of resources in a variety of forms, including services and people supporting the entire life cycle of information. A digital archive works best when it is an integral part of a library that provides its users with access to information organized in the most useful way. A Virtual museum gives a hundred times more useful information and entertainment by enabling the jump from one exhibit to related one at any point in the sequence instead of sequential journeys inside the traditional museum. At the same time digital archives,

libraries and virtual museums share common goals with their traditional counterparts. Many of the existing definitions consider that digital archives, libraries and virtual museums are mere stores of digitized content available in CD or hard disk or through Internet. A digital archive should be a seamless extension of their traditional counterparts that provides scholars with access to information in any format that has been evaluated, organized, and preserved. Access to this evolving digital collection is enabled by the services of information professionals. The digital archives add value and save time while extending the hours of access. It reduces the need for physical nearness to information resources, but still emphasizes the quality of those resources. Digital Archives is a library or museum that can be individually customized and, ultimately, will be easy to use.

One of the simple and interesting example of digital archives, library and a virtual museum in one is the virtual Bird and Cat produced by Dorling Kindersley in Eyewitness Encyclopedia series meant for children. They are excellent virtual museums on those Cat and bird species which contains text, images, video, animation and sound files. The user feels like entering a big museum on cat species moves around the exhibits and standing in front of the exhibits in which he is interested calls for and uses information on a specific item as text, graphics, sound files, posters, video or maps. The same method can be used for any large museum or archives any advance nature also.

4. International and National Efforts

There are many national and international ventures for archiving of monuments, text, and multimedia. Some examples of international magnitude are; UNESCO Memory of the World Programm, American Memory, Digital Libraries Initiative (DLI), Ankorvat Heritage Conservation Project, Ragamela, Indian National Mission for Manuscripts National Library, and Kerala History DL

UNESCO memory of the World Programme seeks to safeguard documentary heritage of the world. Naturally, it reflects the diversity of languages, peoples and cultures. 'It is the mirror of the world and its memory'

American Memory is American history in words, sounds and pictures and extensive online exhibitions of historic photographs and documents.

Digital Libraries Initiative (DLI) of the National Science Foundation, Department of Defense Advanced Projects Agency and the National Aeronautics and Space Administration funded six research projects in 1994 aiming at developing new techniques for creating digital libraries.

Ankorvat Heritage Conservation Project is part of World Heritage Inventory and Management project supported by UNESCO. It is meant for ancient site of Angkor, Kampuchea and specifically related to conservation of Preah Khan, a twelfth century temple complex in the ancient city of Angkor, Kampuchea.

5. Developments in India

In India, several institutions and national level organizations like Ministry of Culture under government of India, Indira Gandhi National Centre for Art, The National Library, Centre for south Indian Studies, and Tamil University have taken steps to develop digital libraries. Government of India has initiated a project named Manuscript Mission. However, all projects aimed at creating digital libraries concentrate only on special collections.

National Mission for Manuscripts was launched in February 2003 by the Ministry of Culture, Government of India, to save the most valuable but less visible items of heritage. An ambitious five-year project, the Mission seeks not merely to locate, catalogue and preserve India's manuscripts, but also to enhance access, spread awareness and encourage their use for educational purposes. Working with specially identified Manuscript Resource Centres (MRCs) and Manuscript Conservation Centres (MCCs) in states all over the country, the Mission has collected data on manuscripts located in a variety of places, from universities and libraries to temples, mathas, madrasas, monasteries and private collections. The Mission also brings manuscripts and the knowledge they house

Ragamela is a multimedia archiving project initiated with support from UNESCO at Indira Gandhi Centre for Arts (IGNCA) which has developed a programme for archiving cultural and historical environment for enabling the future to experience the past using virtual technology. A digital archive on Indian culture has already been developed at IGNCA.

National Library at Calcutta has initiated programmes to digitize its rare collections. A number of utilities have been developed there to facilitate creation of digital libraries using solutions offered by various groups.

Kerala History is a digital library of basic texts on Kerala History, which is developed keeping in mind the vision of UNESCO ("developing digital content of archival and rare materials at regions where they originated"). It is a joint venture of Centre for South Indian Studies and CIRD, Thiruvanathapuram, Kerala.

But presently none of the projects are following any standard for these activities. It is necessary to make professional approaches towards the development of appropriate utilities, national level standards and to train a new generation of archivists and information professionals to create and maintain digital archives and museums.

6. Digital Archiving Packages

Keeping these professional goals in mind UNESCO and a few other organizations working on conservation of heritage have developed various open source packages to achieve selective access to information from digitally organized text, graphics, sound and video. Of this the important items relevant to conservation of heritage items of India are GenISIS, Greenstone, DSpace and Nitya. Of this Nitya which can process a few regional scripts of India developed by Centre for Informatics Research and Development is a best example that can be used to describe the features of a digital archiving solution.

Nitya developed in India combines high-level text compaction technique and highly sophisticated free text search and retrieval procedure. In Nitya any piece of information can be searched out from a huge store of information within seconds. Searches can be by title, by author, by date of publication, by tablet, roll or leaf number, by a keyword or combination of keywords using the Boolean logic (OR, AND, NOT). It permits proximity searches ("NEAR"). Although, several systems are in vogue for the digital storage of information, they lack efficient search mechanism. Devoid of a search mechanism, digitally stored records or information in hard disks and compact discs remain very much like unorganized collections of paper documents or microforms. The full potential of digital technology can be exploited

only if there is a strong search mechanism at front end. Then only it can perform the traditional functions of an archival collection.

7. Features of Digital Archiving Solutions

Documents like palm leaf manuscripts, handwritten paper records, old books, etc. are first scanned and converted to Adobe's Portable Document Format (PDF) or a similar one which can control permissions. Images of the leaves or pages are then book marked and annotated. Adobe PDF enables extensive navigation inside a document. A database is created for the digital documents and page by page content index is also prepared using a powerful database management system. For this Nitya uses CDS\ISIS, a powerful bibliographic DBMS developed by UNESCO. The efficiency of the search depends on indexing techniques applied to build up Inverted file.

Nitya offers full functionality of windows. At front end it exhibits 4 distinct areas which are static windows for index, query, hits and references. The same is shown in Figure-1.

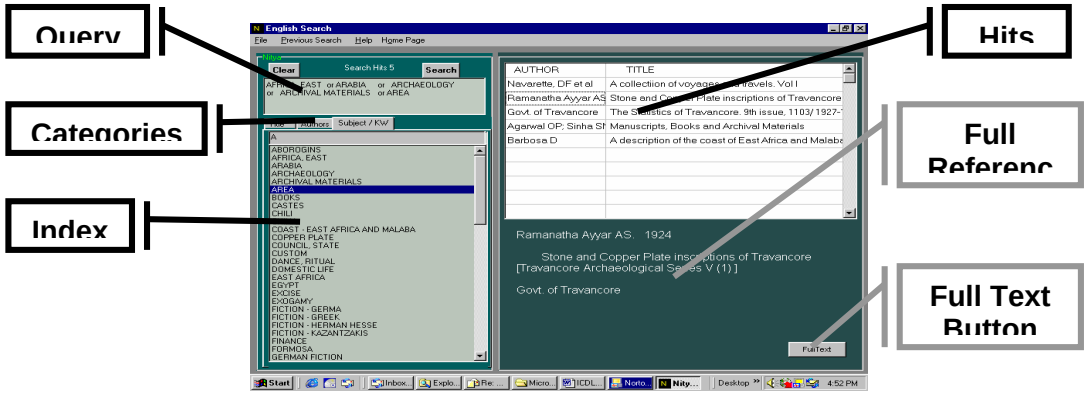


Figure -1: Search Mechanism of Nitya

Index contains searchable terms arranged under different categories such as subject, author, etc. Number of categories depends on the type and characteristics of the collection. For example, categories of an archive of palm leaf manuscripts may be author, title, leaf number, first line, subject, etc, where as categories of an art gallery can be artist, media,

title of picture, collection number, collection or owner etc. For an archive of theses categories may be the research scholar, guide, departments, university, subject, etc. Search terms from the Index can be selected by a few alphabetical strokes. Clicking on the Index, terms are transferred to `Query' area.

Queries are formulated using Boolean/proximity operators. Queries are submitted to search and results are first exhibited in the Hit area and then in Reference area exhibiting full bibliographic details.

When the 'Full Text' button is clicked, the original document that is book, manuscript, image, sound or video is opened in concerned programme. If it is text it opens in Acrobat Reader and there after the navigation through the document is made possible using book marks and other hyper links. Manuscripts like those in palm leaf can be digitized along with there word processed files which will be easy to use. To check the authenticity or correctness of the word file the leaf also can be viewed along with the word file. Palm leaf manuscript with content transcribed in traditional Malayalam script: in text format kept as pdf file retrieved from a digital library and opened in Acrobat Reader is shown in figure -2.

Any part of the record or palm leaf, manuscript or painting can be zoomed up to 1600 percent. Also, any word can be located from any part of the displayed record with the help of Acrobat Find / Search option.

DL packages like GenISIS, Greenstone, Nitya etc are based on user approach to knowledge and books in traditional library and archive environment. When a user enters a library, he first goes to the catalogue and searches the catalogue under authors, titles, subjects, etc. This leads one to the shelf where the document he is looking for is shelved. Once he gets the document he locates relevant chapters / pages / portions of the book with the help of the content page. In the same way, when one opens the Digital Archives in the computer, an Index of Author/ Title/ Subject appears and terms selected from the Index are used to formulate queries. Search produces a hit list of relevant documents. List leads to records with full bibliographic details.

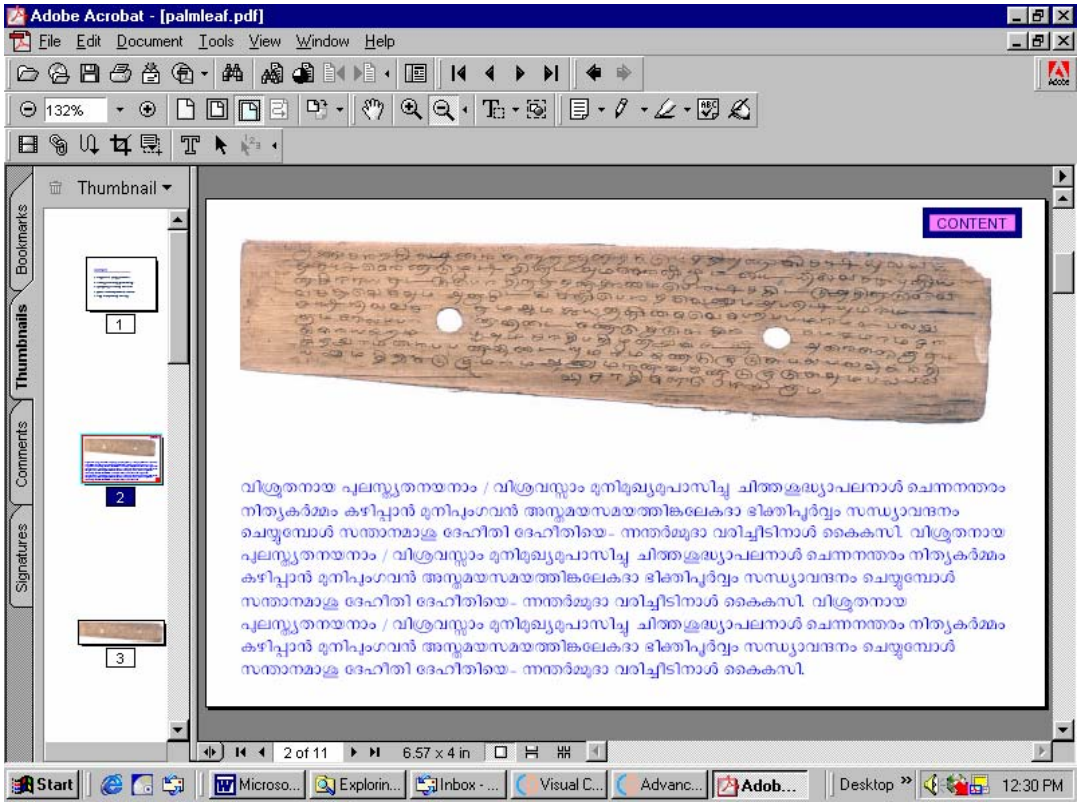


Figure - 2: Palm leaf with its content transcribed in traditional Malayalam script retrieved and shown in Acrobat Reader

Records lead to full text which is opened in Acrobat Reader. Here the entire text can be navigated through book marks. In a similar way navigation in the museum or archive or art gallery is achieved in virtual systems.

8. Possibilities of Digital Archiving Systems

Unlimited number of documents can be digitized, stored and managed using digital archiving packages like GenISIS, Greenstone, DSpace, etc. Older books, manuscripts in palm leaves, dissertations, old journals, paintings, drawings, etc can be digitally archived and stored in compact discs or special storage systems and media meant for that. Audio and video files can also be archived by them and they can be opened in their own

multimedia programs. For scanned texts, paintings, palm leaves, etc. Acrobat and PDF format is found to be most adaptable. Retrieval of a document in such systems is carried out in a few seconds even when total documents or records in the collection are large.

Digital Archiving packages like Greenstone, Nitya etc makes use of proven documentation principles and procedures for organizing the archives. Different kinds of materials demand different methods of indexing in order to achieve maximum benefits out of archiving. For example, documentation methods applied to archive and retrieve paintings of a museum should be different from that of archiving palm leaves of a manuscript library. It will be still different in archiving old legislative proceedings. Unless proper documentation methods are applied, any archival system would be ineffective. Since Greenstone, GenISIS, Nitya etc uses the database management system named CDS/ISIS, defining databases and applying indexing methods are naturally adhered to cataloguing and documentation principles. These programmes uses similar technology committed to translate age old practices of traditional archives and libraries to modern ICT practices in organizing knowledge.

9. Conclusion

In developing digital archives, libraries and virtual museums the solutions developed for them should adhere to international standards prescribed by organizations like UNESCO. For example the Common Communication Format (CCF) developed by UNESCO can be followed for developing databases at the front end of digital archives, libraries and museums. Following such standards can enable developing different categories of archives like those for theses, research reports, technical reports, conference proceedings, historical documents, and manuscripts in palm leaves, journal articles, paintings, and paper clippings, sound files, film an video etc and even integrating them at any time based on sound documentation principles to check the proliferation of non-documented digital collections that ‘dumps’ thousands files in digital storage systems without proper retrieval mechanisms.

Following standards like CCF will enable easy exchange and pooling of information and documents generated in archiving systems at national and international level. Digital collections developed can be distributed to any number of users. This will facilitate electronic publishing and marketing of special information sources of a library, archives or

museum. Customization of similar packages for Indian languages is also a very important area where developments needs to be initiated for organizing and preserving our cultural heritage which exists in different locations and are partly or in full based on regional languages and scripts.

References

Griffin, Stephen M. Digital Libraries and the NSF/DARPA/NASA Digital Libraries Initiative. Libraries for the New Millennium, Ed by David Raitt. London, Library Association, 1997.

http://www.loc.gov/

http://www.unesco.org/webworld/mdm/index

Hussain K H, Raman Nair R, and Raveendran Asari, K. Importance of Search and Retrieval in CD-ROM Full Text Publishing: Experiments Using PDF Documents and 'Nitya' Archival System. Information Studies, Vol. 8, N o. 3, July 2002

Raitt, David. Ed. Libraries for the New Millennium: Implications for managers, London, Library Association, 1997, 288p.

Raman Nair, R. Charithrathinte Uravakalkku Samrakshanam (Conservation of Sources of History). Veekshanam, 19 July 1987.

Raman Nair, R. Microfilimum Granthasamrakshanavum (Microfilm for Conservation of Documents). Vijnanakairali 29(7), July 1998.

Ravindran Asari, K.; Hussain, K.H. and Raman Nair, P. Nitya archives: Innovative blending of Techniques for selective Access to Information from Digitally Organized Text (SAIDOT) Proceedings of the National Conference on Information Management in e-Libraries, 26-27 February 2002, IIT, Kharagpur. Allied Publishers Limited, New Delhi, 2002.

Rowley, Jennifer. The Electronic Library. 4th ed. London, Library Association, 1998.

Sulochana Devi, L. Information System for Research in Sanskrit and Indology. In the Proceedings of the National conference on Vedic Sciences. Madras, JGRCVS, 1997

Witten, Ian H. Visions of the digital library. International Conference on Asian Digital Libraries, 4th, Bangalore, 2001. Proceedings: Digital Libraries, ICDAL 2001.