Proceedings

Conference on
Recent Advances in Information Technology

READIT 2005

14-15 July

Theme:
Digital Libraries to Knowledge Systems

Organised by

Madras Library Association- Kalpakkam Chapter
&
Scientific Information Resource Division
Indira Gandhi Centre for Atomic Research
Kalpakkam-603 102 Tamil Nadu

Proceedings of the Conference on

Recent Advances in Information Technology

(READIT - 2005)

July 14-15, 2005
IGCAR – Kalpakkam

Conference Theme

Digital Libraries to Knowledge Systems

Organised by
Madras Library Association - Kalpakkam Chapter
and
Indira Gandhi Centre for Atomic Research , Kalpakkam.

Foreword

Technological innovations in general and the revolutionary advancements in the Information Technology field in particular have necessitated the Library and Information Service personnel to be on their toes during the last few decades. Catering to the demands of the patrons, especially of those in Science and Technology has been a daunting task these days.

To be abreast with the changing environment and to meet the user demands by proper assimilation and application of technology is what is required of the Library and Information Services now. For this, interaction and exchange of information amongst peers and with the trendsetters in the field is absolutely necessary. The motto of the READIT series of conferences had been this. Forums like READIT provide an opportunity to share our ideas and experiences. Our past experience shows that this exercise is really a success.

Computers and Information Technology have given us this profound 'Digital World'. As we are wading through the 'Digital World' new horizons of 'Knowledge World' is beckoning us. We hope READIT-2005 will equip and enlighten the Library and Information Service professionals to migrate to the new pastures.

The Conference Proceedings brings to highlight the deliberations to a wider audience. Also it will serve as a reference document providing an opportunity for us to learn and evolve to meet our vision and mission needs.

M.Somasekharan

Convener, READIT-2005

CONTENTS

Technical Session III : KNOWLEDGE MANAGEMENT

Technical Session IV : KNOWLEDGE SHARING TECHNIQUES

Technical Session V : LIBRARY THE GATEWAY

Facets of Digital Library

Digital Collection Development

Dhanasegaran G.

Abstract

Explains the components of digital library - Discusses the contents of a digital library collection – Describes the digital collection infrastructure - States the criteria for collection development of digital information resources – Elucidates E-Journals, E-Books, E-Reference books-Talks about net library – Enumerates the steps in building digital collections- Discusses Digital Information Collection in India.

1. Introduction

Libraries have a number of machine-readable publications and databases that facilitates remote access. Their aim is to provide instant access to digitized information. In digital libraries, the information is stored in electronic or digital medium such as digital books, scanned images, graphics, textual, numeric data, films, audio and video clippings etc. The basic concept of digital library is to exploit the facilities of Information Technology. Further, it has to provide the opportunity of accessibility to every piece of information in any desired format from any remote location.

2. Components of digital Library

In a digital environment, only digital information is disseminated. Retrieval Software is produced locally and most of the information is obtained by remote accesses. The three major components of digital libraries are

1. Documents (Print, audio, video)
2. Technology
3. Operations (Research, education and others)

3. Digital Library collection

A digital library collection contains fixed and permanent documents available in the digital format. A digital library contains different types of materials that include multimedia objects representing different subject fields. A Hybrid library is a combination of both the traditional as well as the Digital Library resources range from printed resources and documents available on CD-ROM to online databases and internet sources. Some digital library collections may be already in digital form while others are digitized from their analogue versions.

3.1 Collection infrastructure (acquire, create convert and access)

The term ***acquire*** is to hold all kinds of digital material such as discovery, searching, selection, ordering and receiving each of the five important ingredients which constitute books, journals, films, newspapers, pamphlets or microforms.

Dr.T.P.M.Library, Madurai Kamaraj University, Mdurai-21

Create is the process that includes statistics, mathematical calculations, filing tasks and all traditional activities such as automated teller machine and also converting the raw data into information.

Similar to the provision of information in the Traditional Library environment, ***access*** is the new way of providing rights to access the digital resources in the Digital Library. Digital library has the potential coverage of working in different network platforms and gain access to information using devices like telephone, internet etc.

3.2 Collection Development of Digital Information Resources

Selection of digital information resources is a responsible task of library professionals. The process of selection has to be made carefully keeping in view of the characteristics and goals of the institution and also its user community, their nature and information requirements. New digital library services are on their way to create a union catalogue of specific type of digital information resources. The information professionals usually scan the web using standard search engines for selecting relevant digital information sources like e-books or e-journals and E-Reference books

Important guidelines have been suggested by eminent scholars for the selection of library materials in the Traditional Library environment. They are

- The authority of creations(Author, Publishers, etc.)
- Scope (the breadth and depth of coverage)
- Treatment and level (Suitability for the intended audience)
- Arrangement (Organisation of Content)
- Format (Accessibility , searchability, readability, portability, durability etc)
- Special features (What shows their individuality)

The following major factors are considered to be more useful for the selecting digital information sources.

1. Hardware , software and network requirements
2. The content of digital networks
3. The version of the product
4. The number of concurrent users allowed
5. Access control through password, proxy server, authentification etc.
6. Price, licensing and copy right agreement
7. Database features, the retrieval engine and user interfaces
8. Ease of use and user training
9. Archiving and preservation of the digital information sources

For proper collection development, it is essential to know whether a particular information product has to be used only with the interface that comes with that product. Many digital libraries provide users with only one interface to search a wide range of products. The equipments used for digitization and the design architecture of a digital library influence the selection of digital information resources. The internet archival collection has tremendous impact on the selection of digital information resources through websites.

3.3 E-Journals

Electronic journals form a major part of a digital library collection. The publishers of E-Journals provide access to their journals to the licensed users. Some publishers bring out their journals both in print and electronic version. Access to e-journals is an important issue and several models are now available. Users need to use a password to get access to a particular e-journal from an agency or provide their IP addresses to the publishers if they have leased line facilities.

3.4 E-Books

Electronic books normally appear on CD-ROMs. E-books are cheaper and can be delivered instantly. They are portable and one can carry hundreds of books at a time. They are easily searchable and are easy to use. Some of the packages necessary for reading e-books are

1. Microsoft reader
2. 'RCA e-book
3. Adobe Acrobat –Book reader
4. Mobi pocket reader
5. Gemstar e-book
6. Soft book reader

3.5 E-Reference books

Many reference books are also brought out in CD-ROM formats and available on-line through payment. There are a number of reference sources available freely on-line through Internet. Since the reference books are supposed to be used by all types of users, the selection procedure must be liberal.

3.6 Net Library

In the digital environment, net-library concept has come up as a model for e-book. It offers access to the full text of a large number of books. The users of this library can search, retrieve and read those books on line. The users of each member library are asked to register for the first time and then the users can select one or more e-books for their current and future use.

4. Building a digital collection

Digital Library Development is not a single process. It involves various steps like

- **Scanning** - This process involves conversion of paper documents and other media in the existing collection to digital form

- **Indexing** - Identification of meta data of the scanned images, storing them in the database and indexing.

- **Storage** - Transferring scanned images to external storage devices like CD's DVD's etc.

- **Retrieval** – Development of a search software for the scanned images and displaying them in a user friendly format

5. Digital information collection in India

In India the concept of digital library is in its infant stage. Now many commercial sectors have taken the initiatiative in publishing the documents in electronic(CD-ROM) form. The popular Business India magazine is available in CD-ROM since December 1995. The other magazines like 'Time' and 'PC Magazine' are already issuing CD versions. A few publishing companies are now releasing the sources in multimedia which are available for a cost ranging from Rs. 900/- to 2000/-. Further, there are also certain digital information packages providing enormous wealth of information on current aspects. Some of the examples are Provess 2.0, Industry Analysis Service Version 2 etc by CMIE, Chennai (Centre for Monitoring Indian Economy Pvt. Ltd.). Such commercial databases are encouraging the librarians to develop their own local databases.

There is no legislative provision to make it mandatory for the publishers to deposit electronic and optical publications with a national centre. The National Centre for Digital Publications was established at the Foundation Innovation and Technology Transfer (FITT) at Indian Institute of Technology, Delhi.

The products that need to be covered are :

1. CD Media
2. DVD and
3. Internet
4. Audio tapes and discs
5. Audio video tapes and discs

Based on the availability of resources the centre is concentrating on the following areas:

1. Knowledge and information products
2. Titles published on CD media.

The centre has adopted a mechanism for locating publications. In the process of acquisition, the centre procures all the products costing up to Rs 20,000 after reviewing the contents by a committee. The centre compiles a union catalogue of CD titles held in Indian libraries. The members get a copy of union catalogue free of cost. The website being set up exclusively for the centre helps the users to get a general description of Indian CD publications available in the market. The publishers would get national and international publicity for their products. The national centre for digital publications is a milestone in the digital information collection of India.

Also in India, the Sustainable Development Networking Programme (SNDP) is funded jointly by the United Nations Development Programme (UNDP) and International Development Research Center (IRDC) of Canada and is being implemented by the Environmental Information System(ENVIS) of the Ministry of Environment and Forests with technical support from the National Centre for Software Technology(NCST), Mumbai. There are more than 25 ENVIS centers spread all over the country and

disseminate information on specialized subject areas in digital information on the environment.

The National Informatics Centre , India extends digital packaged delivery service through its network. In today's changing scenario, Information and Library Network (INFLIBNET) centre, Ahmedabad is playing a vital role in bringing information technology culture and establishing infrastructure in Indian Universities and Research Institutes. Earlier INFLIBNET had been concentrating more on automation and now the main focus is on digitization of potential resources such as Doctoral Dissertations, Working papers, Reports, Seminar proceedings etc generated by the Universities.

UGC has taken the initiative to strengthen the INFLIBNET Centre with required manpower to implement, monitor and execute UGC INFONET programme in many University Libraries. At present, most of the university libraries are on the move to provide web-based reference and information services in the digital environment.

6. Conclusion

The Digital Library has been a mirage in Indian environment. But the concept has gained momentum by the UGC's INFLIBNET, SNDP and the like. Traditional libraries are now on their way to digitization in a phased manner. In India, collection development of digital libraries is greatly influenced by a number of stakeholders such as Library and Information Science Professionals, Publishers, Subscription Agencies, Database managers and Information Service Providers. The day is not far off to visualize a library with only digital resources.

References

Mahapatra, M …[etal.] ed. Information Management in Academic and Research Libraries.-Ahmedabad: INFLIBNET centre, 1998.pp322-351.

Murthy, T.A.V ed. First Convention Planner-2003.-Ahmedabad: INFLIBNET Centre, 2003. pp.1-384

Murthy T.A.V ….[etal.] ed. Road map to New Generation of Libraries using Emerging Technologies.-Ahmedabad: INFLIBNET centre,2004. pp1-613.

Urs Shalini,R., Rajashekar,T.B, Raghavan,K.S. The 4th International Conference of Asian Digital Libraries .-Chennai: Multicraft ,2001.pp.1-456.

The Tamilnadu Dr.M.G.R. Medical University Virtual Library

Krishnamoorthy G. and Jayamani N.C.

Abstract

The TamilNadu Dr. M. G. R. Medical University (TANMMU) is the largest Medical University in India. The Ministry of Health and Family Welfare, Government of India has accredited this University's library as a Regional Medical Library (RML) for South India. The University has established a digital library known as – MEDICAL INFORMATICS CENTRE since 1992. TANMMU has proposed to establish a virtual library for wider access within its campus and also to all the medical colleges affiliated to TANMMU. This paper presents the proposal draft for setting up one such virtual library. It has the support from the State, Central and its own source of funds. It is at the planning stage and envisaged to be implemented by 2005.

1. Background

The Tamil Nadu Dr. M.G.R. Medical University Act, 1987 (Act No.37 of 1987) received the assent of the President of India on 24th September 1987. This affiliating University started functioning from July 1988 and is governed by the said Act. It is one of the premiers Medical Universities of India and it is the second largest Health Sciences University in India. The chief mission of this university is to achieve and maintain uniformly high standards of needs-responsive medical education in the State of TamilNadu and to develop research facilities and to provide for research and for the advancement and dissemination of knowledge in the field of medical sciences. The motto "HEALTH FOR ALL" reflects all the objectives of this Medical University. To accomplish its prime objective of this University to promote research in the core and allied areas of Medicine and Health Care & Delivery System, The TamilNadu Dr. M.G.R. Medical, as a first has set up Central Medical Library in mid-90s.

The Director General of Health Services, Government of India, New Delhi has given accreditation to this University Library as a Regional Medical Library for the Southern Region. The Regional Medical Library has established Medical Informatics Centre, a digital library in 1992. This Centre plays an important role in supporting the mission of the University by enabling it to advance, transmit and sustain knowledge and understanding by providing and promoting access to recorded knowledge.

To this end, it is imperative that this Centre needs to be improvised to extend its services to a wider access. Hence, it was decided to plan for a Virtual Library. Virtual Library is a library in which the holdings are found in electronic stacks. It is a library that exists, without any regard to a physical space or location. It is a technological way to bring together the resources of various libraries and information services both internal and external, all in one place, so users can find what they need quickly and easily.

The Tamil Nadu Dr MGR Medical University Library,Guindy,Chennai.

2.Justification of the Project

The changing information environment with increasing availability of electronic sources, particularly in the medical field, and the increasing information demand from the medical fraternity warrants the existing medical libraries to redefine its objectives, scope and services to suit the current trend. Hence, the need for a Virtual Medical Library to facilitate access to all types of electronic sources in the field. TANMMU being the mother medical university holds the responsibility to initiate requisite action to establish a virtual library for the benefit of its user community. TANMMU has already received adequate funds from the State and Central Governments plus its own funds to initiate this project. As a first step towards this, is the preparation of the proposal document placed before the Committee for its approval. It has been approved and implemented from January 2005.

3.Objective of the Paper

This paper intends to:

1. highlight the features of the proposed project; and
2. present the progress of the Virtual Library project at The Tamil Nadu Dr. M.G.R. Medical University, Chennai.

4.Features Of The Project

4.1. Project Objectives

The focus of the proposed project is to digitalize the TANMMU – Housekeeping functions, Infrastructure, Resources and Services.

4.2. Scope

The project attempts to network at two levels.

1. Phase-1 - Campus Network (LAN)

Local Area Network (LAN) plans to link all the departments in the Dr. M.G.R. Medical University campus, Guindy, Chennai.

2. Phase-II - Regional Network (WAN)

Regional Network plans to link all the 209 institutions in Tamil Nadu affiliated to TANMMU (Appendix-1).

4.3. Duration

The total period of the proposed project is set for two years.

Phase-1—January, 2005 – December, 2005

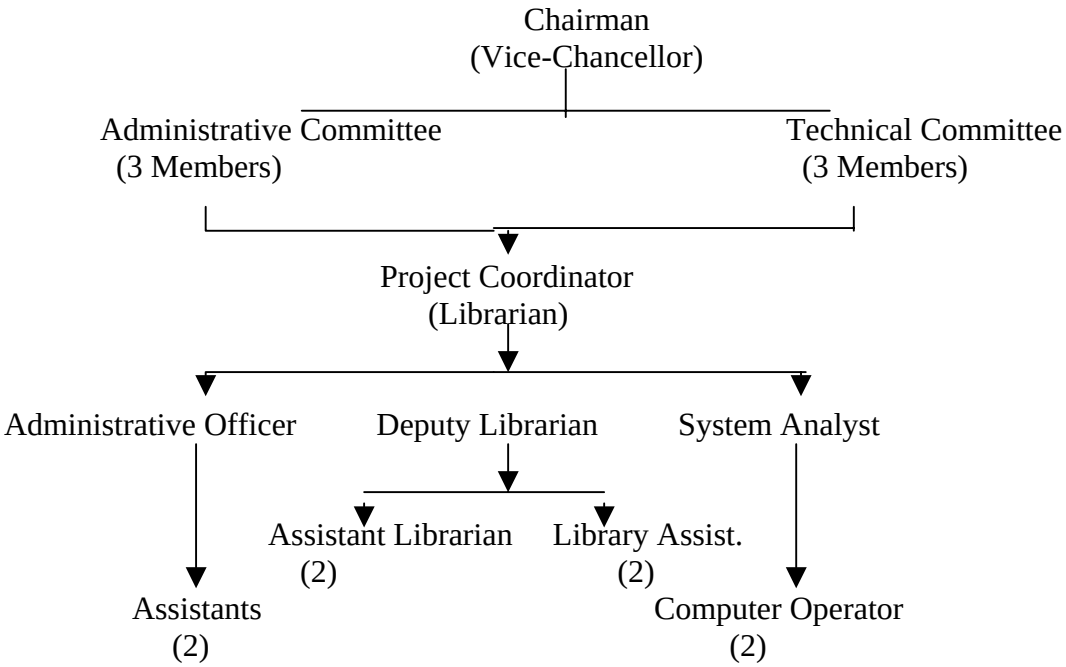
Phase-2--- January, 2006 – December, 2006

4.4. Requisites

4.4.1. Manpower

The staff structure involved in the project is as presented in the following chart:

Staff Structure- Virrual Library



These project personnel need to continue after the set Phase period also towards the sustenance and maintenance of the project.

4.4.2. Materials

The materials required for the project include both equipments - hardware, software and fittings & furniture - and the reading materials- information resources. The details and description of the required items of each category is furnished below:

A**. Equipments**

HARDWARE

S.No	Description
1	Servers with RAID-5
2	Thin Client
3	MS Windows 2003 Server
4	Win 2003 Client Access Lic.
5	Terminal Server Client Access License
6	Automatic Monitoring System including one CCTV with rotating Colour Camera (5 Nos.), Pan & zoom facility.
7	Dot Matrix Printer
8	A4, 52 ppm Laser printer
9	A4,20 ppm Color Laser printer
10	10 KVA online UPS, SMF batteries with one hour backup
11	CAT6Network cabling work, 24 port L2 switch (2 Nos.), 12U wall mount

	rack
12	Leased Line 1 mbps port charges
13	Router
14	Access control system

SOFTWARE

The supporting software required for the design, development and implementation of the Virtual Library is under discussion. It may be decided in favour of outsourcing the entire process.

FURNITURE & ELECTRICALS

1	Computer Table for 35 Thin client computers & Revolving chairs
2	Partition work
3	False ceiling
4	Split Air-conditioners
5	Electrical work
6	Illumination

B. Information Resources (Reading Materials)

- Full text databases
- Bibliography databases with abstracts
- E-Journals- under Negotiation
- E-Books
- E-learning tutorials.

4.4.3. Finance

The source of funds and finance support and the budget estimate of the proposed project is detailed below; Source of funds The two major sources of funds for this project include the Central and State Governments. The total grant is Rs.85, 00,000. AboutRs.50, 000/- is from the Centre and the balance amount of Rs 35, 000. - is supported by the State.

A. Expenditure Estimate

The grant money 85 lakhs has been distributed as: 50 lakhs towards equipments and 35 lakhs towards purchase of information resources. An additional amount of 50 lakhs required towards staff salary, maintenance and updating the information resources has to be met by the TANMMU

5. Work Plan

The project work involves:

- Planning, Preparation and implementation towards installation and implementation of LAN and WAN for the proposed Virtual Library
- Consortia development
- Design and development of web portal
- Conversion of traditional/print to electronic version
- Updating of E-Resources
- Evaluation of the project

6. Project Progress

The project of the Virtual Library has commenced with effect from January 27th 2005 in the newly built, separate building in the university campus. The project is located in the second floor of the building equipped with 10 systems with necessary fittings and furniture. A webpage in the university website has been created to access the following information contents.

E-Journals
E-Books
E-Databases
E-Question Bank
E-Dissertation

The project team visited few other states in the country for observation and consequences of the proposed virtual library and based on the visit a detailed report has been submitted with Technical committee for the future progress of the project.

7. Future Plan

The proposed Virtual Library envisages:

1. To create separate website for easy access
2To impart training for the Librarians of all the affiliated institutions of this University to create awareness of the Virtual Library.
3. To conduct workshop for hands on exercise for the Librarians of the all the affiliated institutions of this University.
4. To organize seminar for updating knowledge on E-resources for different types of user community and professionals.
5. A series of meetings (Specialitywise) of the affiliated institutions of TANMMU has been planned towards consortia development.
6. The project on completion by 2006, shall be sustained, strengthened and steered for better information services.

Central Institute of Indian Languages library Goes On-line

Suman Kumari R and Sharada B.A.

Abstract

The Central Institute of Indian Languages library and its regional seven regional centers in India are unique in terms of information resources which are specialized in linguistics and Indian languages. Recently the CIIL library was formally inaugurated as the second digital library with a computerized system that allows its users to obtain coherent means of access to organized, electronically stored resources of information on their desktop. The paper attempts to brief the automation and the digitization stages of the CIIL library. For the library automation , an international software package i.e. VIRTUA management system was chosen developed M/S VTLS Onc. Virgina, USA which is certified by ISO 9001 with a main feature of UNICODE support to Indian languages .The library has just completed its retro conversion work. Also other small databases have been prepared for the thesis and dissertation, non-book materials, gratis publications and maps. The software has five menus for the house keeping operations. The package is being used for the day to day operations. Several Information Services has been offered to the staff on intranet. For the digitization of the resources, the available non-book materials are being digitized The CIIL has plans to slowly transform its digital library into a hybrid library with the objective to make make it as a national information centre for linguistics and Indian languages .

1. Introduction

Today we are living in an information era where information as a commodity is playing a central role in our daily life. We are amidst an information explosion and information technology revolution leading to the emergence of an electronic information era. Rapid advances in information processing, storage and communication technology have revolutionised the role of worldwide libraries in the dissemination information services to their users. (Gopal, 2000)

Emerging information technologies are allowing the libraries to transform the way of providing access to their resources and consequently, changing the way users find and utilize information. A digital library offers the prospect of access of electronic sources at their convenience temporally and spatially. (Shiuav, Su, 2002)

The decade of the nineties has brought a mature spectrum of automated library system software solutions to the libraries of all sizes and types. (Iyer, 1999). Librarians have experienced a series of changes over the past decade with the movement of information resources from the physical to the electronic resources. The user perception of the librarian today is not mere custodian but a friendly expert, resource provider, creator and distributor of information. On the other side, the role of the libraries are not just a mere storehouse of books but to act as promoters of information. Today in the digital era, a user desires a fruitful experience in a newly wired world to access information. (Tanner, 2001)

Central Institute of Indian Languages , Dept. Secondary and Higher Education, Ministry of HRD, Mysore – 570 006 ; ciil@sancharnet.in

The library initiative for digitization is to dramatically advance the means to collect, store and organize information in digital form, and make it available for searching, retrieval and processing via communication networks, all in a user-friendly way. The reasons behind this is the rising acquisition and the subscription rates which have forced libraries to seek other ways to make information available, and content aggregators and e-book publishers are providing the means. Perhaps more importantly, digital libraries support service improvement. Digitization presents opportunities for long-term preservation of information resources. Digital environment enables cross community interactivity and collaboration regardless of physical location (Schwartz, 2000).

2. Central Institute of Indian Languages: A Research Institution, www.ciil.org

The Central Institute of Indian Languages (CIIL), situated at Mysore was established on 17th of July 1969, as a national institution of advanced research in the area of linguistics and Indian languages. The aims and objectives of the institute is to coordinate the development of Indian languages, to bring about the essential unity of Indian languages through scientific studies, promote inter-disciplinary research, contribute towards emotional integration of the people of India.

During the past 35 years, CIIL has created a legend of its own by producing and archiving studies on 118 Indian languages, including 80 tribal languages and published 500 books, monographs, including language learning materials and created courses on cassettes in almost all the major Indian languages.

The institute has its seven regional centers imparting instruction in fifteen Indian languages. The main objective of these centers is to provide all the states and union territories of the country facilities to implement the three language formula in order to promote national integration through multilingual education.

3. CIIL Library: The Treasure House of Information Resources in Linguistics and Indian Languages

The CIIL library was established in 1970 to support the institute's objectives as a specialized research library in Indian languages, linguistics and its related areas. The motto of the library has been to provide information support to the staff of the institute in their research activities. It is viewed today as a premier advanced research centre for Indian languages. Its role is to collect, maintain and make available the information from print to electronic form, to its users.

Recently the library has launched its web site as http://www.ciil.org/aboutus/index.htm, which provides the total overview of the library.

The total collection of books at present is 80,000 volumes out of which 60% of the books are in Indian languages, which represent 60 Indian languages including tribal languages and 15 foreign languages. It is acquiring about 350 journals and 25000 back volumes covering all aspects of linguistics and languages studies. Besides this, the library has subscribed to electronic journals through "EBSCO Information Services Group, New Delhi, which provides access to its databases like "Academic Search Premier", "Professional Development" and "MLA Bibliography". This has enabled the users to

access the full text articles from scholarly journals in the area of Humanities, Social Sciences, and Education etc. The annual library budget is Rs.50 lakhs.

Among the resources collection, the library maintains unpublished theses and projects reports in Indian languages and linguistics: proceedings of seminars, conferences, workshops etc.

Besides, it has a few special collections of resource materials which include 1) Language text books from standard 1 to standard 10 in all Indian languages 2) College level text books published by 'grantha' academies 3) Children's literature in major Indian languages 4) Indian literature in translation 5) Adult education and literacy primers 6) Reprints, off prints and language acts.

Another important feature of the collection is the acquisition of the complete set of census reports and imperial gazetteers from the beginning from1872 in microfiche form and print form from 1961-2001.

The non-book materials on linguistics, languages and related areas include microfilms, microfiches, cassette tapes, filmstrips, photographs and charts on the development of Indian scripts and writing, maps which are for restricted and unrestricted use.

Recently, the library has completed digitization of a rare book on Telugu and made available as an e-book for its users electronically. This process will continue in future for the benefit of users for rare books with copyright permission.

The external users from India and abroad include students and research scholars from government and other institutions, Universities, etc.

4. Bhasha Bharati and Library Automation Project

Under the tenth plan of the CIIL, the Bhasha Bharati and Library Automation Project (BBLA) was initiated during the first week of August, 2003. The first and foremost task of the project was to take up the "Library Automation Work" of libraries of CIIL and its seven regional language centres. The project aim was how effectively and efficiently the information is made available to the users at their work place.

The project plan was to automate all the library house keeping operations such as

- Acquisition
  - Books and Serials
- Circulation
- Reference Service
- Library Management
  - Planning & Budgeting
- Stock Verification
- Networking and Resource Sharing of RLCs libraries

4.1 Goals

- To create a multilingual database compatible with international standards

- To promote resource sharing among the centre and its seven regional centres.
- To develop a repository platform to preserve the linguistics and Indian languages resources including tribal languages.
- To prepare a union catalogue
- To provide information service
- To prepare authority list of titles and Indic names.

5. Need For Digitization: Initial Steps

The CIIL library is a multilingual, multi-media specialized research library. Since its inception the library has acquired information resources from print to electronic form in the area of linguistics and its related areas, Indian languages including tribal languages. This valuable collection includes rare documents, theses and project reports, microfiche, microfilms, maps, cassettes, and special information resources in special areas. With this background set up a strong need was felt to preserve these valuable resources for posterity as well as to enable the users to access the resources in the digital form by adopting modern technology in the digital era.

Further the goals and the user needs were defined in the project. A survey was made for the selection of suitable software from the point of its support for the Indian languages. After much scrutiny an international software known as "Virtua Management System" developed by M/S VTLS (Visionary Technology for Library Solutions), Virginia was selected as a suitable one as it has UNICODE support to Indian languages and certified by ISO 9001

5.1 Installation of VTLS Software: Its Features

VTLS with its head quarters in Blacksbury, Virginia is a global company that creates and provides visionary technology in library solutions. The Virtua Integrated library system software version 43 on Solaris platform 9 was installed in the Sunfire E 450 server. The software has been designed on oracle.
The software is based on the following techniques:-

- Relational database management system
- Three -tier client /server architecture
- Built on z39.50 protocol
- UNICODE language support
- Z39.76 (data elements)
- ISO 10160/10161

5.2 Virtua ILS features

- Supports MARC system including UNIMARC, MARC21, and its authorities
- Full support of UNICODE Language
- Liable to customize the library operational requirement
- Incorporates Z39.50 protocol standard for information retrieval which enables the user to search the library catalogues remotely
- Provides powerful security management at all levels i.e. system manager, individual user, user group and patron group
- Works under graphic user interface

- Integrated authority control and cross reference structure

5.3 Library Retro-Conversion in Action

With the infrastructure set in at CIIL, the first phase of the automation work was executed on 19th of Jan 2004 at the CIIL centre library to be followed in the next phases for the regional centre libraries. The library retro conversion work was assigned on contract basis to the VTLS, Noida branch in India.

5.4 Retro-Conversion Work by the VTLS, Noida

The VTLS had a team of library professionals engaged in the retro conversion work. One set of team worked for the entry of worksheets, whereas the other team worked for the verification and validation of the worksheets. A team leader was appointed for the supervision of the work.

As a first step the CIIL library and the VTLS using the MARC tags designed a user need based worksheet. These worksheets were filled in with all the required bibliographic details by the VTLS team of staff. Physically each and every book was provided with a worksheet for further verification and validation.

Later these verified sheets were collated and forwarded to the VTLS office at Noida.

The VTLS team of library professionals, computer specialists and language specialists further scrutinized the information recorded for the final record. Once the entire validation process was over, the data was input in the virtua client system to prepare the CIIL database and directly transported to the Sun server.

The CIIL database was uploaded for testing during August 2004.

6. CIIL Digital Library Goes Online

On 5th of May 2005, the CIIL library as the second digital library in the country in its 37th year of service to Indian languages, converted its conventional multimedia library into modern partial digital library. This will allow the users to obtain coherent means of access to an organized, electronically stored repository of information resource materials at their desktops.

The retro-conversion work of the CIIL Library, and its regional language centers libraries at Mysore, Pune and Patiala have been successfully completed. The bibliographic records of these centers have been provided online in English along with the parallel information in the original languages. In its subsequent phases, the library plans to take up the retroconversion work of the other regional language centre libraries at Solan, Lucknow, Bhubaneswar and Guwhati shortly.

While building digital library aspects such as collection infrastructure, access infrastructure, computer and network infrastructure, digital resource organization and training programmes for staff and users were taken into consideration.

The multimedia database consists of all the books and their accompanying materials such as digitized gramophone records, audio tapes, video cassettes, film strips, etc, using 856 tag in the work sheet, current/back volumes of journals, theses and dissertations and their digitized version, gratis books and maps. For maps "High resolution" facility is provided in the software for viewing and navigating.

Virtua Software is being used for all the house keeping operation. The library has replaced its card catalogue with OPAC. Bar-code technology is being used for circulation transactions.

The information services provided on intranet are Current Awareness Services, New Acquisition Services, Contents of Periodicals in linguistics and Allied Services, Online News-papers clippings Services, Gratis books, Theses and dissertations etc.

The library has 25 terminals installed exclusively connected through LAN to the Sun Server. The user has a wide range search capability by author, title, subject, keywords, language etc along with the electronic database

After the completion of the retro-conversion task of its centers, the ambitious plan is to make it as a "National Information Centre for Indian Languages and Linguistics". The CIIL database which includes bibliographic records of its seven centers will serve as an Indian Languages and Linguistics union catalog which will be an important asset of information to the users in networking and resources sharing.

7. VTLS Software Support for Indian languages

- With its UNICODE support document titles are made available in both Roman transliterated form and Indian language script.
- The VTLS iPortal has easy to use facility, which requires no programming and training.
- It supports many languages including right to left languages such as Arabic,etc.
- At the data input stage provides languages code support such as ass- Assamese, ben - Bengali, etc.
- For Language display, the user can change the language on the screen at all levels including menu tool bars. There are about twenty languages in the combo box.
- While configuring notice- generating reports, under info station, Honor language and Honor Address Preference parameters appear in the Report Specific parameters. Honor language preference is based on the language preference specified in the patron record, 100 tag subfield 1 and language of the report configuration.
- MARC labels can be edited, translated into any language.

8. Conclusion

With the onset of the CIIL digital library, the future plan is -

- To act as a clearing house of information on all aspects of languages, linguistics and its related areas.
- Digitization of books in Indian languages
- To transform the CIIL Digital Library to a Hybrid Library.

References

1. Gopal, Krishan, Digital libraries in electronic Information Era, Delhi: Authorpress, 2000.
2. Iyer V.K., Library Information Technology: Tools for tomorrow, Delhi: Facet Publishing, 1999.
3. Schwartz, Candy, Digital Libraries: An overview, Journal of Academic Librarianship **26:6**:2000.
4. Shiuan Sm, Shery, Web-Based reference services; the user intermediary interaction perspective IN The digital factor in library and information services/edited by G.E.Gorman, London: Facet Publishing: 2002.
5. Tanner, Simon, Librarians in the digital age: Planning digitization projects. Program: **35:4**:2001.p327-337.

Digital Library: An Unavoidable Need in Today's World

Manisha R Amrelia, Murthy T.A.V. and Satyabati T.

Abstract

Over the past few years, libraries throughout the world have been in a state of transformation as a result of the impact of information technology. No area of the library has remained untouched. The impact on collections, services, staff, and facilities has had major ramifications on budgets, planning, and training. Thus the concept of Digital Library has come which focus on access and service not on buildings and volume. Libraries should support users in their searching and acquiring of information and their organization will reflect services rather than physical location. Technology, law and economics are all becoming more important for libraries requiring new expertise in library staff. Perhaps the most important issues for the long term will be the ability of libraries to co-operate in the delivery of the new services. This paper will highlight in brief the needs of the digital library in this age of Information technology. It will also give an overview of the requirements of a digital library.

Keywords: Digital Library, Digitization, Hardware, Software, preservation.

Introduction

A digital library is a highly organized collection of electronic resources. Digital libraries share an important characteristic with search engines - they can both be accessed online. However, while search engines cover a wide range of subject areas, digital libraries are more narrowly focused around one or a specific group of disciplines. Unlike search engines, digital libraries attach content-specific and highly descriptive metadata to describe each item in the collection. When a user conducts a search in the digital library it is this metadata that is searched. Search engines, on the other hand, search "blindly" on an item's content and the results obtained may only indicate that a particular search terms appears somewhere in the item, and not whether the overall content of the item is relevant to the search. Therefore, searches in a digital library produce more useful results, save users' time and effort in searching, and users can access the information found instantly. Digital library is a very complex and dynamic entity. It has brought phenomenal change in the information collection, preservation and dissemination scene of the world. There are many definitions and they are synonymously used as electronic library or virtual Library. They require technology to link the resources of many and the linkage between them and the information services are transparent to the end users. There collections are not limited to document surrogates they extend to digital artifacts that cannot be represented or distributed in printed formats. Another thing is that digital libraries will not be single, completely digital system that provides instant access to all information, for all sectors of society, from anywhere in the world. This is simply unrealistic. This concept comes from the early days when people were unaware of the complexities of building digital libraries. Instead, they will most likely be a collection of disparate resources and disparate systems, catering to specific communities and user groups, created for specific purposes. They also will include, perhaps indefinitely, paper-based collections.

INFLIBNET Centre(UGC) , Near Gujarat University Guest House, Ahmedabad
E-mail: manisha@inflibnet.ac.in

1. Characteristics of Digital Libraries

Digitization has benefits beyond improved accessibility. Institutions can protect originals from excessive handling and repeated copying; digitization can be a preservation strategy for the institutes.

The different characteristics of a digital library are as follows:

Collections: Digital library collections contain fixed, permanent documents. Not only those current libraries have more dynamic collections, but digital environment will enable of quick handling and/or ephemeral information.

Technology: Digital libraries are based on digital technologies. The underlying assumption is that the digital libraries will contain only digital materials, may be wrong. It is likely that both digital and non-digital information material will have to coexist.

Work: Digital libraries are to be used by individuals working alone. There is work oriented perspective focusing on group of information analysts, work being done and the documents and technologies that support it.

Transbordering of information: Breaking the physical boundaries of data transfers within and outside the countries. It is viewed that the support for communications and collaboration is as important as information seeking activities.

2. Digital Collection

Libraries began to create digital content; the impetus came from a number of areas. The desire are not to be left behind, opportunities presented by funding sources and/or faculty interest, need to develop local expertise, the desire to bring special collections to a broader community, or the hope of preserving the physical artifact. While all these interests are worthy, the long-term implications of the costs/benefits of creating and maintaining digital collections now must be seriously reviewed. The success of a digital library depends largely on the nature, content and quality of its digital collections. The basic requirement in creating a digital library will be the building of digital collections. The digital collections of information include various resources such as electronic journals, books, full text, CD-ROM databases, etc. The benefits of information collections in digital form for preservation, access and managing large quantities of information have been recognized by both library professionals as well as users.

3. Digital Preservation

The long-term maintenance and upgrade of digital files on digital storage medium is called digital preservation. Although technology is a key element in digital preservation, we believe it isn't the greatest inhibitor. The important issue is keeping digital information available in perpetuity. In the preservation of digital materials, the real issue is technical obsolescence. Technical obsolescence in the digital age is like the deterioration of paper in the paper age. Libraries in the pre-digital era had to worry about climate control and the de-acidification of books, but the preservation of digital information will mean constantly coming up with new technical solutions.

When considering digital materials, there are three types of "preservation" one can refer to:

- The storage medium.
- Access to content.
- Fixed-media materials through digital technology.

For these libraries jointly can coordinate a scheme and

- Create policies for long-term preservation.
- Ensure that redundant permanent copies are stored at designated institutions.
- Help to establish preservation standards to consistently store and share materials preserved digitally.

4. Digitization

Digitization refers to the conversion of an item in printed text, manuscript, image or sound, film and video recording from one format (usually print or analogue) into digital. The process basically involves taking a physical object and captured using a scanner or digital camera and converted to digital format that can be stored electronically and accessed via a computer.

One of the primary methods of digital collection building is digitization. Digitization is an electronic process of converging information from an analog format to the digital format. The Analog-to-Digital conversion means the transformation of continuous variable signal into the discrete variable signal. The process of digitization involves the scanning of the materials to be digitized. The scan images are collected in various standard formats like JPEG, MHEG, MPEG, HYT etc. depending upon the type of documents to be digitized. Optical character recognition technology is needed to transform the scanned image into hyper-text document.

The primary method of digital collection building is digitization. Digitization is also a high-speed data transmission technique. It is the conversion of any fixed or analog media (such as books, journals, articles, photos, painting, maps, microforms etc) into electronic forms through scanning, sampling or rekeying by using various technologies.

Digitization provides solutions to traditional library problems such as conservation, preservation, storage, space, multimedia documents, remote access to information collections, and acquisition of original digital works created by publishers, agencies and scholars, access to external materials not held in-house by providing pointers to websites, other library collections and publisher’s servers.

5. Hardware and Software for Digital Libraries

Hardware:

The minimum requirements of hardware for digital Library are:

- Computer
- CPU, PCI Bus, Ethernet, Silicon Graphics, RAM
- Storage Devices
- Hard Drives, Removable hard drive, Optical drive, DAT drive (Digital Audio Tape), CD ROM drive
- Monitors
- Digitizing devices
- Scanners, Digital Camera
- Out Put Device
- Printers Modem and CD Writer

Software:

There are number of software, which can be used for different function some of them are as follows:

- Editing images
- Page layout programmers: to integrate text and graphics
- Page transferring utilities : to share files between computer platforms
- File translation programmes: to convert files from graphics
- File compression software.

The software which are used mainly for digital library are Greenstone, Dienst Eprints Archive software, Dspace etc.

6. Future of digital Library

As we venture into a more digital environment, many of the traditional measures of an excellent library have become eroded. We know that we have been successful and what benchmarks might we use to compare ourselves with peer institutions and against ourselves. In the digital world of information highway, there should be stress on three things: awareness of information, awareness of technology, awareness of needs. The awareness of information gives the breadth of vision; awareness of technology gives the power to make the visions manifest; and awareness of needs gives the insight to use professional skills and talents to the greater effect.

7. Conclusion

The information and communication technology has changed the complexion of today's libraries on a large scale and we are amidst in information explosion and information technology revolution leading to the emergence of electronic information era. Rapid advances in information processing, storage and communication technologies have revolutionized the role of worldwide libraries in disseminating information services to their users. As a result, libraries are facing new challenges, new competitors, new demands, new expectations and variety of information services from users tailored to their wants and needs. Libraries around the world have been working on this daunting set of challenges for several years now. They have created many digital library initiatives and projects, and have formed various national schemes for jointly exploring key issues. With several years' accumulated experience, the initial enthusiasm surrounding the development of the digital library has been replaced by sober second thought. Librarians

have discovered that, with a few exceptions, making a business case for digitization and investments in digital technology is more difficult than first envisioned, especially given the technical and legal constraints that must first be overcome. As with most other technical developments in libraries over the years, we will have to move forward in small, manageable, evolutionary steps, rather than in a rapid revolutionary manner.

References

Gopal Krishan: Digital Libraries in Electronic Information Era. Delhi. Authors press. 2000 (http://www2.edc.org/ GDI/effectiveaccess.htm)

Stern David: Digital Libraries:Philosophies, Technical Design Considerations, and Example Scenarios. New York. The Haworth Press, Inc.1999

Sasikala, C … [et al.] Delivery of information services through distributed digital environment, SIS 2005.Visakhapatnam,27-29 January 2005.

Singh, Yogendra … [et al.] New Challenges in Information Management and E-learning in the age of Globalisation: Issues and Oppurtunities; 21st Annual Convention and Conference , SIS 2003. Roorkee, 9-11 April 2003.

Murty, T A V … [et al.] Multilingual Computing and information Management in Networked Digital Environment, CALIBER 2005. Kochi. 2-4 February 2005.

Murty, T A V … [et al.] Content Creation, Access and Management in Networked Environment, PLANNER-2004. Imphal. 4-5 November 2004.

Digital Library Collection Development and Management Policies

Rajendran. L[1], Venkatesan.M[1] and Kanthimathi.S[2]

Abstract

The Processes of Collection Development and Collection management are undergoing a transformation as well. This is due to the effect of great many and diverse digital resources and tools that can be used in collection development and that are generally and easily available through the Internet. The Digital Library Collection Policy is a standard library practice for publicly declaring a library's intent for breadth and depth of the material; it will collect within certain subject areas, genres, or physical formats. Such declarations are useful tools that scholars can use to determine the relative utility of a collection for their purposes, as well as to assist in cooperative collection development with other libraries.

Key Words: Digital Collection policy, Challenges to Digital Collection, Principles for Digital Collection.

1. Introduction

Libraries are about many things. But, Collections have always been at the heart of libraries, be they digital, traditional brick and mortar, or hybrid between the two. Moreover, collections will retain that role in the future as well. However, the concept of what constitutes a collection in the networked environment of digital libraries is undergoing a transformation from the age-old concept of library collection signified by ownership. A new concept of a digital collection is evolving incorporating adaptations of many old features and standards, and creation of many brand new ones. This conceptual and pragmatic evolution is far from over. The concepts and processes of collection development and collection management are undergoing a transformation as well. This is due to the effect of great many and diverse digital resources and tools that can be used in collection development and that are generally and easily available through the Internet. New processes and tools for collection development has emerged, used for development and management of both, traditional and digital collections. In addition, the process of collection management became more closely connected than ever with means, ways, and policies for access, adding an additional dimension.

2. Challenges

Digital Library faces some challenges at collection and development processes as follows:

- Technical Architecture
- Copy right Management
- Preservation

1. Tagore Engineering College, Chennai-600 048. (rajendranlak@yahoo.com)
2. Rani Anna Govt. College for Women, Tirunelveli

2.1 Technical Architecture

The Technical Architecture that underlies any digital library system. Libraries will need to enhance and upgrade current technical architectures to accommodate digital materials. The architecture will include components such as:

- High-Speed local networks and fast connections to the Internet
- Relational databases that support a variety of digital formats
- Full text search engines to index and provide access to resources
- A variety of servers, such as Web servers and FTP servers
- Electronic document management functions that will aid in the overall management of digital resources

One important thing to point out about technical architectures for digital libraries is that they won't monolithic systems like the turnkey; single box OPAC's with which librarians are most familiar. Instead, they will be a collection of disparate systems and resources connected through a network, and integrated within one interface, most likely a Web interface or one of its descendants. For example, the resources supported by the architecture could include:

- Bibliographic databases that point to both paper and digital materials
- Indexes and finding tools
- Collections of pointers to Internet resources
- Directories
- Primary materials in various digital formats
- Photographs
- Numerical data sets
- Electronic Journals

Though theses resource may reside on different systems and in different databases, they would appear as though there were one single system to the users of a particular community. Within a coordinated digital library scheme, some common standards will be needed to allow digital libraries to interoperate and share resources. The problem, however, is that across multiple digital libraries, there is a wide diversity of different data structures, search engines, interfaces, controlled vocabularies, document formats, and so on. Because of this diversity, federating all digital libraries nationally or internationally would an impossible effort.

2.2 Copyright Management

Copyright has been called the "single most vexing barrier to digital library development". The current paper-based concept of copyright breaks down in the digital environment because the control of copies is lost. Digital objects are less fixed, easily copied, and remotely accessible by multiple users simultaneously. The problem for libraries is that, unlike private businesses or publishers that own their information, libraries are, for the most part, simply caretakers of information they don't own the copyright of the material they hold. It is unlikely that libraries will ever be able to freely digitize and provide access to the copyrighted materials in their collections. Instead, they

will have to develop mechanisms for managing copyright, mechanisms that allow them to provide information without violating copyright, called rights management.

Some rights management functions could include, for example:

- Usage tracking
- Identifying and authenticating users
- Providing the copyright status of each digital object, and the restrictions on its use or the fees associated with it
- Handling transactions with users by allowing only so many copies to be accessed, or by charging them for a copy, or by passing the request on to a publisher

2.3 Preservation

Keeping digital information available in perpetuity. In the preservation of digital materials, the real issue is technical obsolescence. Technical obsolescence in the digital age is like the deterioration of paper in the paper age. Libraries in the pre-digital era had to worry about climate control and the de-acidification of books, but the preservation of digital information will mean constantly coming up with new technical solutions. When considering digital materials, there are three types of preservation are:

- The Preservation of the storage medium
- The Preservation of access to content
- The preservation of fixed-media materials through digital technology

2.4 Preservation of the Storage medium

Tapes, hard drives, and floppy discs have a very short life span when considered in terms of obsolescence. The data on them can be refreshed; keeping the bits valid, but refreshing is only effective as long as the media are still current. The media used to store digital materials become obsolete in anywhere from two to five years before they are replaced by better technology. Over the long term, materials stored on older media could be lost because their will no longer have the hardware or software to read them. Thus, libraries will have to keep moving digital information from storage medium to storage medium.

2.5 Preservation of access to content

This form of preservation involves preserving access to the content of documents, regardless of their format. While files can be moved from on physical storage medium to another, what happens when the formats (e.g., PDF, Adobe Acrobat) containing the information become obsolete? This is a problem perhaps bigger than that of obsolete storage technologies. One solution is to do data migration that is, translate data from one format to another preserving the ability of users to retrieve and display the information content. However, there are difficulties here too-data migration is costly, there are as yet no standards for data migration, and distortion or information loss is inevitably introduced every time data is migrated from format to format.

2.6 Preservation of fixed-media materials through digital technology

This slant on the issue involves the use of digital technology as a replacement for current preservation media, such as microforms. Again, there are, as yet, no common standards for the use of digital media as a preservation medium and it is unclear whether digital media are as yet up to the task of long-term preservation. Digital preservation standards will be required to consistently store and share materials preserved digitally.

3. Collection development principles

General principles related to quality are defined and discussed, and supporting resources providing further information are identified. These resources may be standards, guidelines, best practices, explanations, discussions, clearinghouses, case studies, or examples. Every effort has been made to select resources that are useful, current, and widely accepted as authoritative. However, the list is not exhaustive and, given the dynamic nature of digital information, can be expected to change over time. The resources listed will in some cases serve as a starting point to lead the reader to additional resources. This framework is intended to be flexible enough to accommodate new principles, considerations, and resources, and to absorb the contributions of others. At the same time, it is intended to be a concise introduction to core considerations for the building of good digital collections and to serve as a springboard to encourage further research and innovation by its readers.

- Collection should be described so that a user can discover characteristics of the collection, including scope, format, and restrictions on access, ownership, and any information significant for determining the collection's authenticity, integrity and interpretation.
- A collection should be sustainable over time. In particular, digital collections built with special internal or external funding should have a plan for their continued usability beyond the funded period.
- A good collection is broadly available and avoids unnecessary impediments to use. Collections should be accessible to persons with disabilities, and usable effectively in conjunction with adaptive technologies.
- A good collection respects intellectual property rights. Collection managers should maintain a consistent record of right shoulders and permissions granted for all applicable materials.
- A good collection has mechanisms to supply usage data and other data that allows standardized measures of usefulness to be recorded.
- A good collection fits into the larger context of significant related national and international digital library initiatives.

4. Conclusion

Collection management and organization research is the area where traditional library missions and practices are reinterpreted for the digital library environment. Progress in this area is essential if digital library collections are to meet successfully the needs of their user communities. Policies and methods for incorporating information resources on the network into managed collections, rights management, payment, and control issues were all identified as central problems in the management of digital collections. Approaches to replication and caching of information and their relationship to

collection management in a distributed environment need careful examination. The authority and quality of content in digital libraries is of central concern to the user community; ensuring and identifying these attributes of content calls for research that spans both technical and organizational issues. Research is also needed to clarify the roles of librarians and institutions in defining and managing collections in the networked environment.

References

Chepesuik, R. (1997). The future is here: America's libraries go digital. American Libraries, vol. 2, No.1, pp. 47-49.

Shreeves, E. (1997). Is here a future for cooperative collection development in the digital age? Library Trends, vol. 4 , No.3, pp.373-390.

Waters, D.J. (1998). What are digital libraries? CLIR, July/August Issues. URL: http;//www.clir.org/pubs/issues/issues04. HTML

James R. Davis and Judith L.Klavans. (1997). "Workshop Report: The Technology of Terms and Conditions. D-Lib Magazine ,June.

Arms, W.Y. (1995). Key concepts in the architecture of the digital library. D-lib Magazine, July.

Erway, R.L. (1996). Digital Initiatives of the Research Libraries Group. D-Lib Magazinde. URL: http://www.dlib.org/dlib/december96/rlg/12erway.html.

Lynch, C.A. (1998). Identifiers and their role in networked information applications. Felciter, January, pp.31-35.

Digital Collection Development

Florence Chandran

Abstract

The transformation from print media to electronic media has a tremendous impact on the way the information is owned, shared and accessed. The commercialization of digital information has long-term implications for the acquisition and development of library collections. The purchases are replaced by licensing and business practices of software companies replace those of publishers. Access to information on demand supersedes collection building and cooperative acquisitions supplement local collection building.

Growing demand for full text online content that can be easily searched and remotely accessed has led libraries to depend on a host of intermediary agents and cooperation. There is proliferation of information and diminishing buying power. The greatest challenge faced by research libraries is the digital collection development.

In this paper, efforts have been taken to identify how libraries have responded to this paradigm shift by pioneering new collection development strategies and they examine the changing responsibilities of collection development libraries in an electronic environment.

Electronic Acquisitions and Collection Development

Collection development policies are framed based upon an understanding of strengths and weaknesses of the collection, the availability of shared resources and the information needs of the community. Policies must consider the virtual library from a dual perspective. It is both a dynamic collection in its own rights and a hybrid collection created by merging the virtual and physical libraries. The goals of collection development in any library are to meet the immediate and anticipated information needs of users. This is accomplished through strategically selecting sharing, retaining, duplicating, archiving and facilitating access to intellectual content.

Content

The foundation of the virtual library is intellectual content. This includes indexing, abstracting and full-text databases; electronic journals and books; resources in multimedia formats, numerical and geospatial data; digitized special collections and free internet sites. Resources that contain full text articles generally selected by the vendor form a variety of sources are often referred to as aggregator databases. Some databases are multi-disciplinary, while others offer integrated access to multiple resources by discipline.

The budget allocations have to be made for electronic resources. Initially additional funding my be available for the initial acquisition of electronic materials, eventually electronic resources will take up major percentage in allocation.

Librarian, Women's Christian College, Chennai

Electronic Collections Collection Team

The complexion of evaluating and comparing electronic resources, especially in large libraries or networks makes it difficult for one person to select materials. Hence an electronic collections selection team may be able to do a better job. Members of this team should possess both functional and subject area expertise. Having representatives from technical services and systems will be especially advantageous when comparing similar resources. An individual must be responsible for negotiating price and modifying signing the license agreement. An additional person must be identified to interact with vendor's technical staff and be responsible for mounting new resources on the virtual library.

Before beginning the identification and selection processes the team should select background information and formulate preliminary policies and procedures. It is both expedient and wise to adopt polices and standards established by national and international organizations.

Budget

The selection team should begin with a clear idea of the funds available for electronic collections. To take advantage of the reduced pricing structures available through consortia, the team should identify which professional systems are affiliated with the library. These include national and regional library networks; city, country and state educational systems, special interest groups for special libraries, professional associations and statewide technology initiatives. Finally the selection team should contain basic information on the hardware and software need to access the resources and existing technology within the library. The cost of new technology must be considered in the development of the budget for electronic resources.

Establishing the Types of Resources Needed

The selection team needs to determine the types of resources to be collected. A frequent issue involves the replacement of print indexing and abstracting materials with online versions. A second issue considers the licensing of aggregated full text databases or journals that are only available online. The formats such as numerical and multimedia are also to be determined. Libraries require an evaluation of resource even if it is available (without cost) in the internet. However there is a cost associated with the time and energy used by professional staff to evaluate, catalogue and maintain these "free" resources.

Establishing Workflow

After establishing a selection team, a budget, preliminary policies and the types of resources to be considered, it is time to create the actual workflow process. The team must determine what criteria will be determined for evaluation. Evaluation of resources can be done making use of the free-trial offers. Published reviews of electronic resources have become regular features in both scholarly and popular prices. Using trial access, published reviews and vendor documentation the evaluation should be completed within a given time frame.

Critical Issues in Evaluating Electronic Resources

There are a number of critical issues to be considered including content, access, timeliness, cataloging, and sustainability, usability, usage assessment and statistics, technical performance and service levels, pricing structure and licensing term.

Content

The quality of digital content like print content is judged by a number of factors including the authority of the resource, comprehensiveness, completeness, currency, accuracy, clarity, uniqueness and conformity to academic standards and conventions. If the product has a print, it is important to determine whether the electronic version contains all of the content that is available in the print version. If graphics are included in the electronic version, major concerns focus on the image clarity and consistency. In the databases care should be taken to check whether access to all the titles and articles are possible. JAKE (Jointly Administrated Knowledge Environment) is a freeware metadata management system and online database used to find, link and compare journal titles and union lists (http://jake.med.yale.edu) if is particularly useful for identifying duplicate full-text holding.

Sustainability

Sustainability requires that the cost of acquiring and maintaining a resource reflect lasting value and contributes to the integrity of the collection. Availability of the archiving service needs to be checked. The policy on retaining electronic back files of online data should also be specified in the evaluation.

Usability

During product trial it is possible to look at the usability of the resource. Usability includes ease of use, graphic design features and navigability. Since patrons' access to resources depends upon hardware, software and network connections, resources must be evaluated using a mix of operating systems and browsers. Both basic and advanced search should be available and easily located.

Usage Assessment Statistics

Usage statistics theoretically offer a quantitative method for evaluating the use of electronic resources. In addition, usage statistics are used in basic cost-benefit analyses to determined cost per use of a resource and to justify its expenses.

Access

In any well based resource, access is a critical component. Two major issues surrounding access include copying restriction (copying, lending or electronic reserve) and authentication of institutionally affiliated computers network and remote users. It is also important to consider whether the users are able to print download, e-mail the content as well as cut and paste from the resource.

Timeliness

The electronic resources have to be received on time. The electronic journals also should be checked whether they are required on time.

Guiding catalogues

When considering electronic materials the availability and quality of cataloging records are important. Descriptive information about electronic resources should include aquatic item-level descriptive metadata. For examples, licensing information lacking metadata is like purchasing a book without a title page, table of contents or index. Cataloging resources also enhances its visibility. Users dislike having access multiple gateways to find a pertinent resource. Mainstreaming electronic resources into the online catalog is best, avoiding separate gateways whenever possible.

Technical Performance and Service I Levels

According to the 1999 International Coalition of Library consortia guidelines on technical performance, vendors should provide information about performance levels, including response time, server downtime and disconnection. Product trial can be a valuable tool for identifying technical problems.

Features that Add Value

There are a number of valuable interoperable features found in electronic resources that are not possible in printed versions. Chief among these is full -text searching and linking which directly connects the text or images of one document or resource to the text or images of another document or resource. For example, some interfaces offer advanced search features that allow users to store and combine searches, map search terms to thesauri and manipulate search results by limitations. Others organize and display search results in particularly useful ways or allow the user to customize the display. However, the sheer number of features is often problematic for users who must adapt to different operators, search terms and screen displays.

Internet sites with substantive content increasingly offer a variety of added value services to an identified community of users. These services may include current awareness alerts (via e-mail), continuous revisions, topical online forums, e-mail lists and options for creating personal profile online.

Pricing Structure

Corporations create content, subject headings and pricing structures dynamically while struggle to define content and price. Unlike printed materials, which have a set cost with a possible discount, electronic resources are regularly priced in a flexible manner. The cost of adding online access to a print subscription varies considerably from a nominal charge to more than double the cost of the subscription.

Conclusion

Considerable attention has been given to the role of electronic resources in library collections. Less attention has been paid to the widespread reallocation of library funds to acquire and maintain electronic resources at the expenses of all other library materials. Hence enough resources should be allocated for building up the electronic resources. But the selection of the resources should be made by the team having a thorough knowledge of the library mission and collection parameters of the library. This helps in creating meaningful criteria for the guidance and development of research quality academic collections.

Future Trends

Theories on the future of libraries and electronic publishing are many. In some cases technology moves so quickly that future trends rapidly become past practices. Three consistent trends are the emphasis on managed information, increased collaboration with vendors and the creation of intellectual content.

Managed information, like managed health care attempts to contain costs and improve outcomes through a combination of approaches that focus on integrated, networked systems and services. These include cooperative collection development, on-demand publishing and purchasing by the article. It is important to have shared collection based on document delivering expanded access to virtual resources and the elimination of duplication.

References

Evans G.E (2000) Developing Library and Information Center Collections Englewood Colo: Libraries Limited

Lakos A.R. Gray, (2000) Personalized Library Portals as an organizational culture change agent. Information Technology and Libraries 19 (4) : 169-174

Manoff .M (2000) Hybridity, Mutability, Multiplicity: a Theorizing electronic Library Collections. Library Trends, 48 (h) 857-876

Martin M.S (1998): Collection Development and Finance A Guide to strategic library-materials budgeting, Chicago: American Library Association

Thorton, G.A (2000) Impact of electronic resources on collection development, the roles of librarians and library consortia. Library trends 48(4) 842-856

Digital library Infrastructure and Architecture

Prasad B.S[1] and Swarnalatha[2]

Abstract

Information technology has changed the modern day libraries as compared to traditional libraries. The conventional set up of brick, stone and mortar libraries that store information within a constrained physical space (Books/print materials on the iron racks/wooden racks) have given way to digital multimedia information store houses that integrate data resources around the globe through the effective deployment of Information technology, without straining the financial resources. Today it is obvious that the most effective way this is to create digital libraries, distributed information systems ensuring reliable storage and effective use of various collections of electronic documents (text, graphics, video, audio etc) via global telecommunication networks in a way convenient to the end users. The information explosion has direct impact on the libraries as they have to devise ways of performing their tasks better and faster even when the volumes are increasing at a high pace. Terms such as electronic library and virtual library are often used synonymously. The emergence of Internet and wide availability of affordable computing equipment have created tremendous interest in the Digital library and electronic publication concept.

1. Introduction

Traditionally, libraries have been collecting various kinds of sources and holding them in readiness for use by users. After World War II, there has been a knowledge explosion and consequent exponential growth of literature and information. There have been complexities in the information generation, handling and use. Libraries in the developed and developing countries witnessed the introduction of computers and IT from the 1960's. The growing impact of ICT (Information and communication technology), web technology and database technology has compelled libraries to use these technologies effectively to provide services to users. With growing number of e-sources, it has become imperative for library and information professionals to properly play their roles in disseminating information to their users. Information can be saved digitally and therefore this helps in immediate access to high demand and frequently by users. Digital library provides access to digital information collections, and includes a combination of structured/unstructured, text/numeric/graphical data, scanned images, and graphics, audio and video recordings. Digital library is now-a-days the most widely accepted term and implemented in all libraries

Librarians, better known as library managers, are required to keep-up and satisfy the demand of the faculty, students and researchers against a diminishing budget.

2. Definition: Digital Library

What is Digitization? It is the process of conversion of an analog signal/code into digital signal/code i.e., the analog information is captured into digital form. Here technology is readily available in the form of cameras, DTP, other capturing devices like scanners that allow one to convert analog images into digital images.

1 New Horizon College of Engineering, Bangalore-560087, Karnataka: Email-prasadbld@yahoo.com
2 St. Joseph PU College, Richmond Road, Bangalore: E-mail: swarna7768@yahoo.com

Digital library stores materials in electronic format and manipulates and disseminates the large collection of information effectively. Various documents saved in digital format that can be accessed through the intranet or the World Wide Web.

The term digital library covers the creation and distribution of all types of information over networks ranging from converted materials to kinds of information that have been generated in the physical world.

Digital library consists of information in digitized form, where the information stored are electronically accessible using a computer which can store, provide access and disseminate without any conversion process. Information in digital form is stored electronically and accessed, where access to digital library has no boundary or particular restrictions in access with respect to space/age/time. This digital library helps to provide access to information through electronic gateways to remote digital database.

Digital libraries are well accessed with establishing the wide range of Internet accessible sources of information. To the user interest of information needs are accessible and with greater speed, 100% accuracy and reliability. This helps to access the 3-dimensions graphics, created by computer imaging, which leads to virtual library.

Most logical definition from the view point of librarians, which was proposed by American Digital Library Federation, 1998, which says "digital libraries are organizations that provide the resources, including the specialized staff, to select, structure; offer intellectual access to; interpret, distribute, preserve the integrity of; and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of computers."

Based on the above definitions, Cleveland (1998) gave some of its characteristics. One of these characteristics is "Digital libraries are the digital face of traditional libraries and include both electronic (digital) as well as print and other (i.e., film, sound) materials."

Cleveland also says "In reality, digital libraries will not be a single, complete digital system that allows users to promptly access all information, for all disciplines, from any where around the world. Instead, they will most likely to be a collection of disparate resources and disparate systems, catering to specific communities and user groups, created for specific purposes. They will also perhaps include paper-based collections."

Sharma and Vishwanathan (2001) say that "Growth of digital libraries involves digitization of existing library materials; connectivity to the users in the world online and offline; integration with networking; and availability on the World Wide Web."

2.1 Digital library – Goals:

- To increase the access i.e., all types of information availability to users - both offline and online
- To preserve the original documents and manuscripts
- Qualitative and quantitative resource sharing
- To improve the library services
- Ensuring the effective usage of information storage in digital format
- Effective utilization of funds invested on digital library

- User acceptability from their desktops
- To satisfy the five laws of library science

3.0 Historical development of Digital libraries

From the conventional role as static storehouse of information, the library has matured into a proactive model of information generation and effective dissemination. The revolution in computer, information and telecommunication technology is bringing about significant changes in all types of libraries. Availability of powerful computers at affordable cost, spread of telecommunications networks to even remote areas, advent of internet, increasing interest in creating digital content are some of the significant forces accelerating the pace of changes in functioning of our libraries.

Licklider (1965) refer to a digital library "Library of the future" referring to a fully computerized library. Later, Lancaster, F W (1978) termed digital library as "Paperless Society". Presently many synonymous terms like "Electronic Library"; "Virtual library", "Library without walls", "Paperless library" are used.

The three main characteristics of digital library are the storage of information in digital form, usage of communication networks to access and obtain information and copying by either downloading or online/offline printing.

3.1 Developing digital libraries

Very few people will have all the skills required to set up a digital library. Library professional alone cannot develop the digital library. The total technology in developing digital library is too specialized for the librarians or any other layman. In conclusion, the digital library development projects are very much a team effort. Group co-operation and co-ordination can lead to fruitful results in developing the digital libraries. The important point that should be borne in mind before initiating the project of setting up of a digital library is that the main characteristic of a digital library i.e., Information is selected on the basis of quality; should be accessible to everyone and there is no restriction for definite user groups. Information stored in a digital library can be changed when needed and after obtaining the necessary permissions.

The five laws of library science coined by Dr. S R Ranganathan should be implemented in designing of digital library.

3.1.1 First Law: Books/Information are for use

Digital library should be designed in such a manner that it is easy for use, with a web-based user interface that can be customized for the institution, individual or department etc. Digital library comprises the information in digital format that will be easier to access only when necessary technology links, well-built infrastructure and computer systems are available to users. The users must be given orientation/training for searching and retrieval of information in the vast digital library. Digital libraries are expected to play a major role in formal learning as well as by providing the teacher and students with more information in a variety of media. The acquired/stored information in the database of digital library must be qualitative rather than quantitative. The information available in

digital library must be readily available for the user. The stored information must not been soon outdated. Well-maintained networks should be established.

3.1.2 Second Law: "Every Reader his or her Book/Information"

Any user who steps in for information in the digital library, the information acquired, gathered, stored, retrieved or disseminated in the action of digital library must be so relevant to the user, so that the information can find its own user in a short period of time. There should not be any obstacles in the search of information for the user. The digital library can be constructed in such a manner, that it substitutes for the librarian or online/offline databases. A number of open source software available for the implementation of a digital library, which can be reorganized, developed according to our needs with latest developments of Information technology.

3.1.3 Third Law: "Every book/information its Reader"

The user can obtain the required information or data from the database. For the same, the information must be so qualitative and the search path or retrieval path must be designed in such a way that the information can attract the user towards it. Internet is also termed as metadata (data about data or information about information). Over the years, metadata formats have been developed for a wide range of digital objects. Within the range of formats, there is a degree of consistency across all metadata schemes that support inter-operability i.e., most schemes provide for the title field, data field, and identifier field. Metadata creation must benefit the user.

3.1.4 Fourth Law: "Save the time of the user"

The searching tool or path must be well designed, the indexing part in the database must be regularly updated, qualitative index pattern must be used for framing keywords, search techniques must be easy, in such a way that the layman can easy retrieve information. User friendly software can be installed in retrieving the information. Generally search operations provide a large amount of information to the user, which is both related and unrelated to the topic of search. For narrowing the search, he can use the Boolean logic for reduction in search time. The search engines can be framed based upon the key terms entered into the digital conversion library.

3.1.5 Fifth Law: "Library is a growing organism"

The digital library should be designed with an eye for the future. The database storage capacity should be high. A system with good hardware capabilities should be installed for storage of information as Main Server. As 5^{th} law of library science says, that in formal libraries, the books/print media are the main acquisition products. But in a digital library, there are a large number of e-documents, which will be always increasing. Books can be kept in the library for reference, where as the electronic resources are stored in a server, which can be made centralized or distributed over WWW or local intranet.

Information and communication technology has given way to develop a new technology in the field of library science to transform the traditional library into the digital library. The other skills needed to design the digital library by the team (Librarians and IT Professionals)

- Technical skills (Knowledge of IT, Hardware, software etc.,)
- Project management
- Databasebase creations and development
- Computer programming
- Web designing
- Cataloguing/classification/indexing
- Preservation
- Graphic design/digitalization technique and skills

Other than these

- Elicit suggestions from the end users/subject experts/digital librarians
- Good co-operation and co-ordination among the team of library , IT professionals and users
- Computer systems for Providing access to users
- Various databases
- Servers

The parent organization must provide adequate finance and suitable working environment..

The Team : The Team should consist of the following experts.

- Subject
- Management
- Conservation
- Digital and film photographers
- Cataloguers
- IT specialists
- Administration

4. Digital Library – Infrastructure

The implementation of digital library can be achieved after discussing the framework between the librarians, information professionals, IT professionals, hardware/software/network professionals, management professionals, with the help of non-technical professionals. The main steps in the creation of digital library is as follows:

Digital library – Ideology, manuscript creation, literature search,
Digital library – Creation
Digital library – Acquisition
Digital library – Cataloguing, indexing
Digital library – Preservation
Digital library – Access/dissemination

4.1 Digital library – Ideology, Data collection, manuscript creation, literature search

To begin the idea of digital library, there must be some data collected related to it. These data may help us to get an ideology, problems faced by the other libraries suggestions and directions of professionals etc., Also, information about new developments are to be collected.

4.2 Digital library – Creation

Building the digital libraries begins with creating digital content and collections. There are many open source softwares for the creation of digital library like D-space, Green stone etc., which along with free supply of software, also provide the guidelines, basic requirements, procedure etc. Digital creation is the act of producing the information product in digitized format. The creator may be a human or a piece of equipments. There must be some specification for the purpose of standardization in terms of literature. Here in this step, there must be good coordination between IT professionals and Librarians. The digital library infrastructure can be built by IT professionals, as per the requirement of the library.

4.3 Digital library – Acquisition

The acquisition of digital library products is divided into different categories, i.e., print and non-print version, utility for present and future. Many commercial agencies have emerged for providing commercial electronic resources through providing IP address with user name and password. At the same time, based on usage, the print version can be continued.

4.4 Digital library – Identification, Cataloguing and Indexing

As books are acquired by the library, the accession number is assigned after all the normal actions like reviewing, ordering and purchasing, It is then sent to classification and cataloguing for identification purpose. In the same manner, for a document that has been acquired in digital format, it is necessary to assign the identification number and to catalogue it. This helps the digital library to manage the digital/electronic resources smoothly and for easy accessibility. Identification provides a unique key for finding the object and linking the object to other related objects; cataloguing helps in the organization and access.

4.5 Digital library – Preservation

This is the aspect of archival management that preserves the digital content of a digital library. Here preservation of digital content is mostly based on time frame and also depends upon the hardware and software.

These days a new release of database or word processor can be expected every few months for few years by making it better by updating. Software vendors also provide backward compatibility upto 2 to 3 versions. So these problems must be kept in mind while dealing with computer hardware, software and peripherals. For journal articles, the majority of projects reviewed use image files, PDF or HTML. For purely electronic documents, PDF

is most prevalent format. In PDF format, if the document is put in internet, it will consume more bandwidth as compared to HTML font, however the HTML format, the tables and pictures are not possible.

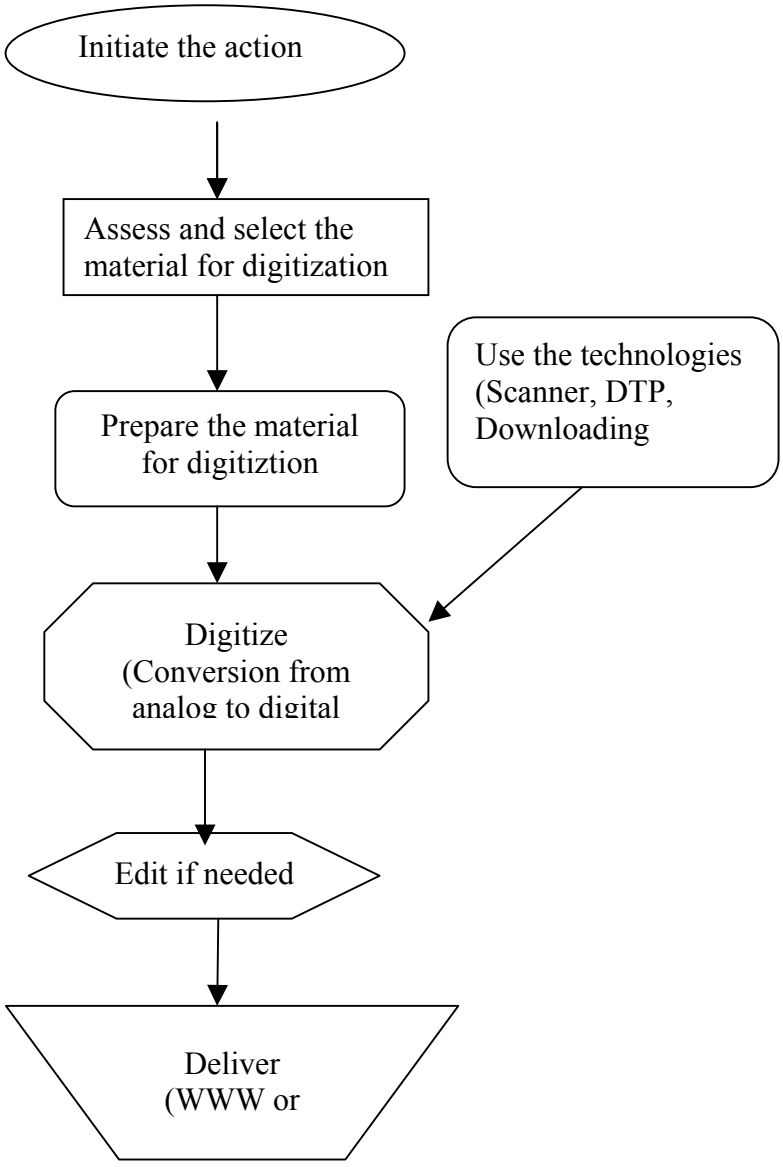
4.6 Digital library – Access

The information stored in a digital library server i.e., document or e-information can be accessed through search or retrieving software. As in a physical library, we use library catalogue for efficient search and to retrieve the books. The software involved in digital library helps to retrieve the metadata after analyzing the contents stored in the server. Appropriate Boolean logic needs to be used to narrow the search from a larger set.

5.0 Life cycle for digitization project

The life cycle of a typical project consists of the following parts :

Initiation of the project is shown in the following chart.



To initiate the project of digitization, following points are essential:

- Are there sufficient funds?
- Have copyright and any other rights related issues been secured?
- Can we protect the authorship i.e., IPR?
- Does the institution have enough expertise?
- Is there a partnership with a commercial provider?
- Do the results of digitization justify the costs?
- Are the necessary equipment available and also easy to use?
- Are any necessary precautions taken for preservation?
- Are steps taken for frequent updating of data?
- What services can be added in digital library to offer to the users?
- How can we establish and control the accuracy of information sources i.e., quality and quantity?
- Authority control problems
- Organizational problem etc.,

6. Conclusion

Digital libraries are going to play a very vital role in this century and are an important component for disseminating the required information. The context of physical sources of information are being transformed into digital for its various advantages which can lead to multiple usage and easy access. Traditional libraries are undergoing rapid changes to meet the targets with demands of fast changing information requests from its users. The advent of computers, IT, advancement in telecommunication and storage devices and disseminating techniques have given new dimensions to collecting, organizing and disseminating the vast amount of information.

NextGen Digital Resource Centre: Online Information Gateway

Manoj Singh, Vijai Kumar, Sanu Mathew and Leena Kanal

Abstract

Online solution with high degree of accessibility, scalability, manageability and security is a better approach to access networked information resources in a digital library. This paper presents a case study of online resources made available through an indigenously developed online information gateway at BARC Library. Technical aspects including merits & demerits are also discussed in detail. Some available technologies and in-the-box-applications for online information models are also discussed.

1. Introduction

Information is highly valuable resource that is being created, produced on a mass scale and disseminated to end-user using various types of methods and technologies. With the introduction of portable storage media right from the floppy to CD-ROM and to the state-of-the-art technology of DVD, the development of information publishing and its access has changed multi-fold.

The World Wide Web has played a major role to change the attitude of the so-called information society and continues shading its IT color to rest of the word. The fast changing technologies in IT also raises the expectations from a resource centre to provide the seamless access to information resources as fast as possible.

The role of a Digital Resource Manager for libraries in dissemination of digital resources varies with regard to technological infrastructure, collection development, financial outlays and manpower among others.

2. Digital Resources

With the emergence of new technologies, its services and latest formats of digital information, the current-generation libraries are being pushed to the outskirts of the non-digital era. However, the real utility of the next-generation libraries will be realized over time, as innovative technologies develop rapidly to get the standardisations. To fulfill the demand of the digital savy users for in time access of today's new services and applications, deployment of digital resources across the net and its smooth access raise new technical challenges to digital managers. The responsibilities, the technical specialisations and problems of digital resources compound the challenges to digital managers in terms of the factors shown in Fig 1.

Scientific Information Resource Division, Bhabha Atomic Research Centre, Trombay, Mumbai 4000 85, India

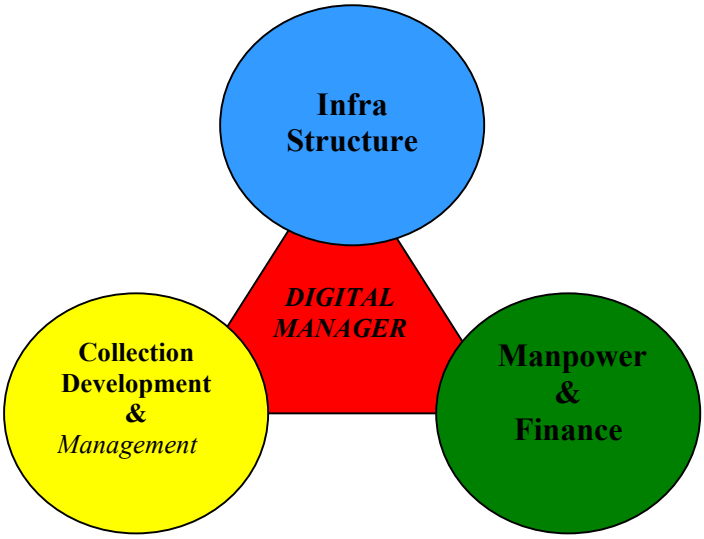


Fig. 1

2.1 Technological Infrastructure

The technological infrastructure of a digital library, which ensures the smooth flow of information across the organisation net, is very significant and important for digital managers. To future-proof the networks, increase bandwidth on demand and QoS requirements for new digital era, digital managers must deploy scalable technological infrastructure under secured network scenario that have been designed to address 21st century digital resources requirements. Hence, it is essential to design, plan and develop technological infrastructure according to latest & modern technologies. Procurement of latest need-to-access equipment with high-end network backbones, Servers, Network attached storages (NAS), operating systems etc. should be implemented. It is also essential to migrate from old technological infrastructure such as low-end desktops to high-end multimedia rich desktops/notebooks, Single Processor Servers with Dual Hot Swappable & RAID based Servers, Standalone Printers to Network Printers, 10 MBPS Media to 100 or Giga or Fiber channel media, Narrowband applications to Broadband ones, thick clients to thin clients, upgradation of Operating systems with latest patches/service packs etc. to new technological dimensions. With the growing number of digital resources and global information technology impact, the needs of effective management of digital era are redefined. The need of huge technological infrastructure poses greater challenges to today's digital managers.

2.2 Collection Development & Management

The history of collection development starts collection from carving on stones/wall to scripts on leaves and to print form of paper media. Later technological revolution made it possible to store in sequential tapes to floppy and CD-ROM and further to now emerging technology of DVD & Optical Discs. In the context of collection development & management, the impact of electronic revolution has been quite high for research and academic libraries. In recent years, there has been a remarkable shift & trends in organisations to work without boundaries and in a better open IT environment. The form and format of collection development also changes with latest trends of publishing. Portability of old collection development is the main bottleneck for today's digital managers while providing wide access to these collections. It is, therefore, essential to re-define the collection development & management policies and functions, Collection analysis, Material selection, Collection maintenance, Fiscal management and User

Liaison". To streamline, some of the collection development & management policies, which should be adopted by a typical library, are:

- **Qualitative selection:** Assess the user requirements either by inviting their collection development needs or search for their requirements from prospective suppliers/publishers. A team of library members should conduct the qualitative analysis from time to time.

- **Mode of access:** Various modes of collection development now are available starting for stand alone to network to on-site to on-line. The policy for mode of access should be carefully chosen in view of future technology migration and simultaneous access of collection development.

- **Media selection:** Formulation of policy for acquisition of print and electronic media should also be the need of the hour, especially for electronic media. The outdated media such as 5.25 floppy or sequential tapes should not be encouraged.

- **Technology:** Various technologies such as Standalone, Client/Server, Distributed, Networked, Intranet, On-site, On-line etc. are available on several platforms such as DOS, Windows, Unix, Linux, Solaris, Banyan Vines or Macintosh. Policies should be firmed up according to organisational need and infrastructure.

- **Updation:** Mode and frequency of collection update is very essential for providing just-in-time access of collection developments. Various modes of updation are available such as downloads through FTP and updation on CD-ROM/ DVD. Provision for frequency update of collection development and its mode should be tabulated as a policy.

- **Copyright & licensing:** Copy Right & Licensing of electronic collection development and management should be evaluated carefully before signing the agreements. Licensing is also available in wide range such as individual access, site wide access, institutional access, multi site organisational or consortia from most of the supplier/publishers.

- **Archive & Access:** Policies for conservation of original digital resources should be framed according to latest technologies and scope for future migration should be planned for better management of collection development. Access is very key factor to be considered while deploying the electronic resources or its archivals by digital managers.

2.3 Man Power and Financial Implications

Human interface to digital or conventional libraries is a fundamental requirement. Libraries without human interface are only conceptual things to happen in the next few decades, but manpower requirement in the context of changing roles & responsibilities of digital managers is growing quite high technically. Starting from acquisition to cataloging, collection development & management to access, storage and retrieval to archival of digital resources, manpower is essential. Appropriate specialists in each field, especially in IT, will enable the digital managers to meet up to some extent the challenges coming into their path.

With increased availability of electronic resources and huge increase in production cost of print media, the digital managers are facing lot of budgeting problems. To meet the annual increase in running cost of digital resources, sufficient budget sanctions should be made available in each financial year. Various libraries face financial problems due to lower priority given this aspect within the organisation. It is high time that plans, polices and strategies should be drafted more firmly to reduce the extra budgetary burdens.

3. The Need

The rollout of value-based Networking is finally emerging, giving digital managers the opportunity to introduce enhanced services. But what type of digital services can be expected from these new networks is a big question for digital managers. There is range of technologies to choose from, for building-up the value-added library networks, type of network access, computation concepts etc. Network access modes are broadly divided into centralized and distributed technologies. Sub Network access categories are:

3.1 **Centralized Network Access:** Networks providing the data & information from central node or nodes with single control & management.

3.1.1 **LAN**: Access, control and management of data & information for limited access area such as within a building.

3.1.2 **Extended LAN**: Access, control and management of data & information for another limited access area such as interconnections of different buildings within an institute / organisation /campus.

3.1.3 **Closed loop Intranet**: Access, control and management of data & information for limited or extended access area using Internet services, tools and topologies in closed loop of institute / organisation /campus network(s).

3.1.4 **Extranet**: Similar to closed loop Intranet but access to other closed or open loop Intranet with proper authorisation.

3.2 **Distributed Network Access**: Networks providing the data & information from distributed node or nodes with independent control & management.

3.2.1 **Internet**: Network of Networks distributed across the globe.

3.2.2 **Enterprise Network**: Similar to Extranet with distributed data access/processing capabilities across the globe.

3.2.3 **MAN**: Limited access to LANs within a metropolitan limit with distributed capabilities. Not necessarily using Internet/Intranet technologies.

3.2.4 **WAN**: Access to MANs or LANs in a distributed scenario. Not necessarily using Internet/Intranet technologies.

4. Online Resources

With careful watch on Library budget, it is advisable to selectively incur expenditure on technologies that will improve the library's productivity in targeted areas, in anticipation of future needs. This is the perfect time for digital managers to maximize resources and manage them efficiently in a digital library. It is wise to choose the most appropriate and cost effective solution & technologies for deployment of digital information across the Net.

Internet Access (On-Line) Model: The Internet community has increased from 15.3 millions to 68.6 millions in the last two years. Many publishers are exploiting the potential of Internet for online access of digital information. This model basically consists of two parts, one is Internet connectivity and other is subscription of online access of resources at publisher/agent/sub-agents web site.

In Indian scenario where the cyber laws are under progress and existing paper-based reading habits, it is very difficult to decide on subscription-based accessibility to digital information through Internet. This model becomes very cumbersome with its complex online subscription, in-demand access speed, users access rights and archival at the end of subscription and, moreover overall operational cost. This model will be more feasible to those academic and research libraries where Internet is easily and widely available within the organisation and may not have any security restrictions. However, it offers overall cost reduction per person when there is substantial increase in targeted users with high bandwidth connectivity. It should also be kept in mind the future existence of ISPs for smooth connectivity with proper bandwidth & upgradation before choosing. Security threats to these models are always in cloud even with implementing multi layer firewall systems. In the ever-changing Internet technologies, this model is highly recommended to libraries with low less security implication. Fig. 2 shows a typical Internet (Online) access model.

Fig. 2

5. Online Information Gateway: Lakshya

Considering the merits of online access model of digital reference a typical library should initiate online digital access with gradual increase in digital online resources.

In late 2002 an online information gateway named 'Lakshya' with online access to few hundred scientific and technical electronic journals was developed and established. Lakshya provides a single point access to the digital collection and services to scientist and engineers of BARC within the Internet network. Currently more than 2500 E-journals are available on Lakshya, a large number of them with full text. In addition many references and bibliographic databases like INIS on Internet, Medline, and Hydraulic pump standards etc have also been integrated into Lakshya. Planning, designing and

development of Lakshya were carried out with an endeavor to provide seamless access of online resources without hassles.

5.1 Lakshya: The front Layout :

Fig. 3

5.2 Lakshya: Infrastructure setup

Lakshya has been setup over BARC Internet Network and can be accessed within the institute either through centralized Internet centers at various locations through LAN or dialup ISDN connection from desktop of individuals. The Lakshya server operates in Windows 2000 environment with IIS 6, MS-Back office, PHP and Perl.

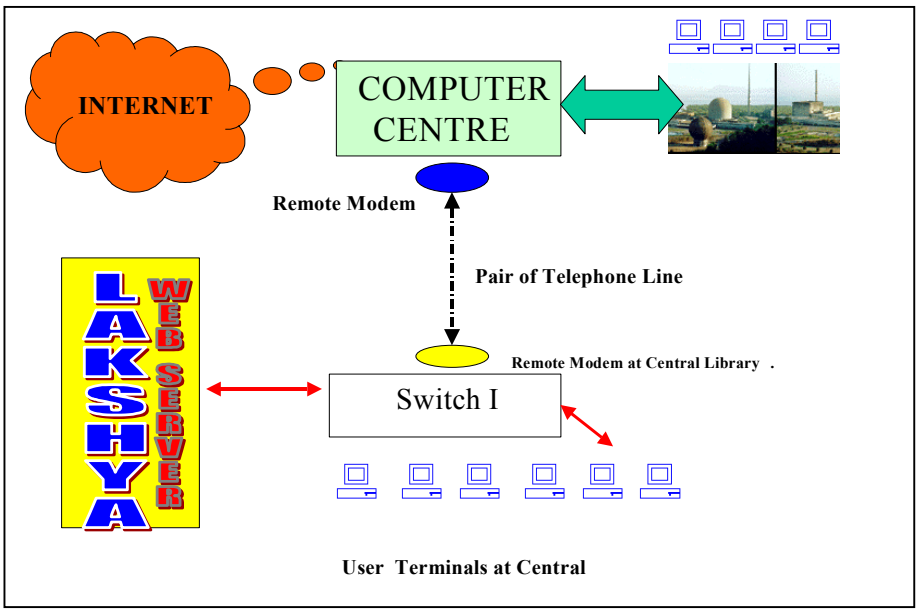


Fig. 4

5.3 Lakshya: Mode of Access

In current practice there are three methods of authentic access of web site as

- **Anonymous access**: mostly web access is anonymous to the global community.
- **Login and password:** to access the resources of web site, web administrator can restrict the access with login & password.
- **Institute IP authentication:** many web resources have provided the IP (internet protocol) based access. In this case all the Internet access from the premises of institute IP address are authentic.

A large number of journals are covered under a Consortium. In addition journals form publishers who are providing complimentary online access to their E-Journal archivals against currently subscribed print journals in the Central Library, are also available.

Apart from these, many journals, which are available freely, are also included in Lakshya. In coming months, many more journals and other E-resources would also be hosted on Lakshya, some of them under consortia.

5.4 Lakshya: An array of value added features

The complete list of available online resources on Lakshya web site is mainly categorized into two sections viz. alphabetic list of online journals wise collection and alphabetic list of publisher wise collection. Some of the major features of Lakshya are

- User friendly and seamless access to E-Journals available on either IP address based or login ID based access
- Downloadable Full text articles in different file formats as supported by publishers
- Search for E-Journals available on Lakshya, BARC
- Title-wise alphabetical list of E-Journals
- Publisher-wise alphabetical list of E-Journal titles
- Online references, encyclopedias, Internet-INIS database, E-Books, Useful web sites etc. are also available
- Bulletin Board for news updates on trial online offers from various publishers/agents
- Online facility for submission of suggestions /Feedback from users
- Consolidated list of useful links

Lakshya : Future Plans

How to manage 3C (changes, challenges and Content) is a very difficult task for a digital manager with on going demands for fast and wide subject coverage. To facilitate better search results from Lakshya, a complete revamping of search interface is being developed and is under testing phase. With new search interface, one can search the title either by keywords or descriptors with one click hyper link to online journal.
An upcoming FAQ facility related to online resource will also help the users to communicate in a better manner. Moreover many technicalities and know-how about online access are also covered in FAQs. Fig. 5 shows the front end of upcoming Lakshya.

Fig. 5

6. Conclusion

The infrastructure of the new value-added online access will allow digital managers to reduce their content development efforts, their digital deployment time, to unlock innovations, and to provide easier management of digital resources, which in turn will reduce the operational and capital expenses. The ability to rapidly introduce digital information that will provide long-term benefits to the research community will be the real achievement of online solution for digital managers.

References

Cogswell (J). The organisation of collection management functions in academic research libraries. (J Acad Librarianship 13; 1987, 268-76)

http://www.citrix.com

http://www.silverplatter.com

http://www.sciencedirect.com

http://www.itcl.com

Citrix: Server Based Computing White paper

Information Management

Content Management System: A Case Study of IGCAR Library

Varathan K., Soundararajan E. and Somasekharan M.

Abstract

As libraries build larger collections of electronic resources, finding ways to manage them efficiently becomes a major challenge. This paper describes how the IGCAR Library, Kalpakkam has been changing its direction from traditional library to digital library. For better understanding of digital content creation and preservation, this paper focuses on three different techniques namely digitization, Meta data harvesting, and preservation of digital documents. This paper also highlights the tools, techniques and standards that are being used in IGCAR Library.

Keywords: Content Management Cycle, Metadata, OCR, OAI-PMH, Information Storage & Retrieval, Digital Asset Management, Web Publishing.

1. Introduction

Digital library is a collection of electronic resources and associated technical capabilities for creating, searching, enhancement of information storage and retrieval systems. The real digitization job for librarians today is to convert available in-house resources into digital form. IGCAR digital library play pivotal role on dissemination of information for our research community. The E-collection of ICGAR Library includes e-journals, e-reports, FBR (Fast Breeder Reactor) conference proceedings, Standards, CD Resources, and other technical documents. There is a heavy demand from Scientists/Engineers of our center that, the organization's in-house document must be made available on the network (intranet) so that they can access them any time from their desktops. As part of digitization activity the following in-house resources are digitized and made available to library patrons on the intranet.

- IGC Research Reports
- IGC News letters
- IGC Annual Reports
- Ph.D Thesis - Abstracts
- Internal Reports

Scientific Information Resource Division , Indira Gandhi Centre for Atomic Research, Kalpakkam – 603 102

2. Content Management Activities

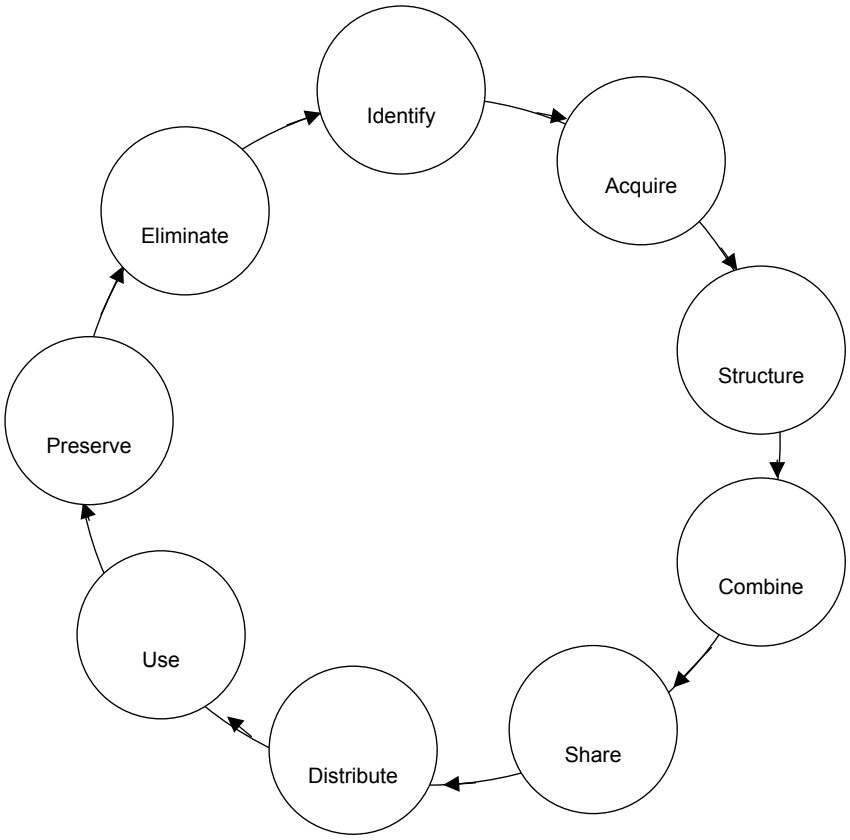


Fig.1 Content management Cycle Activity

Content Management involves a chain of activities on selected contents (Fig-1) such as identify, acquire, structure, combine, share, distribute, use, preserve, and eliminate. It provides dynamic access to large content pools, digital assets, and Dynamic Medias. The CLM (Content Life cycle management) uses standards to store and retrieve content. Metadata plays a vital role for creating centralized repository of controls and efficient retrieval.

2.1 Conversion of Print resources to digital format

First step in digitization process is selecting the content to be digitized. The list of document for digitization should be sent to an expert committee for selection of relevant contents. The approved documents are then digitized using appropriate hardware. The digitized documents are edited by using editing software tools and converted to PDF Format by Adobe Acrobat Suit. The entire PDF document is made searchable by running OCR (Optical Character Recognition) software, published on to web and made available to Intranet. Backup and recovery process will ensure the continuous access to the resources.

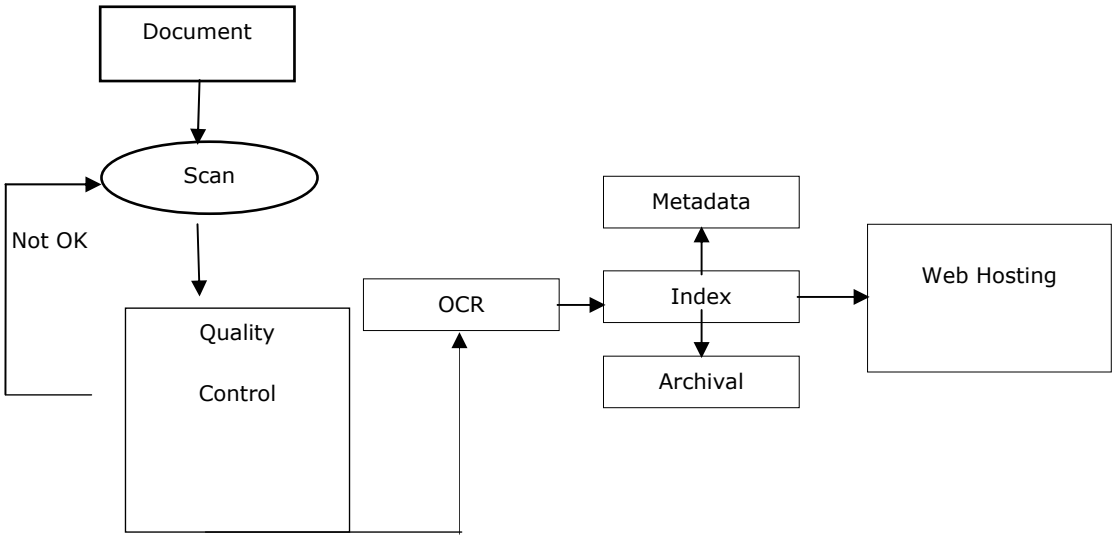


Fig.2 Content management System (Work Flow)

2.2 Infrastructure requirements for Content Creation

Content creation activity requires hardware and software tools. A System with good memory and storage, a scanner, network environment and server class system are required for creating and publishing the content. Software tools are required for image editing, OCR and metadata creation/management. Also a user friendly website is required for easy access by patrons. IGCAR Library has the following computer hardware and software Infrastructure for the content creation activity.

Hardware Tools:

- Flatbed Scanner – hp Scanjet 3570C
- Auto Document Feeder - hp Scanjet 7400C
- Book Scanner - Minolta 7000
- Microfilm/Microfiche Scanner Minolta MS 3000 & 7000
- Servers & Workstations

Fig.3 Book Scanner

Software Tools:

- Adobe Photoshop 6.0
- OCR OmniPage Pro 11.0
- Adobe Acrobat 6.0
- Front Page 2000

Fig.4. Microfilm Scanner

2.3 IGC Reports digitization activity:

The IGC Reports are scanned using suitable scanner with required resolution. The Scanned tiff image files are edited through Photoshop and compiled as PDF document. The PDF (Portable Document Format) is chosen as a global format for content because PDF format has been a de facto Internet standard. It guarantees that the image seen by the viewer is congruent across all platforms. Because of the compression methods used by PDF, the size of converted documented into a PDF reduce considerably of its original size. PDF documents can be quickly retrieved on an Internet/Intranet. Digital document must be made read-only which provides good security against virus & other attacks.

The digitized documents are moved to a folder for uploading them to server using file transfer protocol (FTP). The created contents in PDF format are indexed with Report No., Title, Author, Abstract and Subject (Fig-5). The created meta-data are stored in the database systems for efficient retrieval.

Content management systems often provide support for integrated indexing and search functionality. By embedding categorized metadata it is possible to fine-tune desired search results and ensure that patron can easily find what they want.

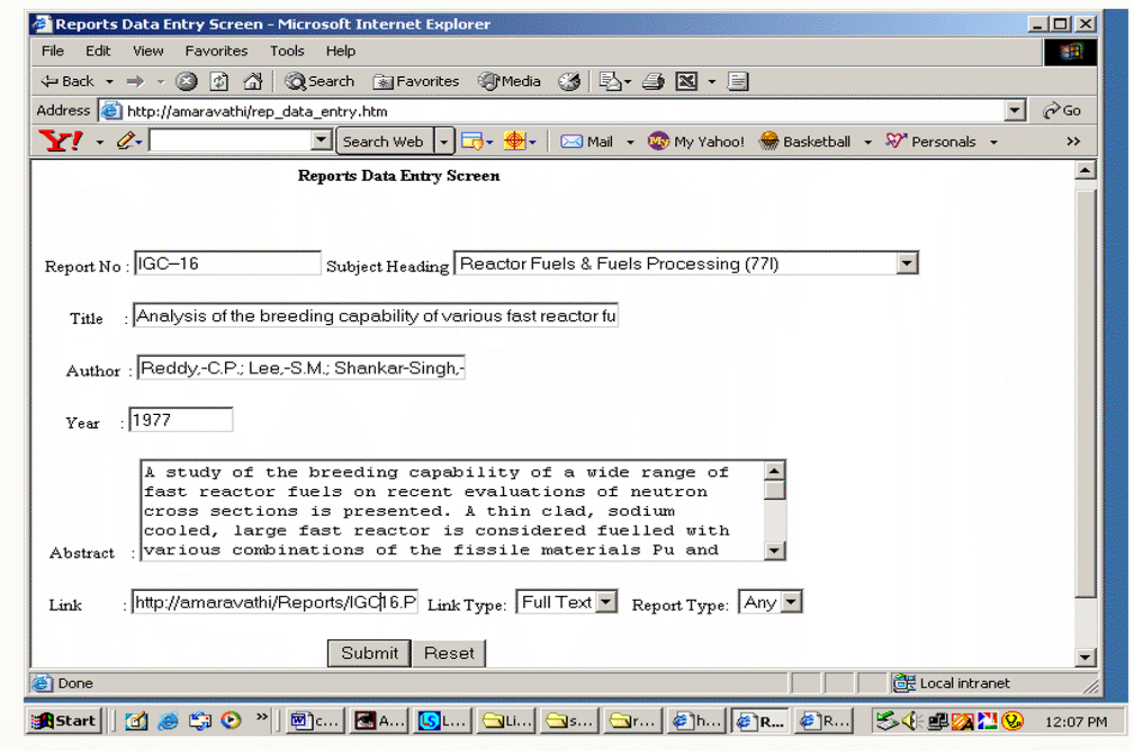


Fig.5: IGC Reports metadata Input form

3. Metadata Harvesting

The common definition is that metadata is “data about data.” Metadata is one of the important concepts for the description, organization, exchange and retrieval of information in the network environment. When creating metadata records to describe a resource,

thesauri and controlled vocabularies are widely used as a basis for both resource description/discovery and information management purposes. The minimum metadata sets are required for harvesting such as Document Identifier, Title, Description, and Subjects.
OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is mechanism for harvesting XML-formatted metadata from distributed collections of metadata. This Protocol provides the basic information discovery environment that relies on transferring metadata en masse from one server to another in a network of information systems based on the open standards HTTP (Hypertext Transport Protocol) and XML (Extensible Markup Language).

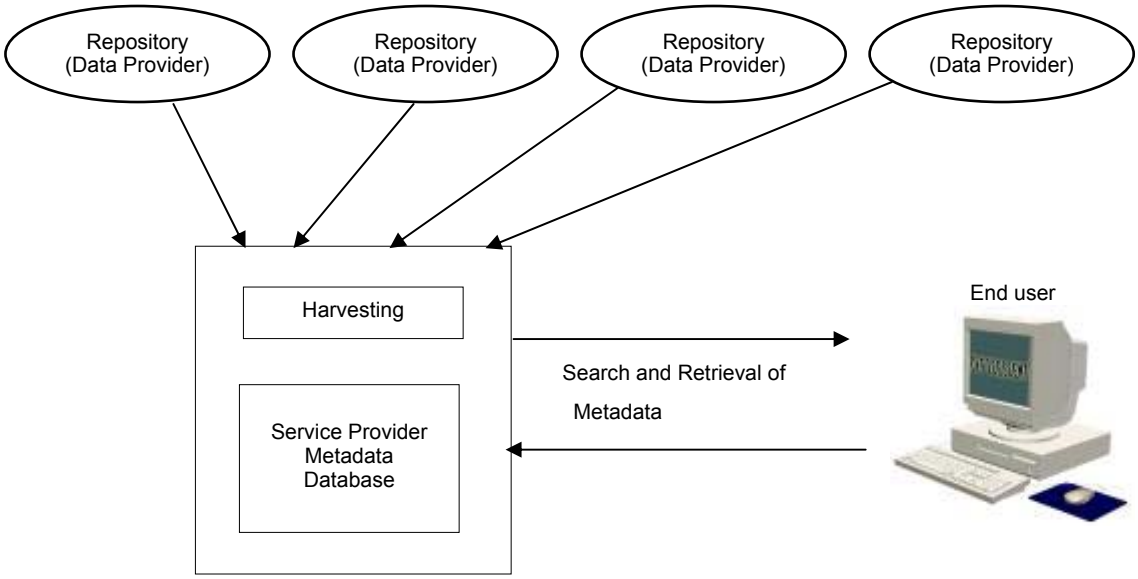


Fig.6 Basic approach of OAI-PMH

The metadata that is harvested may be in any format that is agreed by a community (or by any discrete set of data and service providers), although unqualified Dublin Core is specified to provide a basic level of interoperability. Thus, metadata from many sources can be gathered together in one database, and services can be provided based on this centrally harvested or "aggregated" data. The link between this metadata and the related

Content is not defined by the OAI protocol. It is important to realize that OAI-PMH does not provide a search across this data; it simply makes it possible to bring the data together in one place.

4. Storage & Retrieval

There is an increasing need on providing search on the full text of digitized documents in addition to the bibliographic level. It is required to develop a storage & retrieval system which takes the advantage of structural knowledge from the assigned index terms on various search fields. There is an increasing need to research on the impact of this capability and to develop a storage and retrieval system, which takes advantage of structural knowledge. The integration of structural and textual information can allow one to achieve a higher quality of retrieval results.

Standard Generalized Markup Language (SGML) provides a very powerful tool for describing document. Based on SGML, Hypertexts Markup Language (HTML), Hypermedia/Time-based Structuring Language (HyTime), and Text Encoding Initiative

(TEI) have been proposed as encoding standards. SGML documents have information about document structure, as well as the contents.

The digitized documents must be made available to users easily through web, any time, anywhere, and to anyone. IGCAR library has indexed documents based on selected subject categories and titles alphabetically. Documents are then organized under Report Number, Title, Author, Year of Publication, and Abstract. A web server is required to host the digitized contents and make it available to user desktops. Organized access to information is provided through a web interface.

4.1 Web Publishing of IGC Reports

Digitized IGC reports of various disciplines mainly nuclear science & engineering are published through Web Interface and made available for users desktop with search facility. There is also a collection of Conference proceeding available on FBR (Fast Breeder Reactor), IGC Publications like reports, newsletters, and thesis are also accessible from here. The open source software like Linux, Apache Web Server, PHP, MySQL are being used for storage& retrieval and web publishing of contents.

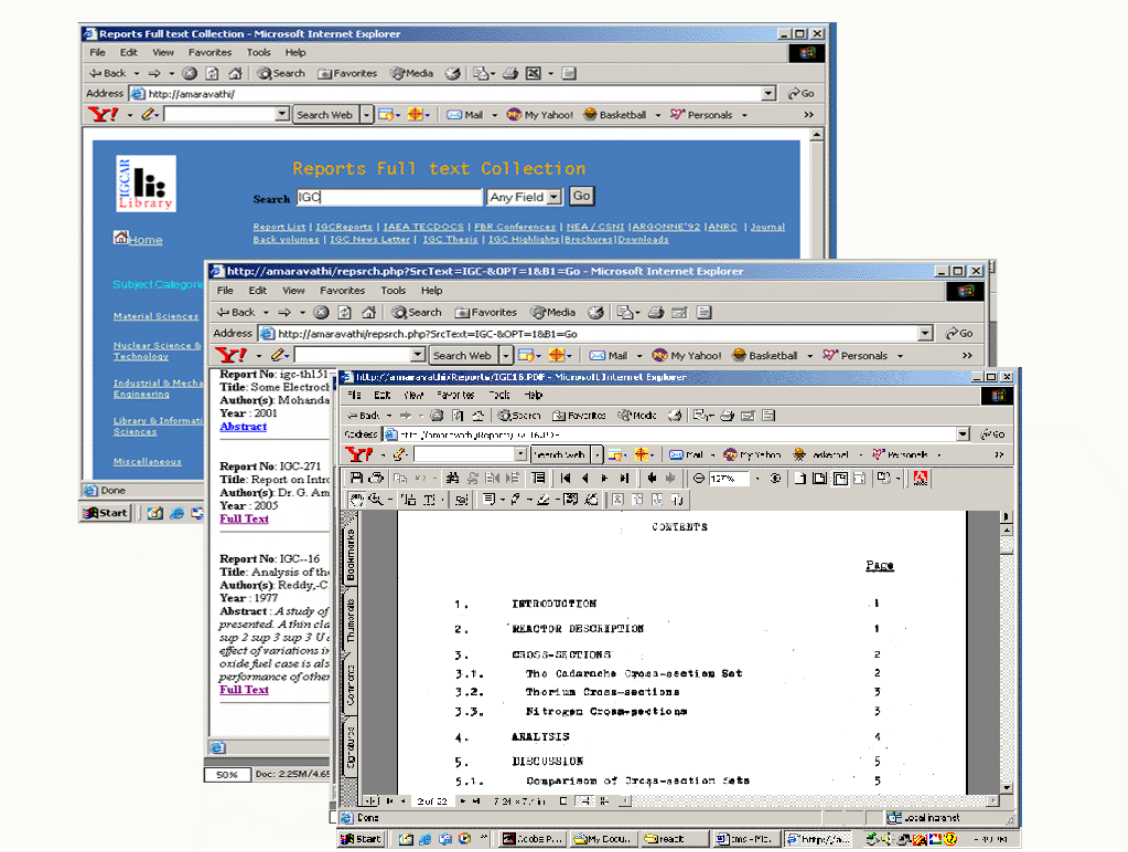


Fig. 7: Web Publishing of IGC Reports

5. Preservation & Digital Asset Management

Preservation is the immediate issue for digital resources, which ensure their continued accessibility. Digital Asset Management (DAM) is a set of coordinated technologies and procedures that allow the efficient storage, retrieval and reuse of the digital resources. There are two important activities of DAM i.e. Migration and Up gradation. Migration

ensures the usability of digital resources in spite of changing technologies. Up gradation maintains the versions of software and hardware being used for content management and retrieval. Outdated materials can either be archived offline or removed from the system. Hence there is a need to constantly store, transfer, and backup data. The data can be backed up on CD/DVD media or DAT Tape depending upon volume of information. The new generation 'blue-ray' DVD technology support up to 25GB of storage capacity. The ultrium generation 3 Tapes provide up to 100GB native capacity. It is required to maintain a separate database on backup activities. This will be helpful in finding out the version of backup, date etc. Also more recent backup will be taken from this database for recovery incase of server crash.

6. Conclusion:

Library collections will continue to grow with materials in both physical and digital formats and there are challenges for us to organize those materials so that they can be maintained and accessed efficiently. Content preservation and retrieval are the important issues for librarians in the network environment. Constant migration of contents in terms of storage media, technology and software tools will ensure not only the preservation of contents but also their accessibility in spite of ever changing technology. Although lots of vendor based CMS tools are available in market, open source solutions will be of great help to Digital Librarians. Notably the open source projects like e-prints, Dspace, fedora play vital role in repository building & content management. MySQL & PHP are the most widely used open source database and scripting technologies on the Web today and provide a good platform for implementing and developing open source content management systems.

Reference:

1 Marilyn Deegan and Simon Tanner. Digital Futures. London: Library Association Publishing, 2002. ISBN 1-85604-411-4

2 D.Kamalavijayan. Information & Knowledge Management. India: Macmillan, 2005, ISBN 1403 92652 2

3 E. Soundarajan. The Integrated Search System. In: Knowledge Creation, Preservation, Access, and Management. An International Conference on Digital Libraries 2004 held at The Energy and Resources Institute, |24 – 27| February 2004. New Delhi: TERI, ICDL 2004, pp 912-916.

4 Mark Needleman, Column Editor. The Open Archives Initiative. Serial Review, vol.28, no.3, Summer 2002, pp.156-158.

5 Usha Mujoo-Munshi. Building Digital Resources: Creating Facilities at INSA. The International Information & Library Review, vol.35, no.2-4, June-December 2003, pp. 281-309

6 Adam Chandler and Elaine L. Westbrooks. Distributing non-MARC metadata: the CUGIR metadata sharing project. Library Collections, Acquisitions, and Technical Services, vol. 26, no. 3, Autumn 2002, pp. 207-217.

7 Hak-Gyoon Kim and Sung-Bae Cho. Structured storage and retrieval of SGML documents using Grove. Information Processing & Management, vol. 36, no. 4, July 2000, pp. 643-657

Overview of Object Oriented Databases

Narayanan K.R.S. and Jayanthi T.

Abstract

Object Oriented Data Base (OODB) has all the features, functionality of a relational database system and also offers an Object Oriented Programming language interface, user defined data types, object identifiers and persistent object maintenance. OODBs provide facility to build complex hierarchical data models easily. This paper presents the object models, standards, various object oriented DBMS, and object oriented programming languages, object query language, performance of OODBs and various application areas where OODBs can play a major role.

1. Introduction

Data management and organization have become so complex and challenging in today's electronic age of information. Databases, be it bibliographic or textual, ought to have the capability of storing graphics, video, audio and other highly structured data. The database technologies have constantly evolved to meet these changing requirements by adopting object oriented programming concepts. During the last two decades, Relational Database Management System (RDBM) has been established as the technology, handling databases up to terabytes. However Relational Databases lack the mechanisms to deal with complex structured data. Their tabular approach does not allow a suitable modeling of complex hierarchical objects. Although most Relational databases support binary large objects (BLOBs) to store graphical data (e.g. geometry, textures), these objects cannot be queried in the same way as other data types. There would be performance degradation if many BLOBs were used. The limitations of Relational database for Geographical Information System, CAD, Multimedia, Engineering etc. have led to the development of Object-oriented Database Systems.

2. Concept

Object-oriented concept stemsfrom object oriented programming. Objects are independent computer structures representing real world entities. They interact with themselves and with other objects. C++ and Java are popular object oriented programming languages. They provide powerful modeling and development interface. Objects carry their own data and procedures. Each object has a few attributes and has a relation with other objects. Object orientation supports abstract, complex and multimedia data types. Objects can send messages to other objects and an object ID uniquely identifies each object.

Each object has its own attributes. Attributes are known as instant variables. Attributes will have a name and associated data type. Methods are the code that operates on object's data. A class is a collection of objects that share common characteristics. A class will have details about data structure and method implementation for the objects. Figure 1 describes a class and it's instances.

Computer Division, IGCAR, Kalpakkam-603102.; E-mail: krs@igcar.ernet.in

An OODB combines object oriented programming principles with database management principles. Object oriented programming concepts such as encapsulation, polymorphism and inheritance are enforced along with regular database management concepts such as the Atomicity, Consistency, Isolation and Durability (ACID properties). OODB is a system while supporting all the functionality of a relational database system (including queries, transactions, backup and recovery mechanisms), also offers an Object oriented programming language interface, user defined data types, object identifiers and the ability to manage objects persistently.

3. Object Oriented Data Model (OODM)

Object oriented data model represents real-world entities as objects. Each object's attributes and methods implementation is hidden from other object. A unique object ID identifies each object. Classes inherits all properties of their super classes.T he OODM is different from E-R and Relational model in the sense that the OODM object has additional properties like behavior, inheritance and encapsulation. Such features make OODM more close to real world entities. OODM classes allow the implementation of Abstract Data Types (ABT). The ABT is used to create new data types and is a powerful modeling tool. In relational model the relationship among entities is established through a common attribute. OODM in contrast uses object ID to establish relation among objects.

Objects can be used persistently and transiently. Persistent instances of objects are automatically and transparently stored by the OODB. Transient instances exist only while the application based on the OODB is running. Locking and transaction mechanisms guarantee security and integrity for multi-user update operations.

4. The Object Data Management Standard (ODMG Standard)

ODMG was founded in 1991 to define the criteria for OODBs, to provide a feature list for OODB comparisons and to introduce standards into the OODB world. Most major OODB vendors are members of this group. The ODMG has published several documents called ´ODMG Standards.' The first version of this standard appeared in 1993. The current version is ODMG 2.0. This document defines an object model, an object definition language (ODL), an object query language (OQL) and interfaces for the programming languages C++, Smalltalk and Java.

5. Object Definition Language (Odl) and Object Query Language (OQL)

ODL describes the database schema, i.e. the hierarchy of object classes, their elements and methods. The ODL is independent of the programming language (C++, Java, Smalltalk). A pre-processor translates the schema into class definitions for the programming language.

The ODMG-OQL is based on O_2SQL (which in turn is based on SQL 92). In addition to SQL-style queries, OQL can be used to query complex types of data. It can make use of the methods in classes. The OQL does not support recursive queries.

6. Object Oriented Database Management Systems (OODBMS)

An OODBMS is the result of combining object oriented programming principles with database management principles. An OODBMS is a full-scale object oriented development environment as well as a database management system. Features that are common in the RDBMS such as transactions, the ability to handle large amounts of data, indexes, deadlock detection, backup and restoration features and data recovery mechanisms also exist in the OODBMS.

In OODBMS accessing objects in the database is done in a transparent manner such that interaction with persistent objects is no different from interacting with in-memory objects. This is very different from using an RDBMS. In that there is no need to interact with SQL nor is there a reason to use a Call Level Interface such as ODBC, ADO or JDBC. Database operations typically involve obtaining a database root from the OODBMS which is usually a data structure like a graph, vector, hash table, or set and traversing it to obtain objects to create, update or delete from the database.

There are concepts in the relational database model that are similar to those in the object database model. A relation or table in a relational database can be considered to be analogous to a class in an object database. A tuple is similar to an instance of a class but is different in that it has attributes but no behaviors. A column in a tuple is similar to a class attribute except that a column can hold only primitive data types while a class attribute can hold data of any type.

In OODBMS it is possible to have a large class, which holds many medium sized classes, which themselves hold many smaller classes. In a relational database this has to be done either by having one huge table with lots of null fields or via a number of smaller, normalized tables which are linked via foreign keys. A join has to be performed every time. Also an object is a better model of the real world entity than the relational tuples with regards to complex objects. The fact that an OODBMS is better suited to handling complex, interrelated data than an RDBMS means that an OODBMS can outperform RDBMS. The ODMBS require less code to develop, reduced development time, and reduced maintenance costs

Object Store, Objectivity, Shore and Versant are some of the popular OODBMS.

7. OODBS in Digital Library

OODBs provide better support and higher performance for hierarchical and complex objects. OODBs play a major role in the development of wide variety of target applications including hardware and software CAD systems, persistent programming languages, geographic information systems, satellite data repositories, digital library and multi-media applications.

In a digital library OODB can be used for document analysis in multivalent document architecture, in using computer vision to extract content from images for indexing and retrieval. OODB In digital library can play a major role in the development of multimedia applications. Text reports and scientific data can be subjected to data mining using OODB applications to extract knowledge, which would help in decision-making.

Multimedia databases include video, images, audio and text media. They can be stored on extended object-relational or object-oriented databases. Multimedia is characterized by its high dimensionality, which makes data mining even more challenging. OODBs can be used in content-based search and video retrieval from digital libraries over computer networks.

Automatic Extraction of Proper names and Keywords from Web Resources

Velumani G. and Sivasamy K.

Abstract

One of the major Challenges facing information professionals today relates to effective mechanisms for retrieval of information from the Web. Traditionally librarians have used metadata as a tool for management of information resources, and their effective retrieval. However, given the volume of information on the Web and the rate at which it keeps growing only mechanisms that can automatically extract such data will be useful. The existing mechanisms such as search engines appear to rely on full text and index practically every word appearing in a text. This leads to the problem of unacceptably low levels of precision during searches. This research is based on the premise that names of persons, corporate bodies and keywords present in a text are important in terms their value as search keys for a document. This paper describes methodologies that have been developed and are under evaluation fort the automatic identification and extraction of Names of persons, Names of Corporate bodies and keywords from web resources. Some of the problems encountered that are being addressed are also discussed.

1. Introduction:

An issue that has attracted considerable attention in recent years relates to indexing of web documents in an effective manner so as to facilitate acceptable levels of precision in retrieval. Much of this research and discussions are based on the assumption that existing tools for information retrieval from the web are inadequate and suffer from major limitations. Search Engines and Subject Directories are the major tools currently available for resource discovery from the Web. The major limitations of search engines in facilitating effective information retrieval derive largely from the limitations of the mechanisms they employ for creating their databases. Of course there are differences between various search engines with regard to what portion of the web document is used to derive index terms. However, by and large most search engines derive their index terms from web documents with the help of specially developed programs referred to variously in literature as robots, spiders, etc. In reality this process of extracting index terms has several major limitations

- It could, and often does, result in extracting several index terms that are not relevant search keys for the document in question leading to unacceptably low levels of precision
- The existing mechanisms also do not have any means of distinguishing between different kinds of terms that may be present in the text being indexed. For example, there is no way of identifying whether a term that is extracted denotes the name of a person or is a keyword indicating the subject matter of the web document.

- In this paper we discuss the results of the experiment carried out to identify and automatically extract

Research Scholor, Dept. Inf. Sci., University of Madras, Chennai.

- Proper names and
- Keywords / key phrases from web documents

2. Scope and Objectives:

The important elements of description (metadata) of any documentary resource from the point of view of their utility as search keys are:

- The proper names associated with the resource; This would include names of persons, names of institutions and other corporate bodies; and
- Keywords and Key phrases representing the subject content of the document.

If effective mechanisms could be developed and implemented to identify these and extract them from electronic resources these will be useful metadata. The experiments reported in this paper were carried out to develop and test algorithms to identify and extract Proper names and keywords from web documents. The mechanism uses reasonably fast and robust heuristics to identify proper names, keywords and extract them from the web resources.

3. The experiment:

There are basically two aspects to this study. The first one focused on identification and extraction of proper names and the second one on keywords. A corpus of HTML documents on the Web was used as the test bed to experiment with the algorithms developed. The first step was to arrive at a corpus of electronic texts to experiment with. In planning such a study it was decided, keeping in mind a wide variety of factors, to conduct the experiments with documents falling in a subject domain. This was necessary as one component of the study focused on extraction of keywords.

3.1 Identification of Proper names: For the purpose of this experiment and study 'name extraction' is defined as the process of identifying and extracting personal names from unstructured web texts in the English language. A set of rules to enable a computer to identify names in all their variations is an essential component of the kind of text processing application described in this paper. Literature on the subject has references to a few rules identified and applied to extract names. Most of these rules are derived from standard conventions that are widely employed by authors of properly edited texts in English. Some of the key indicators that were used included:

- Use of legend words such as Mr. Miss, Ms. etc is a good indicator that the following word / words denotes a name;
- Use of Corporate trigger words such as 'University', 'College', 'School', etc., is a good indicator to identify the organization names. Prepositions are widely used in corporate names to link legend words with other words (e.g. University of California); this has been exploited to identify complete corporate names
- Initial capitalization is also an indicator of names since the convention of English language requires that each word of a name start with an uppercase letter. In this module, therefore, identification of initial capitalization has been used in formulating the algorithm.
- Example: Powell
- First Name, Middle Name and Last Name starting with uppercase letters is also a good indicator for names[2].
- Example: Mohandas Karamchand Gandhi

- Legend words prefixed to a word with initial upper case letter is a certain indicator of personal names.
- Example: Mr. Powell
- A single uppercase letter followed by a dot connecting two words both with initial upper case letters or a one or more single upper case letters each followed by a dot and further followed by one or more words with initial upper case letters are also good indicators of personal names.
- Examples: Bernard I. Palmer
- S. R. Ranganathan
- A.P.J. Abdul Kalam
- Jack Wells
- Two Consecutive words starting with uppercase letters and followed by a comma and single uppercase letter or a word beginning with an uppercase letter.
- Example: Wells, A. J.
- Mills, Jack

3.1 Name Extraction Process*:* The processes involved in name extraction could be summarized as below:

(i) Removal of HTML Tags
(ii) Tokenization
(iii) Checking for Trigger words
(iv) Identifying names without trigger words
(v) Splitting sequences into smaller names
(vi) Alphabetization
(vii) Establishing link with the source file
(viii) Meta-Name Creation

I. Removal of HTML Tags and unformatted characters: Any HTML file will necessarily carry tags beginning with angular brackets [<]. These tags did present certain problems. As a first step, therefore, the program removes all HTML tags and unformatted characters like (){}[] etc., from the downloaded files before executing the rest of the program. If a text file is used instead of HTML file this module automatically proceeds to the next step of tokenization. The module does not support other file formats such as PDF, ps, etc. Such files need to be converted to HTML format before applying this module.

II. Tokenization: Tokenization is the second step. In this process each line of the input text is broken into words and all sequences of capitalized tokens (or words) are collected and stored in a Temp. File for further processing. This process is repeated until the entire input text is analyzed.

III. Checking for Trigger words: A set of commonly used trigger words (also called as legend words) is stored in two text files called legend (personal names) and org names (corporate names). Each and every capitalized token extracted is compared with trigger words already stored in the file. Whenever a match is found, the words beginning with capitals following the legend word are and stored in a file called Name.

IV. Identifying Names without trigger words: After removal of names beginning with trigger words, the remaining capitalized tokens are examined for words that are preceded by initials. These words are also transferred to Nam**e** file.

V. Splitting sequences: When a sequence contains more than one name, it is in fact a linguistic structure that needs to be parsed. An intervening semicolon or colon is used to identify the different names. Such names extracted are also transferred to the Name file.

VI. Link with Source File: Each name is hyper-linked to the source file.

VII. Alphabetization: Each word/phrase extracted and stored in the Name file is subjected to the following processes:

- The trigger words, if any are removed

- Any initial letter that precedes the name, it will be transferred to the end of the name and a comma inserted between the initials and the name

All the words / phrases are now alphabetized.

3.2 Identification of Keyword / Key phrases:

In OPACs and similar databases the elements representing the subject of the resource usually take the form of a set of data fields that may include keywords, descriptors, subject headings, abstract, classification codes, etc. However, in automatic extraction of keywords from a document it is necessary to look for appropriate lexical clues. The major type of lexical clue to the subject of a document is the set of domain terms the document contains. This usually takes the form of keywords. The experiment reported in this paper involves the use of a set of simple heuristics to identify keywords and key phrases in HTML documents. The module involves the following major inputs:

- One or more HTML files constituting the documentary information resources in a domain from which keywords / key phrases are to be extracted automatically; The program requires that all the HTML files be in a single folder
- A database which is in effect a list of domain terms in the subject area / discipline of the HTML files: In this study the *ASIS thesaurus* was employed
- A database of 'stop words' consisting of all non-noun words taken from the *Pocket English dictionary* which itself is derived from the *New Oxford Dictionary of English*.

The principal output of the program is a HTML page consisting of extracted keywords / key phrases with hyperlinks to the HTML pages from which they were extracted.

Keyword Extraction: The major problems involved in KWE are extraction of keywords and omission of non-significant words. The experience with techniques such as those adopted by the popular search engines clearly brings out the need for a different approach. In this study it was decided to experiment with a validation process using two databases of terms to assist in the identification of keywords and non-significant words in the input file. The validation process employed made certain assumptions:

- It was assumed that a word / phrase in the input HTML file that is also part of a controlled vocabulary in the concerned subject domain is a key word / key phrase with a high probability of indicating the subject content of the input file.
- Non-noun words in the input file are assumed to be non-significant words.
- In the present experiment the following inputs / tools were employed:
- A paper entitled '*Information Retrieval and Cognitive Research*' was used as the input HTML document to test the utility and limitations of the Program. An idea of the paper can be had from the details given in the Table 1 below.
- As for identifying keywords and key phrases in the input file online tools were used..
- The ASIS thesaurus **(**http://www.asis.org/Publications/Thesaurus/isframe.htm)

i. Stop-word Terms (ST): Uncontrolled vocabularies have always presented problems in IR. The most common words in English may account for 50% or more of any given text. Their semantic content measured in terms of their value in describing / indicating the subject matter of the text is minimal. Further, such words tend to lessen the impact of frequency differences among the less common words. In addition, they necessitate a large amount of unnecessary processing. In all methods of automatic indexing such less significant words are ignored based on a stop-word list of such words. As already mentioned in the present experiment all non-noun words were

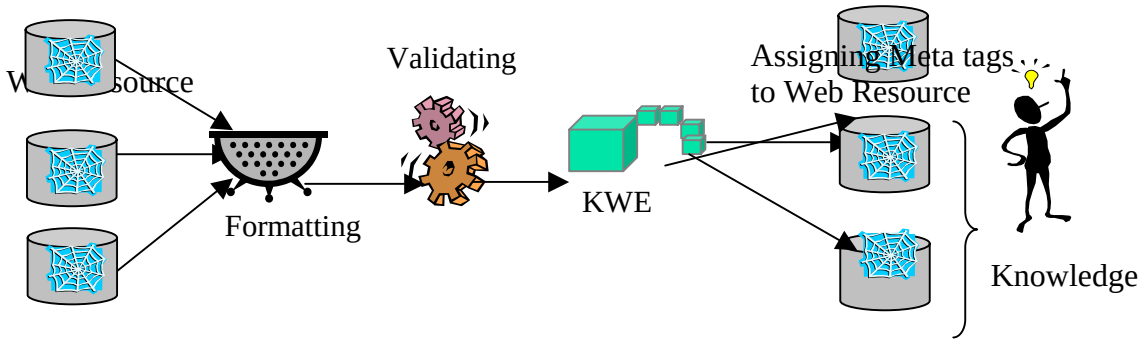
considered as non-significant. A database of non-noun words was created using the *Pocket Oxford English Dictionary* as the base. All the non-noun words in this dictionary were identified and stored in the database of stop words. The database contained about 13,000 non-significant words - adjectives, adverbs, prepositions, articles, conjunctions and verbs.

ii. Auto Index:

iii. *Auto Index* is the program developed in this research to extract keywords and key phrases keywords from HTML pages. It performs an intelligent identification of keywords from a clustered text and parses the output to give a distilled output. Extracting noun phrases (keywords) from a HTML file has been viewed as a pattern-matching task in this research. A detailed report of the module is being published separately. The generalized model of KWE employed in this research involves (*see also* the flow at the end of this paper):

iv. Formatting the Input HTML document: This consists of converting the input file to a format suitable for the algorithm and the processing tasks involved. The reformatted text constitutes the Master File.

v. Validating Process: A rigorous process based on established authority files to accept or reject a term in the input file as a valid keyword or not



4. Output and Discussion

It is relevant to indicate here that it took *Auto Index* about 30 seconds to generate the output and create a link HTML file linking keywords to the source document. An idea of the output, i.e. keywords and key phrases extracted by *Auto Index* can be had from the table below. The table lists the key terms extracted for different pre-determined levels of frequency (the figure against a term indicates the frequency of occurrence of the term in the input files).

Going by the output, it does indeed appear that it is feasible to employ techniques for automatic extraction of keywords from HTML files. Adopting a validation process can substantially enhance the quality of the output. The two principal validation processes experimented with in this research are:

- Using authority files (in this case a vocabulary control device and a database of non-noun words) to admit or reject a term found in the input file
- Defining and adopting a minimum level of frequency of occurrence of a key term in the input file

It is important to realize that the program will be effective when working with documents in a domain. There are a few limitations. The program in its present form can be used only with HTML pages in the English Language. The program requires the online availability of a good thesaurus / glossary of terms in the subject domain for its effective functioning. This

requires that the glossary must be regularly and frequently updated by addition of new domain terms, deletions and modifications that may be necessary.

Input HTML Document	
A paper on***: Information Retrieval and Cognitive Research*** (20 pages, 4660 words, 96 paragraphs, 564 Lines)	
Names of Person	[1] ALLEN, Bryce L. [2] Angel. Miguel [3] Arts, Liberal [4] BELKIN, Nicholas J. [5] BORKO, H. [6] Bibliotecn. R. [7] Bras. R. [8] CASTI, J. L. [9] CHRISTOV O, H. T. [10] Christine L. [11] DAHLBERG, I. [12] DANIELS, P. J. [13] DESCARTES, R. [14] DIAS, Paulo. [15] ELLIS, David. [16] EUGÊNIO, Marconi, [17] FOSKETT, D. J. [18] Fronteira, Nova [19] GARDNER, Howard. [20] Gerais, Minas [21] Graham, Taylor [22] Groce, Pacific [23] Horizonte, Belo [24] INGWERSEN, Peter. [25] JACOB, Elin, [26] JACOB, Elin, K. [27] Janeiro, Calunga, [28] Lafleur. Trad. L. J. [29] MARKMAN, Ellen M. [30] MEDIN, Douglas L. [31] MOREIRA, Walter. [32] MUSTAFA, Solange P. [33] NAVARRO, ESTEBAN [34] OGDENC. K. [35] ORRICO, Evelyn G. [36] Org. Vania R. [37] PEREZ, RUI C. [38] PEREZ, Ricardo O, [39] PIEDADE, M. [40] Paulo, S o [41] Próspero, Idméa Semeghini [42] Raghavan, K. S. [43] Ranganathan, S. R. [44] RICHARDS, I. [45] SARACEVIC, Tefko. [46] SEARLE, John. [47] SHAW, Debora. [48] SMITH, Edward E. [49] VALDEZ, Mark H,
Corporate Names	[1] Federal University of Minas Gerais [2] Georgia Institute [3] Harvard University [4] Havard University [5] Massachusetts Institute of Technology [6] Massachusetts Technology [7] School of Information Science [8] University of Chicago [9] University of Madras [10] University of Maryland

Keywords and phrases	Information processing 10 Information retrieval 22 Information science 17 Information 143 Processing 13 Knowledge 28

Table 1 – Extracted Proper Names Keywords and Phrases

References

1. Chen, Hsinchun. "Medical Text Mining: a DLI-2 status report". In *the 4th International Conference of Asian Digital Libraries (ICADL)*, ed. by Shalini.R.Urs, T.B.Rajashekar and K.S.Raghavan. ICADL 2001, 2001. p. 41-61.
2. Mukhopdhyay, Bikash and Mukhopadhyay, Sripati (2003). "Text mining techniques for analyzing news from Internet and Newspaper", In Library Progess (International). Vol.23 (No.1) 2003; p.35-41
3. Mukhopdhyay, B. (2004) "The Role of Text mining in the document warehousing", In SIS 2004, Digital Information Exchange: pathways to build global information society, papers presented at the 22nd Annual Convention and Conference IIT Madras, Chennai, 22-23 January, 2004. p.96-101
4. Neelameghan, A. (2003). "Content Summarization and Indexing of Texts: use of text mining software", In Information Management: trends and issues. p. 203-210.
5. Panigrahi, Pijushkanti, Prasad, A. R. D. and Basu, A. (2003). "NLP based Automatic classification system for Analytico-synthetic Scheme", *In* SRELS Journal of Information Management. Vol. 40, No. 4, December 2003, p 289-312.
6. Poiheau, Thierry and Kosseim, Leila (2001) Proper Name Extraction from Non-Journalistic Texts www.iro.umontreal.ca/~kosseim/publications/clin.pdf.
7. Ravin, Yael (2001) Extracting Names from Natural Language Text: www.research.ibm.com/talent/documents/20338.pdf.
8. Riaz, Muhammed (1989), Advanced indexing & abstracting practices. New
9. Delhi: Atlantic Publishers & Distributors: 255 – 257.
10. Takahiro (2001) Proper Name Extraction from English News Paper Articles: www.ipsj.or.jp/members/SIGNotes/Eng/14/1996/043/articles001.htm
11. Tkach, Daniel (1998). "Text mining Technology turning information into Knowledge: a white paper form IBM", The Forrester Report, 1995.
12. Wu, Tianhao (2002), Theory and Applications in Information Extraction from unstructured Texts.

Metadata Management &Its Impact on Digital Libraries

Rakhie Joseph

Abstract

The ever-increasing demand for systems to deliver flexible solutions to meet business requirements requires business and IT professionals to proactively share information, adjusting to changing needs in a cost-effective manner. In an attempt to gain a competitive advantage, today's businesses need to gather vital data from new and existing systems to effectively make essential strategic decisions. There are a number of ways for organizations to effectively manage these demands. First, IT assets must be cataloged, documented, and cross-referenced throughout the enterprise. Second, specific IT assets must be related back to the business processes they support in order for IT professionals to efficiently react to continually evolving requirements. Finally, this vital information must be accessible to all members of the organization, providing a unified view of the associations between business and technical assets while stimulating collaboration.In the past, a software component (.exe or .dll) written in one language could not easily use a software component written in another language. Metadata makes component interoperation even easier in the .NET Framework. It also provides the following major benefits in a .Net Environment like Self-describing files, Language interoperability and easier component-based design, Attributes. **Metadata** reduces development, deployment and maintenance costs by helping manage the entire process and associated resources. **Metadata management** helps you understand what information you have, where it is located, and what value it provides to users. Users can view information in a context they understand, providing a more efficient and intuitive way to communicate. Developers can share and reuse existing objects such as data structures, programs, model definitions and more. Metadata management ensures data consistency, enabling better business decisions and promoting more effective communication. Maintaining a library of digital objects of necessity requires maintaining metadata about those objects. The metadata necessary for successful management and use of digital objects, is both more extensive than and different, from the metadata used for managing collections of printed works and other physical materials. Metadata is widely used in library science, computer science, meteorology, geology, Electronics, government, and other domains, for scientific, industrial and commercial purposes. Metadata management provides business value, in all stages of data integration and warehousing projects, including design, development, deployment and management. Metadata ensures that data is only rationalized once and is then made available to all subsequent projects needing access to the data, thereby saving considerable analysis time.

Faculty in the Dept. of I.T ,Alliance Business Academy,Bangalore –560 068.
Email: rakhiejoseph@yahoo.com

In the Internet era, digital libraries represent completely new information infrastructures and knowledge environments. Integrating and utilizing the newest computer and communication technologies and digital content, the digital library builds huge extendable and interoperable collections. In order to manage the massive scale of these digital collections effectively, establishing a metadata model and application profile has become a fundamental part of any digital library project.

In conclusion, Metadata management would Reduce Development Time and Improve Quality, Reduce Maintenance Time, Improve Support Response Time and Improve User Access to Information. In other words, Metadata management would streamline the application development process.

1. Introduction

A major component of Information Management is Metadata management, and Metadata plays an inevitable role in Digital Libraries. Metadata is structured information that describes resources. While the resources are interesting to the end user, the metadata is helpful to the people or programs that have to manage the information. Catalogue records for library materials are a common example of metadata. Usually the metadata describes the content, physical description, location, type and form of the information, and information necessary for management including migration history, expiry dates, security, authentication, file formats and relationship with other versions.

2. Metadata

Metadata is not a new concept; it has existed in the computer science field for decades, and refers to information about electronic computer files. To update the concept a bit, the term "metadata" is now used to refer to information about any digital object that exists on the Internet. The need for certain types of data (such as creation date, file size, etc.) might seem obvious if one is managing a large group of digital objects merely as files. However, the Internet and World Wide Web offer great promise in terms of precision management, discovery and retrieval of digital objects such as images, e-texts, multimedia presentations, and other electronic files. Metadata may manifest itself either as an embedded, integral part of the digital object, to be retrieved and manipulated for various purposes, or it may exist externally from the digital object. Metadata often is broken into three broad categories:

- **Descriptive metadata:** Information that conveys some sense of intellectual content and context.
- **Structural Metadata:** Information that describes the attributes of an object, such as size, electronic format, and digital capture process.
- **Administrative metadata:** Information regarding rights management, creation date of the digital resource, hardware configuration, etc.

A "descriptive" metadata record consists of a set of elements, such as title, creator, format, date of creation, and subject coverage, that are necessary for describing a particular resource. In general, only descriptive metadata is visible to the users of a system, who search and browse it to find and assess the value of items in the collection. Administrative

metadata is usually only used by those who maintain the collection, and structural metadata is generally used by the interface which compiles individual digital objects into more meaningful units (such a journal volumes) for the user.

Metadata can be stored within the resource it describes (e.g. file formats that support descriptive headers), separate from the resource (e.g. an external catalogue) or separate but linked to the resource (e.g. a file linked with the digital object in a repository structure).

An example of storing the metadata with the resource is a World Wide Web (WWW) page. The standard for WWW allows metadata about the page to be embedded within the page. This embedded metadata is not seen when the page is viewed. The advantage of embedding the metadata within the resource is the tight coupling between the metadata and the resource. Whenever the resource is copied or moved, the metadata goes with it. Whenever the resource is modified, the metadata may be modified as well. When the resource is deleted, the metadata goes away as well.

The alternative is to store the metadata and the resources on separate servers. Since the metadata is normally much smaller than a resource, this means that a single server can store a large amount of metadata. This makes searching more efficient: fewer servers need to be accessed to search the metadata. The disadvantage of separating the metadata and the resource is that the automatic linkage is lost. A resource could be copied, moved, deleted or modified without modifying the associated metadata.

3. Metadata Management

Metadata management ensures data consistency, enabling better business decisions. It provides a single place for all users to access accurate definitions of information assets. While metadata can be used to enforce naming standards on new assets, it can also be used to identify and document redundancy and inconsistencies in your existing assets. This enables a consistent taxonomy across users, promoting more effective communication and better business decisions.

Metadata management provides business value in all stages of data integration and warehousing projects, including design, development, deployment and management:

- **Understand your Existing Data Environment** - Metadata documents the location and nature of potential data sources, making it easy for developers to locate, understand, rationalize and use this data. It ensures that the right data is used and that you leverage all of your enterprise-wide assets including information from ERP, specialized and legacy systems. Metadata ensures that data is only rationalized once and is then made available to all subsequent projects needing access to the data, thereby saving considerable analysis time.
- **Reduce Development Time and Improve Quality** - Developers are provided with definitions of available source and target data, transformation rules and code values that can be used/reused in the mapping process, ensuring the consistency and accuracy of delivered information. As you scale your information stores to meet business needs, this documentation serves as a blueprint, dramatically decreasing your development cycle and associated costs. When future iterations of information stores need to be delivered, metadata management accounts for existing systems and creates a map to the new systems.

- **Reduce Maintenance Time** - Developers easily identify impacts of changes in an application or a data structure on data warehouses and their users. This significantly reduces the analysis time in the early stages of a project.
- **Improve Support Response Time** - Support staff can quickly respond to users' questions on the meaning of warehouse data and rapidly trace its origins. They can quickly understand all factors that may lead to questionable data values. They can also store their own documentation.
- **Improve User Access to Information** - Business users are provided with a web-based directory of warehouse information in business terms. Metadata management helps you understand what information you have, where it is located, and what value it provides to users. Users can view information in a context they understand, providing a more efficient and intuitive way to communicate.

Metadata management also streamlines the application development process. The key to success is getting to market quickly. Metadata reduces development, deployment and maintenance costs by helping manage the entire process and associated resources. Metadata provides a single source for logical, physical and process aspects of the application environment, while tracking versions of the code and documenting all aspects of the application development life cycle. By providing a complete, integrated view of the development environment, metadata helps identify redundant processes and applications, thereby reducing duplicated efforts. Developers can share and reuse existing objects such as data structures, programs, model definitions and more. In addition, enterprise impact analysis greatly reduces the analysis and maintenance phase of the development life cycle.

4. Metadata in the .Net Environment

Metadata is the key to a simpler programming model, eliminating the need for Interface Definition Language (IDL) files, header files, or any external method of component reference. Metadata allows .NET languages to describe themselves automatically in a language-neutral manner, unseen by both the developer and the user. Additionally, metadata is extensible through the use of attributes. Metadata provides the following major benefits:

- Self-describing files.

  Common language runtime modules and assemblies are self-describing. A module's metadata contains everything needed to interact with another module. Metadata automatically provides the functionality of IDL in COM, allowing you to use one file for both definition and implementation. Runtime modules and assemblies do not even require registration with the operating system. As a result, the descriptions used by the runtime always reflect the actual code in your compiled file, which increases application reliability.

- Language interoperability and easier component-based design.

  Metadata provides all the information required about compiled code for you to inherit a class from a PE file written in a different language. You can create an instance of any class written in any managed language (any language that targets the common language runtime) without worrying about explicit marshaling or using custom interoperability code.

- Attributes.

The .NET Framework allows you to declare specific kinds of metadata, called attributes, in your compiled file. Attributes can be found throughout the .NET Framework and are used to control in more detail how your program behaves at run time. Additionally, you can emit your own custom metadata into .NET Framework files through user-defined custom attributes

5. Impact on Digital Libraries

The Challenge of Archiving Digital Information

The question of preserving or archiving digital information is not a new one and has been explored at a variety of levels over the last five decades. Archivists responsible for governmental and corporate records have been acutely aware of the difficulties entailed in trying to ensure that digital information survives for future generations. Far more than their library colleagues, who have continued to collect and organize published materials primarily in paper form, archivists have observed the materials for which they are responsible shift rapidly from paper objects produced on typewriters and other analog devices to include files created in word processor, spreadsheet and many other digital forms

Technological Obsolescence

Early attention to the difficulties in preserving digital information focused on the longevity of the physical media on which the information is stored. Even under the best storage conditions, however, digital media can be fragile and have limited shelf life. Moreover, new devices, processes and software are replacing the products and methods used to record, store, and retrieve digital information on breathtaking cycles of 2- to 5- years. Given such rates of technological change, even the most fragile media may well outlive the continued availability of readers for those media. Efforts to preserve physical media thus provide only a short-term, partial solution to the general problem of preserving digital information. Indeed, technological obsolescence represents a far greater threat to information in digital form than the inherent physical fragility of many digital media.

Digital information today is produced in highly varying degrees of dependence on particular hardware and software. Moreover, it is costly and difficult for vendors to assure that their products are either "backwardly compatible" with previous versions or that they can interoperate with competing products. Refreshing thus cannot serve as a general solution for preserving digital information and this conclusion has prompted discussion of other kinds of solutions.

6. Migration of Digital Information

Refreshing digital information by copying it from medium to medium and the possibility of maintaining a complex set of emulators describe two distinct points on a continuum of approaches to preserving digital information. However, neither refreshing nor emulation sufficiently describes the full range of options needed and available for digital preservation. Instead, a better and more general concept to describe these options is migration.

Migration is the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Migration includes refreshing as a means of digital

preservation but differs from it in the sense that it is not always possible to make an exact digital copy or replica of a database or other information object as hardware and software change and still maintain the compatibility of the object with the new generation of technology.

7. Legal and Organizational Issues

Compounding the technical challenges of migrating digital information is the problem of managing the process in a legal and organizational environment that is in flux as it moves to accommodate rapidly changing digital technologies. Consider, for example, the barriers to decisive preservation action caused by widespread uncertainty about legal and organizational requirements for managing the intellectual property that digital information represents. Addressing and resolving the legal and practical questions of migrating intellectual property in digital form necessarily involves a complex set of interested parties, including the creators and owners of intellectual property, managers of digital archives, representatives of the public interest, and actual and potential users of intellectual property. In addition, the parties who represent, for example, owners and users of different kinds of intellectual property (e.g., text and other document-like objects, photographs, film, software, multimedia objects) each operate under very different organizational regimes, with different experiences and expectations.

The costs and the technical, legal and organizational complexities of moving digital information forward into the future raise our greatest fear about the life of information in the digital future: namely, that owners or custodians who can no longer bear the expense and difficulty will deliberately or inadvertently, through a simple failure to act, destroy the objects without regard for future use. Repeated anecdotes about the loss of land use information, satellite imagery or census data, even when false or misleading, feed our general anxiety about the future of the cultural record we are accumulating in digital form. And uncertainty and lack of confidence about our will and ability to carry digital information forward into the future exert a major inhibiting force in our disposition to fully exploit the digital medium to generate, publish and disseminate information.

8. The Need for Deep Infrastructure

Even after more than forty years of growth, the digital world of information technology and communication is still relatively young and immature in relation to the larger information universe, parts of which have been under development for centuries. Our experiences of expense, intricacy and error in digital preservation surely reflect, at least in part, our inexperience with this emerging world as we operate in its early stages. Viewed developmentally, the problem of preserving digital information for the future is not only, or even primarily, a problem of fine tuning a narrow set of technical variables. It is not a clearly defined problem like preserving the embrittled books that are self-destructing from the acid in the paper on which they were printed. Rather, it is a grander problem of organizing ourselves over time and as a society to maneuver effectively in a digital landscape. It is a problem of building -- almost from scratch -- the various systematic supports, or deep infrastructure, that will enable us to tame anxieties and move our cultural records naturally and confidently into the future.

For digital preservation, the organizational effort -- the process of building deep infrastructure -- necessarily involves multiple, interrelated factors, many of which are either unknown or poorly defined. One of the biggest unknowns is the full impact on traditional information handling functions of distributed computing over electronic

networks. The effort to meet the cultural imperative of digital preservation thus requires a complex iteration and reiteration of exploration, development and solution as the relevant factors and their interrelationships emerge and become clearer and more tractable. And the first task in the effort is not to posit answers, but to frame questions and issues in such a way as to engage the many parties already working in various ways with digital information so that they can help us understand the relevant issues and, within the context of their work, help us identify, define and incorporate solutions that contribute to the larger, common goal of preserving our cultural heritage.

9. Information Objects in the Digital Landscape

Information objects in digital form, like those in other forms, move through life cycles. They are created, edited, described and indexed, disseminated, acquired, used, annotated, revised, re-created, modified and retained for future use or destroyed by a complex, interwoven community of creators and other owners, disseminators, value-added services, and institutional and individual users. The digital world is still too new for us to describe fully the life cycle of the information objects that do now or will in the future reside there, but what surely unites the community of actors in their various information-based activities is their common purpose in support of the pursuit of knowledge.
The pursuit of knowledge is a process in which the emergence of new knowledge builds on and reconstructs the old. Knowledge cannot advance without consistent and reliable access to information sources, past and present. It is the archival function in the system of knowledge creation and use that serves to identify and retain important sources of information and to ensure continuing access to them. How reliable the archival process proves to be in the emerging digital environment hinges on the trustworthy operation of digital archives and on their ability to maintain the integrity of the objects they are charged to preserve

10. Metadata and the Warehouse

Managing metadata is actually a complex endeavor that can cause severe problems if you do not handle it in the right way. Focusing on metadata can have substantial rewards for companies, as long as they approach it in a comprehensive manner. Let us analyze why this problem is so complex. First, consider what metadata is. In simple terms, metadata is the information about the data stored in databases, and it incorporates the rules about how to use that information.

Most large organizations have a common problem: They have multiple data sources. Each of these data sources has its own set of predefined rules, naming conventions, and unique file formats. Even when dealing with a single data source, there are complexities inherent in metadata. In some cases, there were as many as 10 or 12 different representations of the same information. Because each version had slightly different sources, processing rules, and formatting, it was almost impossible for the user or administrator to know which data source to use under different circumstances.

As you increase the number of data sources, the complexities involved with a data warehouse increase exponentially. This complexity becomes the problem and concern of those charged with building and managing the data warehouse. Organizations now must take information from several different data sources and push them together to create the illusion of a centralized corporate data source. They can take several steps to accomplish this.

First, an organization must understand what problem it is trying to solve with a data warehouse project. If it is too broad (that is, if the warehouse is being designed to solve too many problems at once), the warehouse will be too complex to implement. If it is too narrow (that is, too focused on a subset of a problem), it may not help the users.

Once you understand the problem, you must then understand the differences between how a particular piece of business information is represented and stored in multiple sources. In most cases, the metadata only represents storage information. Metadata should include information about the business rules, the source of that data, and how that data is being used. Without understanding all of this, it is impossible to bring these pieces of metadata together into a single data source. Therefore, you must first document the business metadata associated with the existing data sources. Once your organization understands the rules, it must agree on a common representation of the data.

Importantly, you must base your data representation on how the data is going to be used. For example, a corporate accounting database might have a large amount of financial data calculated to very precise detail. For purposes of a warehouse, analysis at the more macro level (rounded to the nearest hundreds of dollars) might be sufficient for analysis. The usage should be documented so that users of the warehouse understand how this data may differ from formats and meanings with which they are familiar.

Now you must define the data model for the target database. This can be particularly complex if you are integrating a number of different legacy data models. It is common to integrate relational, hierarchical, and flat files into a single model, but it is also difficult because you must create a data model that reflects the business from the highest level. If you are gathering information from a local data source, such as from a departmental server, it must make sense in the context of this enterprise view.

The next step is the actual data transformation. In this stage, data must be mapped from one source to another. If some data comes from a hierarchical mainframe source, and other data comes from relational databases, you must put them into a common format for storage in the warehouse. At the same time, you must take the format of the warehouse into consideration. Needless to say, this mapping process is complex. Moving data is a two-step process: the initial loading or seeding of the warehouse, followed by the incremental updating from the source database.

During the initial loading process, it is necessary to establish a metadata catalog or dictionary. In many cases, corporations find that they must hire programmers to perform the initial load because so many variables are involved, including platform, database, applications, and so on. The complexity increases as the volume of data expands. If you're talking about terabytes of data, you could have a long night.

11. Metadata Use in Medicine

Metadata is used to organize and process data and information for increased usability and interoperability among different organizations and their information systems.

In medical care, descriptive metadata (in the form of classifications and codes) define demographics, diagnoses and care of patients for the purposes of documentation, communication, transaction and monitoring. Such metadata is created and used within the medical record by care providers, administrators, insurers and regulatory agencies to communicate and document care and to keep track of transactions, payments and

operations. Examples of clinical classifications and codes used include: the International Statistical Classification of Diseases and Related Health Problems (ICD-10), Current Procedural Terminology (CPT) and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT).

In health care transactions, structural and administrative metadata (in the form of messaging standards) define the formats in which coded data (in textual and non-textual formats (such as imaging data) is created, transmitted and stored for use and archiving. Such metadata specifies the ways that data and information are exchanged between electronic systems) for consistent processing (interoperability). Examples are the Health Level Seven Reference Information Model (HL7 RIM) and Digital Imaging and Communications in Medicine (DICOM).

In medical publishing and librarianship, descriptive, structural and administrative metadata are used extensively to archive and index publications (journal articles, books, electronic media, etc.) for identification and retrieval. Examples of descriptive metadata in this domain include: the Medical Subject Headings (MeSH) and the National Library of Medicine (NLM) Classification System. Examples of structural metadata include: eXtensible Markup Language (XML) for electronic business (eb XML). Examples of administrative metadata include descriptions of the different formats used for storage of multimedia.

11. Metadata Tools

Numerous software tools exist for the encoding of metadata, ranging from freeware packages to highly complex and expensive integrated systems. Those designed specifically for XML encoding range from free packages such as emacs to proprietary packages such as XMetal.

12. Metadata Standards in General

There are many different standards governing metadata:

- standards specific to topics or disciplines (such as biology or art)
- standards specific to kinds of materials (such as moving pictures or encoded texts)
- standards to support particular functions (such as discovery or rights management or presentation)

In any of these areas, metadata standards may govern

- what pieces of information are created (semantics)
- how the information is formed (content standards)
- how the information is encoded for computer processing (syntax)

Metadata design is a critical part of the planning for any digital project. Without the right kind of metadata, it will not be possible to find or use digital materials effectively. The LDI Metadata Advisor can help you determine what kinds of metadata you need to get the results you want.

13. Multimedia Metadata Standards

Metadata is an important aspect of the creation and management of digital images (and other multimedia files). Metadata standards for digital imaging can include information about:

- the technical format of the image file
- the process by which the image was created
- the content of the image

14. Conclusion

Although the concept of metadata predates the Internet and the Web, worldwide interest in metadata standards and practices has exploded with the increase in electronic publishing and digital libraries, and the concomitant "information overload" resulting from vast quantities of undifferentiated digital data available online. Anyone who has attempted to find information online using one of today's popular Web search services has likely experienced the frustration of retrieving hundreds, if not thousands, of "hits" with limited ability to refine or make a more precise search. It is often more effective to search through metadata than through the resources. Metadata gives information needed by seekers and managers of information, such as subject keywords, abstract, date created, intended audience, and many others, information that is not always available in the resource itself.
In conclusion, Metadata management would Reduce Development Time and Improve Quality, Reduce Maintenance Time, Improve Support Response Time and Improve User Access to Information. In other words, Metadata management streamlines the application development process and plays a major role in Information management.

Reference sites
http://www.dlib.org
http://msdn.microsoft.com/library
http://www.odl.ox.ac.uk/metadata.htm
http://www.nla.gov.au/padi/topics
http://www.ercim.org/publication
http://www.dag.com/metacenter/overview.asp
http://dbpubs.stanford.edu:8091/diglib/pub/delos.html
http://www.ontoportal.org.uk/snapshot/explore/themes
http://www.e-government.govt.nz/docs/nzglsv2

Articles
- The Technology Behind the New Geodata.gov and the Non-Technology Challenges Ahead - By Adena Schutzberg and Joe Francica
- Pro metadata will lose to folksonomy - Clay Shirky

Online Books
- METADATA MADE SIMPLER by Gail Hodge
- Comprehensive Metadata Management - Whitemarsh Information Systems Corporation

Metadata Management System for Digital Resources

Akhtar Hussain and Javed Asim

Abstract

This paper indicates the concept of metadata. Defined simply, it is "data about data". There are three kinds of metadata associated with digital objects: Descriptive or content, Structural and Administrative metadata. It can generally be viewed as being of various types: Dublin Core, MARC, Global Information Locator service etc. The paper also discusses, how the metadata is structured. Finally, it indicates basic metadata components for data, creation, elements, protocal and a guide for Libraries.

1. Introduction

Metadata is a description of an information resource, and hence can be thought of as 'data about data ;within the context of the world wide web metadata may be used for information discovery ,but metadata is also important in the context of cataloguing resource within the advent of the Dublin core the text encoding initiative (TEI) guidelines the Government Information locator Service (GILS), and other such constructs, the concept of data about data and what to do with it has again become a discussion point in librarianship. As the methods available for describing information grow beyond MARC, it becomes increasingly apparent that librarians, and more specifically catalogers, have a role to play as mediators and creates of an increasingly diverse landscape of descriptive methods. As the choices for providing access increase, the experience and traditions that the cataloging profession can bring to the creation, standardization, and manipulation of metadata systems becomes obvious.

By mediating the use of metadata, cataloguers provide for these developing systems of description and access a strong influence towards standardization. Cataloging exists as an often-invisible process of order making. As constituencies develop systems for ordering in the digital world, what was once a largely invisible process becomes glaringly apparent. Granularity is the level at which metadata is applied to an object or set of object. The expense of creating metadata is directly proportional to the degree of granularity. In depth metadata creation requires more effort, which in turn costs more. For example in a traditional library the entire set of volumes for any journal can be described by one entry. This represents 'coarse' level of granularity. But in case of e-journals, links are provided for all volumes, indexes, individual article, etc. which is a case of 'fine' level of granularity.

2. What is Metadata?

In simple way "Metadata" is a data about data, but basically it is a "structured data about data". It gives information about the data, which are stored on web. Each and every page of any website concerns with the metadata. Those metadata keeps information about the page on which the page talks

Department of Library & Information Science, CCS, University, Meerut (U P).
E-mail: ahussain_47@yahoomail.co.in

3. Definition of Metadata

www.terralink.co.nz/profile/glossary/

Metadata is data about data. Used in the context of digital spatial data, metadata is the background information, which describes the content, quality condition, and other appropriate characteristics of the data.

3.1 Definitions of Metadata

www.nyskwic.org/u_data_terms.cfm

"Metadata provides information about the content, quality, condition, and other characteristics of data."

www.savi.org/savil/documentation/glossary.aspx

(2.2) "Metadata is the term used to described data about data. It describes who collects the data, what the data contains, where (and low) the data is stored, when (and how often) the data is collected, and why."

We can say that information about a data set which is provided by the data supplier or the generating algorithm and which provides a description of the content, format and utility of the data set.

4. Types of Metadata

Metadata generally contains a pointer to the location of the object. Just like in a classification scheme such a Dewey Decimal Classification (DDC) indicates of represents a place on a library shelf, a URL represents the location of a digital object. With digital objects, the metadata can be embedded and then automatically extracted, though the level of extraction depends upon the quality of the metadata. There are three kinds of metadata associated with digital objects:

- Descriptive metadata or content metadata
- Structural metadata
- Administrative metadata

4.1 Descriptive Metadata

Descriptive metadata describes the attributes of digital object such as 'title', 'creative', 'subject', 'date', 'keywords', 'abstract' etc. The description of non-textual data in text is problematic, as it is difficult to know all the possible intended uses for metadata and skilled creators are needed for this. Descriptive metadata can be applied at a fine level of granularity, with descriptions of even one image running into hundreds of words. Application of metadata to digital objects needs to be rigorously controlled with use of thesauri, name authority files, standard subject headings like Library of Congress Subject Heading (LCSH).

4.2 Structured Metadata

Structural metadata attempts to describe the structure and relationships of a set of digital objects. Structural metadata facilitate navigation and resource discovery, as the structure of a digital object is an important indicator of that object's meaning. In absence of good

structural metadata, it is impossible to manage digital objects in a reposting and guide user interaction with these. But as in the case of descriptive metadata, structured metadata creation is also a costly affairs if done manually. The European Union-funded METAe Project is developing automatic process for this as on economical way out.

4.3 Administrative Metadata

Administrative metadata records information about the creation of the digital object such as: initial capture settings, scanner parameters, file formats, programme used to create it, compression technology date of creation, version, etc. It also records information about the legal and financial aspects of access to the object like rights management, payments, costs, and authentication. It encompasses any information that might be of use of future caretakers of that object and it records all the events that happen to it during the entire lifecycle. Administrative metadata may reside within or outside the digital objects and includes all the information needed for presentation to ensure 'digital longevity'.

5. Types of Metadata Standards

Metadata can generally be viewed as being of various types:

- Dublin Core
- Machine Readable Catalogue (MARC)
- Global (Government) Information Locator Service (GILS)
- The Text Encoding Initiative (TEI)
- The Encoded Archival Description (EAD**)**

5.1 Dublin Core:

The Dublin Core Metadata Element Set arose from discussions at a 1995 workshop sponsored by OCLC and the national Centre for Supercomputing Applications (NCSA). As the workshop was held in Dublin, Ohio, the element set was named the Dublin Core. The continuing development of the Dublin Core and related specifications is managed by the Dublin Core Metadata Initiative (DCMI).

Core was to define a set of elements that could be used by authors to describe their own Web resources. Faced with a proliferation of electronic resources and the inability of the library profession to catalogue all these resources, the goal was to define a few elements and some simple rules that could be applied by noncatalogers.

The original 13 core elements were later increased to 15 : title, subject, description, source, language, relation, coverage, creator, publisher, contributor, rights, date, type, format, and identifier. The Dublin Core was developed to be simple and concis, and to describe Web-based documents.

However, Dublin Core has used with other types of materials and in applications demanding some complexity. There has historically been some tension between supporters of a "minimalist" view, who emphasize the need to keep the elements to a minimum and the semantics and syntax simple, and supporters of a "structuralist" view who argue for finer semantic distinctions and more extensibility for particular communities.

All Dublin Core elements are optional and all are repeatable. The elements may be presented in any order. While the Dublin Core description recommends the use of controlled values for fields where they are appropriate, this is not required. However, working groups have been established to discuss authoritative lists for certain elements such as Resource type.

While Dublin Core leaves content rules to the particular implementation, the DCMI encourages the adoption of application profiles (domain-specific rules) for particular domains such as education and government.

There are hundreds of projects worldwide that use the Dublin Core either for cataloguing or to collect data from the Internet; more than 50 of these have links on the DCMI website. The subjects range from cultural heritage and art to mathematics and physics.

Meanwhile the Dublin Core Metadata Initiative has expanded beyond simple maintaining the Dublin Core Metadata Element Set into an organization that describes itself as "dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems."

5.2 Machine Readable Catalogue (MARC*)*

Standardized bibliographic data input, utilizing MARC, insures the integrity of the online public catalogue in storage and retrieval of information. When we talk about MARC, we hope to convey its importance to those who create and maintain MARC data in the online public catalogue. Without good accurate MARC records, patrons cannot find the great resources in the library. An OPAC, to some extent, hides the intricacies of MARC from the patron, but without MARC, the patron would not find the resources. MARC format cataloging has proven, for over the years, to be the most reliable foundation in building databases for the OPAC.

Each MARC format provides detailed field descriptions and guidelines for applying the defined content designation and identifies conventions to be used to insure input consistency. A MARC record is composed of three elements: the **record structure** the **content designation,** and the **data content of the record.**

MARC 21 is the new name of the harmonized CAN/MARC and USMARC formats. From 1994-97, the Library of Congress and the National Library of Canada worked with their user communities, through their MARC committees, to reconcile format differences. In Canada, the Canadian Committee on MARC represents the user community on MARC format issues. MARC 21 is result of the activity undertaken to align CAN/MARC and USNARC. It represents the continuation of the CAN/MARC and USMARC formats in a single edition with a new name. In January 1998, the National Library of Canada issues and update to the CAN/MARC bibliographic format, which included all of the changes resulting from harmonization with USMARC. The principles for content designation in the MARC 21 formats was approved by the American Library Association's ALCTS/LITA/RUSA Machine-Readable Bibliographic Information Committee (MARBI), in consultation with representatives from Unites States and Canadian national libraries and designed bibliographic networks. The statements includes the principles under which the MARC 21 formats were developed and constitutes a set of working principles for the ongoing process of format development.

A MARC 21 format is a set of codes and content and content designators defined for encoding machine-readable records. Formats are defined for five types of data: bibliographic, holdings, authority, classification, and community information. These are widely used standards for the representation and exchange of bibliographic, authority, holding, classification, and community information data in machine-readable form. The

MARC 21 formats are an implementation of the *Information Interchanges Format* (ANSIZ39.2). The formats also incorporate other relevant ANSI standards. All information in a MARC records is stored in character form. MARC communications records are coded in Extended ASCII, as defined in the *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media.*

5.3 Global (Government) Information Locator Service (GILS)

GILS is a U.S. Federal Information Processing Standard (FIPS Pub 192) supported by various guidelines and memoranda from the Office of Management and Budget. GILS grew out of the U.S. government requirement for public access to government information, and it is authorized by the Paperwork Reduction Act of 1995.

Originally called the "Government Information Locator Service", GILS in various forms has been adopted by other governments and for international projects, leading to its current designation, "Global Information Locator Service".

GILLS itself does not formally define metadata elements, rules for representation, and syntax. Rather, GILS specifies a profile of the Z39.50 protocol for search and retrieval, specifying which attributes must be supported.

The original goal of GILS was to provide high-level locator records for government resources, both electronic and nanoelectronic. GILS records were intended to describe aggregated such as catalogues, publishing services and databases. The emphasis is on availability and distribution rather than on description.

Therefore, a GILS record may have data elements such as the name and address of the distributor and the order process. However, some organizations use GILS at the individual item (journal article or technical report) level. Since GILS was an early metadata scheme, evaluations of its implementation and use are available and very valuable in developing other metadata systems.

5.4 The Text Encoding Initiative (TEI)

The Text Encoding Initiative is an international project to develop guidelines for marking up electronic texts such as novels, plays, and poetry, primarily to support research in the humanities. This SGML markup becomes part of the electronic resource itself. In additional to specifying how to encode the text of a work, the TEI Guidelines also specify a header portion, embedded in the resource that consists of metadata about the work.

The TEI header, like the rest of the TEI, is defined as an SGML DTD, a set of tags and rules defined in SGML syntax that describe the structure and elements of type of document. Since the TEI DTD is rather large and complicated in order to apply to a vast range of texts and uses, a simpler subset of the DTD, known as "TEI Lite", is commonly used in libraries.

It is assumed that TEI-encoded texts are electronic versions of printed texts. Therefore the TEI Header can be used to record bibliographic information about both the electronic version of the text and about the non electronic source version. The basic bibliographic information is not dissimilar to that recorded in library cataloguing and can be mapped to and from MARC.

However, there are also elements defined to record details about how the text was transcribed and edited, how markup was performed, what revisions were made, and other non-bibliographic facts. Libraries tend to use TEI headers when they have collections of SGML-encoded full text.

Some libraries use TEI headers to derive MARC records for their catalogue systems, while other use MARC records for the published source texts as the basis for creating TEI header descriptions.

5.5 The Encoded Archival Description (EAD)

In archives and special collections, the finding aid as an important tool for resource description. Finding aids differ from catalogue records by being much longer, more narrative and explanatory, and highly structured in a hierarchical fashion.

They generally start with a description of the collection as a whole, indicating what types of materials it contains and why they are important. If the collection consists of the personal papers of an individual there can be a lengthy biography of that person. The finding aid describes the series into which the collection is organized, such as correspondence, business records, personal papers, and campaign speeches, and ends with an itemization of the contents of the physical boxes and folders comprising the collection.

The Encoded Archival Description (EAD) was developed as a way of marking up the data contained in a finding aid, so that finding aids can be searched and displayed online. The EAD standard is maintained jointly by the Library of Congress and the Society of American Archivists. Like the TEI Header, the EAD is defined as on SGML DTD. I begins with a header section that describes the finding aid itself which could be considered metadata about the metadata; it then goes on to the description of the collection as a whole and successively more detailed information.

If individual items being described exist in digital form, the EAD can include pointers to the digital objects. The EAD is particularly popular in academic libraries with large special collections. Although it is easier to put finding aids on the Web by simply marking them up in HTML than to follow the EAD specification, libraries and archives investing in EAD creation hope that using this metadata scheme will encourage consistency in encoding and give them some measure of search interoperability.

6. How is the meta-data structured?

Structured? And you thought meta-data was just a list of descriptive terms. What complexity rears its ug here? The information structure is sometimes is sometimes called the “data model”. There are two basic types of meta structures that I will call the “clothesline” and the “mobile” models.

Remember clotheslines? Those were cords that were strung up outside from which one hung wet clothes with clothespins. The clothesline model can be thought of as a clothesline with a bunch of wooden sprin clothespins. Each clothespin has a label written on it: Title, Creator, Subject, etc. Each clothespin can no single strip of paper with a data value on it. This is a set of names with values. These are called name-vapairs. Wow. Clever, huh? For example, the meta-data field called “Author” may contain in name of the

Author = Jane Doe.

The equals sign express the one-to-one nature of the name and the value. As Dublin Core meta-data us HTML “meta” tag this would be expressed:

<META NAME= “dc. Creator” CONTENT = “Jane Doe”>

Meta-data may also be structured. Consider a mobile such as those created by the late Calder. they hang single point and have a set of pieces that each might have yet other pieces suspended from them. An autl name can be expressed in a structure:

Name

1--First Name = Jane

1--Last Name = Doe

Or to get closer to reality, Author may be a person:

Author

l - - Person
l - -l - - Name
l - - l- - First Name =
l - - l - -l - - Last Name =
l - -l - - Address
l - -l - - -l - - Organization =
l - - l - -l - - Street1 =
l - -l - - l - - Street2 =
l - - l - -l - - City =
l - -l - - l - - State =
l - - l- - l - - Zipcode =
l - - l - - l - - Country =
l - - l- - Telephone
l - - l - - l - - Home

Familiar, isn't it?

We could also have multiple Authors, repeating Person:

Author

l - - Person

l - - Person

Structured meta-data can become quite elaborate. The resulting richness requires the existence and main of someplace describing the structure so that others can access it. This is often called a "registry". An our hierarchy, is one of the simplest structures. One could also have a single person filling several roles: author publisher, for instance. A "polyhierarchical" structure may result. So we have meta-data about the meta-data. This is particularly true in the geosciences world there are complex datasets.

A structure meta-data system may use the structure to create the definition of a field. Within the IMS at LOM meta-data is a structure called "Classification". This relates to a system for classifying a resource is manner. Part of the structure of Classification is:

Classification

l - - Purpose

l - - Description

l - - Keywords

"Purpose" is definition of the type of classification. For example, a Purpose of "Discipline" is equivalent "Subject". The term is the result of an international negotiation (as noted below), but the definition would understood by Americans under the rubric of "Subject". Therefore, the Description is a description of the area of the resource. The

Keywords relate solely to the subject are. In other words, the Description and Keywords must be interpreted with respect to the Purpose of the classification. The structural relationship meaning.

7. The Metadata Lifecycle Model (MLM)

The MLM is a methodology involving a ten-step process by which digital library projects can design and implement metadata provision. It can be reviewed again once the project requirements have been changed. The purpose of the model is:

- To achieve a consistent method of developing metadata for digital library projects, so that they might achieve greater effectiveness and better quality.
- To conduct a content-based analysis for digital collections, which takes into account knowledge mapping and metadata research needs while adopting any existing metadata formats.

The ten-step of the MLM can be classified into four main groups as follows:

7.1 Groups I Requirement Assessment & Content Analysis

- Step1: Acquisition of Metadata Base Needs
- Step:2 Review of Relevant Metadata Standards and Projects.
- Step:3 Investigation of Deep Metadata Needs.
- Step:4 Identification of Strategies for the Metadata Schemes and

7.2 Groups II System Requirement Specification

- Step5: Preparation of the Metadata Requirement Specification
- Step6: Evaluation of Metadata Systems

7.3 Groups III Metadata System

- Step7: Preparation of Best of Practice Guidance
- Step8: Development of the Metadata System

7.4 Group IV Service and Evaluation

- Step9: Maintenance of Metadata Performance
- Step10: Evaluation of Metadata Performance

The first step of the metadata lifecycle is to interview the content experts or provides about their metadata requirement for each collection project, and analyze the attributes of collection projects through the four-layer for metadata selection.

- Metadata scope: Attributes of metadata are relating to thing, person, event, temporal, vocabulary control, and geographic name.
- Legacy system: basic information about the legacy system, including elements, structure, number of records, storage format, and input method and system. In addition, it is useful to understand the advantages and disadvantages of the system.
- Metadata context: Is only one metadata database constructed for this project? Are any other databases required to integrate with this metadata database, such as geographic information system (GIS).

- Metadata role and function: What kind of metadata role is proposed for each project? What kind of function should be achieved by metadata, such as resources description, discovery, annotation, or administration? After the four-layer for metadata selection is applied, the candidates of metadata schemes for a digital library are thus revealed. Current metadata trends and issues from around the world related to the profile of the digital library project can be collected and analyzed as a reference for practical application and future development.

Through this stage the collection project members could well know what kind of differences exists with other similar or homogeneous collection projects, and also re-arrange the focus of expected project goals.

Metadata scope and context are clarified, and related relationship are clearly drawn and attributed to a diversity of categories. It could ensure what kind of systems and databases are integrated by metadata mechanisms such as GIS.

The purpose of developing metadata services is to guarantee the quality assurance of metadata mechanisms. A service model is constructed to formulate the service items. The assessment of metadata record quality, including completeness, accuracy, record types, including completeness, accuracy, record, types, granularity levels, and records' serviceability. For example, and evaluation of the U.S. Government Information Locator Service (GILS) focused on assessing metadata quality from a set of criteria and procedures.

The evaluation of the effectiveness of adopting a metadata scheme for retrieval, such as its capacity to facilitate access to title, creator, and subject. For many typical applications of metadata. However, the case study shows that there are more functions needed, such as preservation, content rating, rights, data management and e-Learning.

In addition, domain-specific metadata sets are the core development of metadata practices in the case study. In other words, the generic metadata set, such as Dublin Core for resources discovery, is deployed in its top level of the union catalogue.

8. Semantic Metadata

Semantic metadata is the assembly of assets based on explicit or implicit elements. Explicit assembles could include hierarchal structures, such as the ones found within search engines like Yahoo. Implicit assemblies could include inference engines that traverse the corporation looking for relevant assets. Keyword bases search engines provide a basic example of how this process might work in the future. Currently, a user types in a few keywords and the search engine return a set of documents that contain some or all the keywords. This functionality should be expanded to include any asset in the corporation and not just the ones document in a web page of document. In the future, agents will be able to traverse operating systems, XML constructs, interfaces, and other asset that can viewed by the computer system. Asset will then be able to be grouped by context, usage, time, and various other constructs.

9. Metadata Entry System

The metadata entry system is used to collect a sufficient amount of data about individual items to allow any item to be found during a search of the collection, and it performs a number of functions in addition to recording descriptive information about the items in the collection. The design of the system encourages correct, consistent, and standard collection of metadata.

It allows multiple persons to enter metadata simultaneously, providing a common interface for all persons entering data, and enforcing the notion of required fields. Whenever

possible, data are entered by choosing from enumerated lists. Data validation is performed by the system wherever appropriate, and warning messages are generated to alert the use to potential problems

Since most items require that permission be sought before they can be made available in digital form to the public, capabilities are provided in each metadata record for recording information about the status of copyright permissions, as well as any special restrictions imposed by donors. The metadata record also allows the archivist to enter information that relates to the intellectual organization of the collection.

As each item is logged into the metadata entry system, it is assigned a unique identifier. When the archivist enters the system, the next available, unique identifier for the collection being processed appears.

10. Basic Metadata Components for Data

Name	Description
Table Name	The name of the table
Database Name	The name of the parent database
Keywords	Specific keywords and phrases about the table
Description	Detailed description of tables utility
Data	Date the information was published
Schema	Related Sources Schema
Logical Model	Parent logical model
Owner	Owner of SME of table
Server	Physical location of table

11. Metadata Creation

Many participants identified metadata creation as a fundamental challenge. **Lou Rosenfeld** noted the longstanding problem of inter-indexer consistency, and cited evidence that only 10% of human-assigned index terms do not occur in full text. Considering these facts, he was skeptical about whether human-created subject metadata merits investment in many cases. **Sandy Hostetetter** reported that many end users in her organization aren't interested in creation metadata, leading to poor results when they are asked to do so, **Julie Martin** from Boeing expressed optimism about the ability of better metadata creation tools, such as a plugin for Microsoft Word her company has been developing internally. But she noted that even excellent tools are of little use until a "content culture" has developed, in which understanding of the value of metadata and its contribution to information retrieval are pervasive.

13. Dublin Core Elements

The user would enter the Dublin Core information about the item as appropriate: Title, Creator, Subject (keywords), Description, Publisher, Contrib- utors, Date, Type, Format, Identifier Source, Language, Relation, Coverage, and Rights.

Element Name: Title

Label:	Title
Definition :	A name given to the resource.
Comment :	Typically, Title will be a name by which the resource is formally known.

Element Name: Creator

Label:	Creator
Definition :	An entity primarily responsible for making the content of the resource.
Comment :	Examples of Creator include a person, an organization, or a service, Typically, the name of a Creator should be used to indicate the entity

Element Name: Subject

Label:	Subject and Keywords
Definition:	A topic of the content of the resource
Comment:	Typically, Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Element Name: Description

Label:	Description
Definition:	An account of the content of the resource
Comment:	Examples of Description include, but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

Element Name: Publisher

Label:	Publisher
Definition:	An entity responsible for making the resource available
Comment:	Examples of Publisher include a person, an organization, or a service, Typically, the name of a Publisher should be used to indicate the entity.

Element Name: Contributor

Label:	Contributor
Definition:	An entity responsible for making contributions to the content of the resource.
Comment:	Examples of contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.

Element Name: Date

Label:	Date
Definition :	A date of an event in the lifecycle of the resource.
Comment :	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 and includes (among others) dates of the form YYY-MM-DD.

Element Name: Type

Label:	Resource Type
Definition :	The nature or genre of the content of the resource.
Comment :	Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary. To describe the physical or digital manifestation of the resource, use the FORMAT element.

Element Name: Format

Label:	Format
Definition :	The physical or digital manifestation of the resource.
Comment :	Typically, format may include the media-type or dimensions of the resource. Format may be used to identify the software, hardware, or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types defining computer media formats).

Element Name: Identifier

Label:	Resource Identifier
Definition :	An unambiguous reference to the resource within a given context.
Comment :	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Formal identification systems include but are not limited to the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Element Name: Source

Label:	Source
Definition :	A Reference to a resource from which the present resource is derived.
Comment :	The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

Element Name: Language

Label:	Language
Definition :	A language of the intellectual content of the resource.
Comment :	Recommended best practice is to use RFC 3066, which in conjunction with ISO 639, defines two-and three-letter primary language tags with optional subtags. Examples include “en” or “eng” for English, “akk” for Akkadadian”, and “en-GB” for English used in the United Kingdom.

Element Name: Relation

Label:	Relation
Definition :	A reference to a related resource.
Comment :	Recommended best practice is to identify the referenced resource by means of a string or number conforming to a formal identification system.

Element Name: Coverage

Label:	Coverage
Definition:	The extent or scope of the content of the resource.
Comment:	Typically, Coverage will include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity). Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names (TGN) and to use, where appropriate, named places or time periods in preference to numeric identifiers such as sets of coordinates or date ranges.

Element Name: Rights

Label:	Rights Management
Definition :	Information about rights held in and over the resource.
Comment :	Typically, Rights will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. I f the Rights element is absent, no assumptions may be about any rights held in or over the resource.

14. Metadata Protocol

The Open Archive Initiative (OAI) for metadata harvesting is a vital protocol dedicated to solving problems of digital library interoperability by defining simple protocols, most recently for the exchange of metadata. Although Z39.50 Information Retrieval Protocol is also one the harvesting protocol.

Some digital library systems are using for retrieving and searching the records for example Maxwell System. It is crucial to explain, discuss, and disambiguate the concepts and terminology used among OAI implementers to harvesting approach to interoperability.

The harvesting protocol defined by the OAI is a request/ response protocol with 6 request types, viz. Identity, ListMatadataFormats, ListSets, GetReocrds List Record, and ListIdentifiers. It has defined in such a way to support a common set of principles and technical framework to achieve interoperability functionality wise the multingual digital library should have flexible with:

13. 1 Comfortable to the user

Digital libraries are created in countries and, in general may be somehow spcialized. It is expected that the users in the country where the library operates will want to use it in their native language. At the same time, users with other native languages than that of the country under consideration may need more international languages, as for example, English or Hindi.

13.2 Flexible in terms of the chosen languages

Digital libraries in different countries and aiming at different sets of users may operate with distinct sets of languages. As an example, in India a good set of languages of Hindi, Bengali, Gujarati, Marathi, Oriya, Sanskrit, Kannada, Telugu, Tamil and English etc. But in USA or in UK, it must be different.

13.3 Wide and accurate in terms of the information

In order to fulfill its functions, the digital library must be accurate in terms of information. This requires that the languages be kept track at any moment of operation by the use; he/she must know the languages of the catalogue entry and of the content, regardless the navigation language in use.

For the search to be effective, no matter the navigation language, the user may submit search arguments in any language and all points of access must be possible to be searched in all languages, regardless of the navigation language of the session. In order to identify contents in original languages and corresponding translations into other languages, a strict translation control must exist.

13. 4 Easy to operate

The digital librarian must control all cataloguing (original languages and translations) and all the contents (original languages and translations). For data integrity to be achieved, translation interfaces for the cataloguing and translation control applications must be available.

13. 5 Language control Parameters

Language of the content: it is one of the metadata in metadata schemes The language of the content is the language in which the content is written and / or spoken.

13.6 Language of the catalogue entry

The language of the catalogue is the language in which the attributes of the content and its instances are written.

13.7 Language of navigation

The language of navigation is the language of all the interfaces and messages of the digital library system.

14. Advantage of metadata

Metadata has proved to be a boon to the users as it gives good precision through, structured i.e., search and Boolean support. It allows users to determine if the data is what they want. It also helps in preventing some users (e.g., children) from accessing data. It is also beneficial to the site indexes mentioned earlier like the search engines, subject directories or information gateways, who need nog guess the subject or the content of the resource. The task of indexing becomes easier as the data they encounter is highly structured. It also enables them to filter the data, which is obsolete, and provide only current information. It would allow them aware of the features and contents according to various forms it is present it. Publishers of various sites also relay on metadata, which provides them with relevant details about the resource. It makes them aware of the features and contents present or absent in the resource. It also helps publicise and support data, which the organization has created.

15. Metadata: A Guide for Libraries

Metadata is often called data about data or information about information. The term metadata is used differently in different communities. Some use if to refer to machine understandable information, while others use it only for records that describe electronic resources. However, in the library environment, metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital.

Traditional library cataloguing is a form of metadata, and MARC 21 and the rule sets used with it such as AACR2 are metadata standards. Other metadata schemes have been developed to describe various types of textual and non-textual objects such as archival materials, visual materials, geographic information, and science and social science datasets.)

Just as cataloguers made decisions about whether a catalogue record should be created for a whole set of volumes or for each particular volume in the set, so the metadata creator makes similar decision. Metadata can also be used for description at any level of the information model laid out in the IFLA (International Federation of Library Associations and Institutions) Functional Requirement for Bibliographic Records: work, expression, manifestation, or item.

For example, a metadata record could describe a report, a particular edition of the report, or a specific copy of that edition of the report. Metadata can be embedded in a digital object or it can be stored separately. Metadata is often embedded in HTML documents and in the headers of image files. Storing metadata with the object it describes ensures the metadata will not be lost, obviates problems of linking between data and metadata, and helps ensure that the metadata and object will be updated together.

However, it is impossible to embed metadata in some types of objects (for example, artifacts). Also, storing metadata separately can simplify the management of the metadata itself and facilitate search and retrieval. Therefore metadata is commonly stored in database systems and linked to the objects described.

Metadata schemes (also called schema) are sets of metadata elements designed for a particular purpose, for example, to describe a particular type of information resource. The

definition or meaning of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content.

Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, how to identify the main title) and / or representation rules for how content must be represented (for example, capitalization rules). There may also be syntax rules for how the elements and their content should be encoded.

A metadata scheme with no prescribed syntax rules is called syntax independent. Metadata can be encoded in MARC, in "keyword = value" pairs, or in any other definable syntax. Many current metadata schemes use SGML or XML. MXL (Extensible Mark-up Language) is an extended form of HTML which allows for colcally defined tag sets and the easy exchange of structured information.)

SGML (Standard Generalized Mark-up Language) is a superset of both HTML and XML and allow for the richest mark-up of a document. An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, and support digital identification, and support archiving and preservation.

Metadata serves the same functions in resource discovery as good cataloguing does by:

- Allowing resources to be found by relevant criteria;
- Identifying resources;
- Bringing similar resources together;
- Distinguishing dissimilar resources;
- Giving location information.

As the number of Web-based resources grows exponentially, aggregate sites or portals are increasingly useful in organizing links to resources based on audience or topic. Such lists can be built as static web pages, with the names and locations of the resources "hardcoded" in the HTML.

16. Conclusion

The ability to manage the vast arrays of digital resources on the Internet is reality. By utilizing the metadata standards and protocols described in this article, providers of digital objects can create their own databases and participate in national efforts to manage digital objects for educational purposes. The past few years have seen major development in this area, and the next few years will see refinements in processes that will create more efficient and wider ranging improvements in access to digital library resources. By familiarizing ourselves with the standards and databases that have been created to manage digital library material, librarians and information technologists can provide the necessary levels of access for their clients and become an ever more valuable asset to the dissemination of information through the World Wide Web.

17. References

1. www.terralink.co.nz/profile/glossary/
2. www.nyskwic.org/u_data_terms.cfm
3. www.savi.org/savil/documentation/glossary.aspx

4. http://www.dlib.org/dlib/july98/rust/07rust.html
5. http://www.dlib.org/dlib/january99/buckland/01buckland.html
6. http://dublincore.org/documents/dcmi-terms/
7. http://www.getty.edu/research/conducting_research/standards/intrometadata/
8. http://www.ifla.org/II/metadata.htm
9. http://www.ifla.org/II/catalog.htm
10. http://www.ifla.org/udt/dc8/index.htm
11. http://www.mysql.com/it-resources/white-papers/telecom.php

Digital Library: File Formats, Standards and Protocols

Umesha Naik and Shivalingaiah D.

Abstract

The development of the WWW and other networked digital information systems has provided information professionals with many opportunities, while at the same time requiring them to confront issues that they have not had occasion to explore previously. Now the mode of conversion, storage and dissemination of information are changing in electronic era. This paper highlights the various file formats, standards and protocols which may act as a guideline for preservation, maintenance, storage and access of digital materials. This paper also considers in depth the issue surrounding the file formats, standards and protocols of the digital library materials. As with so many issues surrounding digitization the question of choosing which file format, standardisation and protocols to use both store and access the information is somewhat vexed one. There is a great need for adopting various standards and best practices to build interoperable digital libraries.

Keywords: Digital Library, File Formats, Standards, Protocols, Metadata, Z39.50, MARC 21, Image files, Text files, Program Files, Structured files.

1. Introduction

Digital library is a collection of digital objects, including text, video and audio along with methods for access, retrieval, selection, organisation and maintenance of the collection. The important point is that a digital library has materials stored in a computer system in a format that allows it to be manipulated and delivered in ways that the conventional version of the materials cannot be. File formats vary in terms of resolution, bit-depth, colour capabilities, and support for compression and metadata. As with all matters to do with computers there are standards which impinge on the area of digital libraries.
The standards fall into three areas - material description, user access and system architecture. The quality and the formats of the items may vary Lecturer considerably, particularly where results whether they are hit lists or the final required objects, come from a number of sources. The goal is to access relevant information seamlessly, regardless of its type and location. A protocol is the special set of rules that end points in a telecommunication connection use when they communicate.

2. Text and Document Files

A file that consists of text characters without any formatting information it is also known as an ASCII file. A text file can be read by any word processor and file that contains characters in a plain human-readable format.

2.1 Text file

The lowest common denominator of text file formats contains only ASCII characters without any special formatting. A text is not an undifferentiated sequence of words, much less of bytes. There are actually two forms of ASCII: *standard and extended.* Standard ASCII contains codes for 128 characters (i.e. a 7 bit binary code). Ed

Mangalore University, Mangalagangothri-574199; E-mail:umeshai@yahoo.com

It is transportable across all networks and capable of being accessed and manipulated on all computers. Extend ASCII is a non-standard format containing codes for 256 characters. It is the most basic file format used to transfer data on the Internet.

2.2 Portable Document File (PDF)

PDF has become a popular format for producing and delivering electronic files on and off the Web. These files are generally larger than plain text files and the quality of reproduction is higher, especially in print. A special reader, Acrobat Reader software for Windows, Macintosh, and Sun is required and is available free from Adobe Systems Inc.

2.3 Rich Text Files (RTF)

This is supported by the majority of word processing software packages such as Microsoft Word, Word Perfect, etc. and works on any operating system (Windows, Mac, UNIX, etc.). It was not however designed as a full-featured typesetting language. It defines control words and symbols that serve as "common denominator" formatting commands.

2.4 Microsoft Word or Document file (DOC)

It enables us to create, edit, print and save documents for future retrieval and reference. Creating a document involves typing by using a keyboard and saving it. Editing a document involves correcting the spelling mistakes, if any, deleting or moving words sentences or paragraphs. Text is typing into the computer, which allows alterations to be made easily.

3. Image Files

Graphic files are in one of two basics designs, these two designs serve different purposes and it is essential to where and why anyone should use one or the other. Because of the bandwidth issues surrounding networked delivery of information and because image files contain so much information, Web graphics are by necessity compressed. Different graphic file formats employ varying compression schemes, and some are designed to work better than others for certain types of graphics.

3.1 Tag Image File Format (TIFF)

TIFF is the format of choice for archiving important images. it is the leading commercial and professional image standard. It supports up to 48 bits and most colour spaces, RGB, CMYK, YCbCr, etc. It is a flexible format with many options. TIFF with G3 compression is the universal standard for fax and multi-page line art documents. TIFF graphics can be any resolution, in black and white, gray-scaled, or colour and is used un-compressed or LZW compressed. This is the most universal and most widely supported format across all platforms like Macintosh, Windows, UNIX and Linux.

3.2 Portable Network Graphics (PNG)

PNG has been developed to replace the aging GIF format and it is supported by both Microsoft Internet Explorer and Netscape Navigator it also supports progressive rendering, as interlaced GIFs do, and tends to compress better than a GIF. PNG does not support lossy compression because its developers believed that JPEG was a satisfactory standard in

that area. It provides an improved means of progressive display through two-dimensional interlacing and its lossless compression means that it is a good format choice for storing intermediate-stage images. The following table shows that the sizes of the image file formats.

3.3 Bitmap graphics (BMP)

Bitmap files are stored in a device-independent bitmap format that allows windows to display the bitmap on any type of display device. Each bitmap file contains a bitmap-file header, a bitmap-information header, a colour table, and an array of bytes that defines the bitmap bits. BMP images can range from black and white up to 24 bit colour. While the images can be compressed, this is rarely used in practice.

3.4 Joint Photographic Experts Group (JPEG or JPG)

This is a minimal file format which enables JPEG bit streams to be exchanged between a wide variety of platforms and applications. It can reduce file sizes to about 5% of their normal size some detail is lost in the compression and is optimized for the display of photographs and doesn't work as well as GIF for type or line drawings. JPEG compression gives very high compression rates with high perceived image quality and can handle high resolution images with colour depths of 24-bit and above.

3.5 Graphics Interchange Format (GIF)

It is best used for line art such as cartoons, graphs, schematics, logos, and text that have a limited number of colours and distinct boundaries between colour regions. It is a web standard small file size allows images to transmit quickly over the Internet. GIF files incorporate a compression scheme to keep file sizes at a minimum, and they are limited to 8-bit colour palettes.

4. Markup Languages

It is a form of annotation that has been used in publishing for centuries to describe the desired effects or operations to be applied to textual content. In 1974, Charles Goldfarb invented SGML. In the late 1980s and early 1990s, following Tim Berners-Lee's proposal, a simplified markup language, based upon SGML, was developed for the encoding of documents for distribution via Internet. With the effect of marketing changes SGML having proper languages called HTML.

4.1 Standard Generalized Markup Language (SGML)

For most digital library or digital publishing SGML content, the texts are encoded not in HTML but in more powerful and descriptive tag sets, such as the Text Encoding Initiative Guidelines. It is a “meta-language” used to define structural markup languages standardized (ISO 8879) in 1980s. SGML texts are comprised of plain ASCII text, combined with items in angle brackets, e.g. <title> Digital materials: file formats, standards and protocols </title>. It is both a language and an ISO standard for describing information embedded within a document. It identifies document elements such as titles, paragraphs, tables and chapters as distinct objects, allowing users to define the relationships between the objects for structuring data in documents.

4.2 Hyper Text Markup Language (HTML)

HTML is a non proprietary format based on SGML. It can be created and processed in a wide range of software programs, from simple plain text editors. It is a mark-up language that uses tags to structure text into headings, paragraphs, lists, and links. It tells a Web browser how to display text and images. As it was combined with embedded images and other media , it became increasingly popular. However, its fixed set of features, or elements and attributes, and its emphasis on presentation proved limiting. It presents data in a browser window in a multitude of shapes, sizes and colours, and provides extra features, such as forms and hyperlinks.

4.3 Dynamic Hyper Text Markup Language (DHTML)

DHTML is not a language in and of itself like HTML or JavaScript. It's a method used to design a web page by mixing HTML, Cascading Style Sheets, Document Object Model and some scripting language such as JavaScript or VBScript. It is the combination of several built-in browser features in fourth generation browsers (Internet Explorer 4 (or higher), Netscape Navigator 4 (or higher) that enable a web page to be more dynamic. It is not a standard defined by the World Wide Web Consortium, it is a "marketing term" that was used by Netscape and Microsoft to describe the new technologies the 4.x generation browsers would support.

4.4 eXtensible Markup Language (XML)

XML is a lightweight cut-down version, simple, very flexible text format of SGML (ISO 8879) which keeps enough of its functionality to make it useful but removes all the optional features which make SGML too complex to program for in a Web environment. It designed to meet the challenges of large-scale electronic publishing. SGML and XML represent a document by tagging the document's various components with their function or meaning. It supports a wide variety of applications. The number of optional features in XML is to be kept to the absolute minimum, ideally zero. XML documents should be human legible and reasonably clear. And Terseness in XML markup is of minimal importance.

5. Structured data

5.1 Metadata

Metadata is often described as “data about data” or information known about the image in order to provide access to the image. Usually includes information about the intellectual content of the image, digital representation data, and security or rights management information. The Special tagged fields in a document that provide information about the document to search engines and other computer applications. A metadata record can include representations of the content, context, structure, quality, provenance, condition, and other characteristics of an IBO for the purposes of representing the IBO to a potential user - for discovery, evaluation for fitness for use, access, transfer, and citation.

5.2 Dublin Core

Dublin Core Metadata Element Set consists of 15 descriptive data elements relating to content, intellectual property and instantiation. The elements are title, creator, publisher,

subject, description, source, language, relation, coverage, date, type, format, identifier, contributor and rights. They are to be supplied by the producer of the resource. The Warwick Framework set out a conceptual approach to implementing the Dublin Core, one of which is embedding the data in an HTML document using the META tag.

6. Communication

6.1 TCP/IP

TCP/IP is a communications protocol used to transfer digital data around the internet. TCP and IP were developed by a Department of Defence research project to connect different networks designed by different vendors into a network of networks. IP is responsible for moving packet of data from node to node. IP forwards each packet based on a four byte destination address. The Internet authorities assign ranges of numbers to different organizations. The organizations assign groups of their numbers to departments. IP operates on gateway machines that move data from department to organization to region and then around the world. TCP is responsible for verifying the correct delivery of data from client to server.

7. Protocols

In information technology, a protocol is the special set of rules that end points in a telecommunication connection use when they communicate. There are protocols between each of several functional layers and each corresponding layer at the other end of a communication. Both end points must recognize and observe a protocol.

7.1 HyperText Transfer Protocol (HTTP)

HTTP is the set of rules for exchanging files (text, graphic images, sound, video, and other multimedia files) on the WWW and it is both simple and powerful network protocol of the web. Usually, HTTP takes place through TCP/IP sockets. A feature of HTTP is the typing and negotiation of data representation, allowing systems to be built independently of the data being transferred. A browser is an *HTTP client* because it sends requests to an *HTTP server* (Web server), which then sends responses back to the client. The standard (and default) port for HTTP servers to listen on is 80, though they can use any port. It is used to transmit *resources*, not just files. A resource is some chunk of information that can be identified by a URL.

7.2 File Transfer Protocol (FTP)

A Network protocol FTP is a standard Internet protocol (a formalized procedure for transferring digital information) that provides web authors and other web users with an easy way to upload and download files between computers that are connected to the Internet. These servers store a tremendous amount of information, and they are available to anyone on the Internet using FTP. When an FTP server has an access restriction, it requires a user identification code and password before allowing access to its files. If it has no access restriction, the FTP server allows any and every user to access its files; we call these servers anonymous FTP servers.

7.3 Z39.50

Z39.50 has to allow standardized means of cross-database searching among a handful of

(rather homogeneous) major bibliographic databases hosted by organizations. Z39.50 is the ANSI retrieval application service definition and protocol specification for open systems interconnection, application protocol for search and retrieval of information in databases. It is a standard way for two computers to communicate for the purpose of information retrieval. It is one of a set of standards produced to facilitate the interconnection of computer systems. The standard specifies formats and procedures governing the exchange of messages between a client and server, enabling the user to search remote databases, identify records which meet specified criteria, and to retrieve some or all of the identified records and is concerned, in particular, with the search and retrieval of information in databases.

8. Scripting and programming

8.1 Java Script

JavaScript is a programming language that can be included on web pages to make them more interactive and was written with the Internet in mind. It is a scripting language developed by Netscape and continues to mature with each version of Netscape Navigator. JavaScript adds interactivity and power to web documents. The JavaScript fundamentals, including variables, if-then statements, link events, and image swaps. JavaScript DOM, windows and frames, JavaScript syntax with loops, arrays, and functions, and forms. JavaScript only execute on the page(s) that are on your browser window at any set time.

8.2 Practical Extraction and Report Language (PERL)

It is an interpreted high-level programming language and used most often for web programming in CGI scripts. It has become the premier scripting language of the Web, as most CGI programs are written in Perl. It is popular with system administrators who use it for an infinite number of automation tasks. It is an interpreted programming language known for its power and flexibility. It is strong as a tool for programming on the WWW because it is a quick and effective way to prepare interactive applications.

8.3 Common Gateway Interface (CGI)

It is a standard for interfacing external applications with information servers, such as HTTP or Web servers. A plain HTML document that the Web daemon retrieves is static, which means it exists in a constant state, a text file that doesn't change. A CGI program, on the other hand, is executed in real-time, so that it can output dynamic information. Languages used to write CGI scripts include compiled languages C, C++ and interpreted languages perl, JCL (e.g. The Unix sh command language) The CGI script is executed when an anchor tag <A ... > or an image tag <IMG ...> refers to the CGI script file rather than a normal file.

9. Web Browsers

Netscape and Microsoft Internet Explorer are standards for Web browsers, which are software that acts as an interface between the user and the WWW. Browsers are also referred to as web clients because in the client/server model, the browser functions as the client program.

9.1 Internet Explorer (IE)

This is a graphical web browser, which enables a user to experience the hypertext, photographs, sound, video, etc. available on the WWW. It works as part of the Windows operating system. It has some features not available in Netscape. IE can be used on PCs running Windows or on all Macintosh systems running System 7.

9.2 Netscape Navigator (NN)

It provides a useful web site HTML editor, allowing complex HTML to be written for you in a WYSIWYG environment. It works as part of the Windows operating and Linux operating system. Netscape Communicator is a suite of software applications that includes Navigator, Messenger, Collabra, Composer, and Conference etc.

10. Audio/Video files

10.1 AUdi (.au)

AU is short for Audio, a common digital sound file format used on UNIX machines and the standard audio file format for the Java programming language. The file has a very simple structure, the file header specifies the basic parameters of the sound sampling rate, sample size, number of channels and type of encoding - followed by the sound data.

10.2 WAV

The WAV file format was developed by IBM and Microsoft as the Resource Interchange File Format WAV to store arbitrary sound data in a structured format. The format has become the de facto standard for storing sound files on Windows machines: an operating system which uses WAV files for basic system sounds.

10.3 Musical Instrument Digital Interface (MIDI)

This file will store a number of aspects of the sound, including the note's pitch, length and volume, as well as additional characteristics such as vibrato and delay time. A MIDI-compatible computer records music as keystroke and control codes, which, when saved as a MIDI file, can be easily manipulated by any other MIDI-compatible device. The MIDI standard allows for musical pieces to exchanged and edited by different computers on different platforms in a way that conventional digitised sound (actual waveforms) cannot.

10.4 Moving Picture Expert Group (MPEG)

A compression standard for video, which transfers all information of every tenth frame, with the subsequent nine frames being transmitted only as significant changes to that reference frame. It works on all platforms, but of lower quality. MPEG uses a type of lossy compression that is generally imperceptible to the human eye.

10.5 Audio Video Interleave (AVI)

An AVI file plays on a PC via an application capable of parsing the AVI file header and consecutively pulling in the video frame and accompanying audio. The video is then decompressed and displayed in sequence with the audio sample which has been sent to the soundcard for output. AVI is a specialization of the RIFF and it is the most popular format for audio/video data on the PC and is widely supported in the Windows platform. AVI is

of relatively low quality (when compared to MPEG-1, for example) with a limiting frame rate and mono sound.

11. Database Technology

11.1 Oracle

Oracle is a fourth generation RDBMS. In general, a DBMS must be able to reliably manage a large amount of data in a multi-user environment so that many users can concurrently access the same data. The ORACLE Server provides efficient and effective solutions for the major database features. Oracle consists of many tools that allow us to create an application with ease and flexibility. The features and tools that choose to use to implement the application can significantly affect the performance of application.

11.2 Microsoft Access

It is a RDBMS. At the most basic level, a DBMS is a program that facilitates the storage and retrieval of structured information on a computer's hard drive. It is a powerful program to create and manage the databases and it has many built in features to assist in constructing and viewing the information. Some keywords involved in this process are: *Database File, Table, Record, Field, and Data-type.* Access has tools to enter, edit, and index data and to retrieve it via custom forms and reports.

11.3 Open Database Connectivity (ODBC)

It is an open standard application programming interface (API) for accessing a database a set of functions that provide access to client-server RDBMS, desktop database files, text files, and Excel worksheet files through ODBC drivers. A common standard of Microsoft that allows relational and non-relational database programs like SQL, dBase, Oracle, Access, etc, to access the information independent of its file format. The goal of ODBC is to make it possible to access any data from any application, regardless of which DBMS is handling the data.

11.4 Structured Query Language (SQL)

It was formally defined by Dr. E. F. Codd in 1970 and is a specialized, standard, interactive programming language for updating, deleting, and requesting information from databases. It is an ANSI and ISO standard, and is the de facto standard database query language. It is widely used in both industry and academia, often for enormous, complex databases. SQL has now been institutionalized by the creation of an ANSI standard for the language. It is a mature, powerful, and versatile relational query language.

12. Conclusion

This paper has identified a variety of issues relating to digital file formats, standards and protocols. Digital library is an emerging field in the area of information and library sciences, with rapid development in IT, practically, the web technology, the world of digital information resources has changed rapidly and exponentially. The formats that offer the greatest stability and promise to endure for several years are the many image formats that reproduce documents and pictorial materials, and those used for searchable texts, including formats that employ SGML and HTML. The use of standards can pose several well-known problems. Several versions of a standard may be in use. A standard may not

be well specified and may be differently implemented in software. Some standards may have more features than are likely to be used in practice; inter-working problems may arise where different subsets are used in different implementations. For these reasons, it is common to have an additional layer or layers of agreement above the 'base' standards.

References

1. The Digital Information Office. Retrieved September 2, 2004, from http://dio.cdlr.strath.ac.uk/standards/fileformats/
2. *Digital library standards and practices.* Retrieved July 1, 2004, from http://www.diglib.org/standards.htm
3. Haigh, Susan. *A Glossary of Digital Library Standards, Protocols and Formats* Retrieved January 15, 2004, from http://www.collectionscanada.ca/9/1/p1-253-e.html
4. *Learn the net: File Formats & Extensions.* Retrieved July 1, 2004, from http://www.learnthenet.com/english/html/34filext.htm
5. *Sun Microsystems Digital Library Toolkit.* Retrieved December 10, 2003, from http://www.sun.com/products-n-solutions/edu/libraries/digitaltoolkit.html
6. *World Wide Web Consortium (W3C).* Retrieved April 05, 2004, from http://www.w3.org
7. *Graphic File Formats.* Retrieved July 15, 2004, from http://extension.oregonstate.edu/esoc/ectu/services/lessons/GraphicFileFormats.php
8. *Web Style guide- Graphic file formats.* Retrieved July 15, 2004, from http://www.webstyleguide.com/graphics/formats.html
9. Marilyn, Deegan, & Simon, Tanner. (2002). *Digital futures: strategies for the information age*. London: Library Association.
10. Stuar, D. F. (2001). *Digital Imaging: A practical handbook*. London: Library Association.
Witten, Ian. H, & Bainbridgw, David. (2003). *How to build a digital library*. San Francisco: Morgan Kaufman Publishing.

Mobile Communication Technologies

Jehadeesan R. and Rajan J.

Abstract

The field of mobile communications has witnessed tremendous growth in recent years and it has become one of the fastest growing segments of telecommunications industry. This paper aims to provide details on the

- Evolution of mobile communications
- Principles of mobile communications covering modulation and access techniques
- Overview of 2G and 3G cellular systems
- Insight into GSM and CDMA technologies and how they work
- Comparison of widely used GSM and CDMA technologies and future trends.

This paper highlights the currently dominating second generation mobile technologies namely Code Division Multiple Access (CDMAone) and Global System for Mobile Communication (GSM). Also it covers the standards used for data transmission over 2G communication networks like GPRS, CDMAtwo. The paper concludes with a brief description of evolving third generation standards like EDGE, WCDMA and futuristic fourth generation technologies.

The mobile-internet combination will shift the nature of wireless systems from today's voice oriented one to data centric. This can be used for transmitting data/video contents over wireless platform and meet the challenge of accessing rich multi-modal information on wireless handheld devices.

1. Evolution of Mobile Communications

In 1895 a few years after invention of telephone, Marconi demonstrated the first radio based wireless transmission. The first radio based conversation was used in ships during 1915. The first public mobile telephone system known as Mobile Telephone System (MTS) was introduced in United States in 1946.

AT&T Bell laboratories devised the cellular concept which replaced high power high coverage base stations used by MTS with a number of low power low coverage stations. The area of coverage of each such base station is called a cell. The operating area of the system was divided into a set of adjacent, non-over lapping cells.

The first generation (1G) of cellular system was designed in late 1960s and deployed in early 1980s. The first commercial analog system in the United States known as Advanced Mobile Phone System (AMPS) went operational in 1982 with only voice transmission.

The disadvantages of Analog systems were overcome by second (2G) generation of cellular systems which represent data digitally. The first commercial deployment of 2G system called GSM was made in 1992. In 1993 as other 2G system also known as CDMAone (IS-95) was standardized and commercially deployed in SouthKorea and Hongkong in 1995, followed by United States in 1996.

Upgrade to 2G systems offering higher data speeds called 2.5G systems was developed. GSM has two such technologies called High Speed Circuit Switched Data (HSCSD) and

---

Computer Division, Indira Gandhi Centre for Atomic Research, Kalpakkam. *Email : jeha@igcar.ernet.in*

General Packet Radio Service (GPRS). Similarly in CDMA an extension of IS-95 known as IS-95B or CDMATwo was developed.

To meet the future bandwidth hungry services 3G cellular systems was standardized in 2000. The different 3G standards evolved include EDGE, CDMA2000 and WCDMA.

It is envisioned that future of mobile communication will be towards an integrated system which will produce a common packet switched possibly IP-based system. This might be called the 4G of cellular networks which targets the market of 2010 and beyond.

Mobile Gen.	Dates	New Features
1G	70's to 80's	Wireless phones (cellular) are introduced, primarily for voice only.
2G	90's to 2000	Increased performance achieved by allowing multiple users on a single channel
2.5G	2001-2004	Enhanced multimedia and streaming video are now possible with web browsing
3G	2004-2005	Enhanced multimedia and streaming video capabilities are increased. universal access and portability across different device types (Telephones, PDA's, etc.)
4G	2006+	Speeds reach up to 40 Mbps. Enhanced multimedia, streaming video, access and portability are increased still further. world-wide roaming.

2. Principles of Mobile Communication

Each mobile uses a separate, temporary radio channel to talk to the cell site (base station). The cell site talks to many mobiles at once, using one channel per mobile. Channels use a pair of frequencies for communication – one frequency, the forward link, for transmitting from the cell site and one frequency, the reverse link, for the call site to receive calls from the users. Radio energy dissipates over distance, so mobiles must stay near the base station to maintain communication.

Codecs convert an analog speech signal to its digital representation by sampling analog signal at regular time intervals called pulse code modulation. For analog data to be transmitted analog modulation techniques like amplitude modulation and frequency modulation are used. The modulation techniques used for digital signal transmission are amplitude shift keying, frequency shift keying, phase shift keying

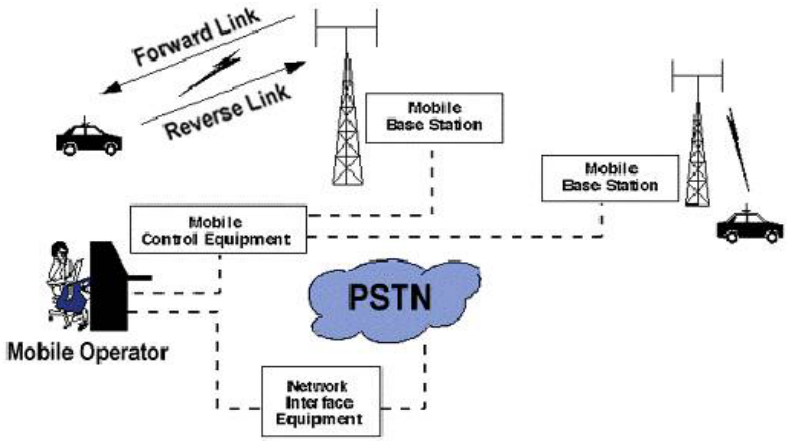


Figure 1: Basic Mobile Telephone Service Network

.

3. Cellular System Architecture

Because the amount of frequency spectrum available for mobile cellular use was limited, efficient use of required frequencies was needed. Provisioning for each region is planned accordingly that includes cell, clusters, frequency reuse, cell splitting and handoff.

3.1 Cells and Clusters

A cell is the basic geographical unit covered by cellular radio antennas. This is represented in the form of simple hexagon. A cluster is a group of cells. No channels are reused within a cluster. Cluster is usually seven.

3.2 Frequency Reuse

Because only a small number of radio channel frequencies were available, the solution industry adopted was called frequency planning or frequency reuse. The concept of frequency reuse is based on assigning to each cell a group of radio channels used within a small geographical area. Cells are assigned a group of channels that is completely different from neighbouring cells.

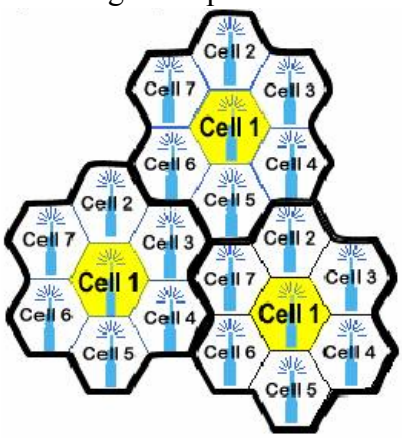


Fig. 2: Frequency Reuse

The coverage area of cells are called footprint. This footprint is limited by a boundary so that the same group of channels can be used in different cell that are far enough away from each other so that their frequencies do not interfere. Cells with the same number have set of frequencies. Here the number of available frequencies is seven, the frequency reuse factor is 1/7. Each cell is using 1/7 of available cellular channels.

3.3 Cell Splitting

Economic considerations made the concept of creating full systems with many small areas impractical. To overcome this difficulty, the idea of cell splitting was developed. As a service area becomes full of users, a single area is split into smaller ones. Urban centers

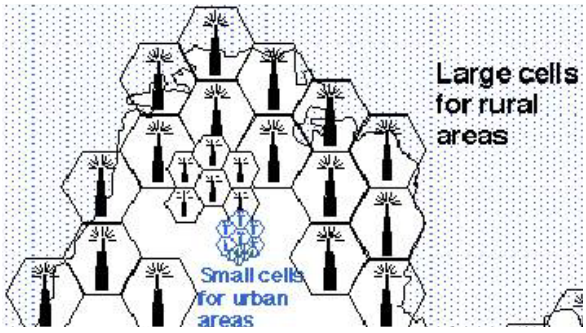


 Figure 3: Cell Splitting

can be split into as many areas as necessary in order to provide acceptable service levels in heavy traffic region, while larger, less expensive cells can be used to cover remote rural areas.

3.4 Handoff

This is done when a mobile subscriber travels from one cell to another during a call. As adjacent areas do not use the same radio channels a call must either be dropped or transferred from one radio channel to another when a user crosses the line between adjacent cells. Because dropping the call is unacceptable, the process of handoff was created. Handoff occurs when the mobile telephone network automatically transfers a call from one radio channel to another radio channel as mobile crosses adjacent cells.

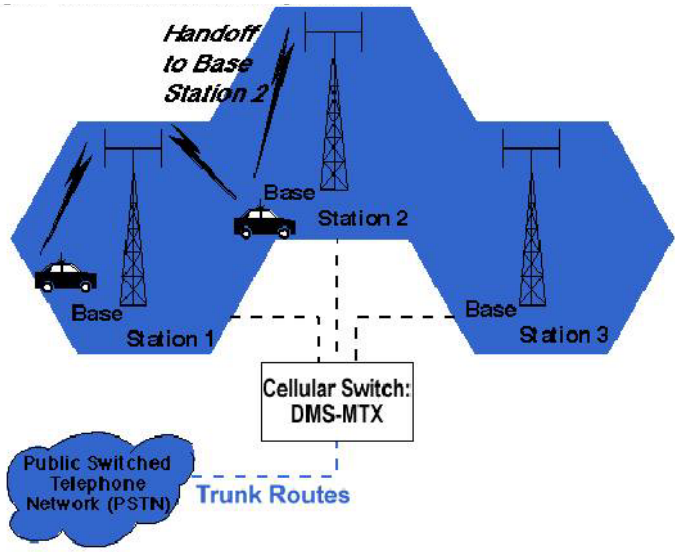


Figure 4: Hand-off between adjacent cells

During a call, two parties are on one voice channel. When the mobile unit moves out of the coverage area of a given cell site, the reception becomes weak. At this point the cell site in use requests a handoff. The system switches the call to a stronger frequency channel in a new site without interrupting the call or alerting the user. The call continues as the user is talking and the user does not notice the handoff at all.

4. Access techniques

The various access techniques used in mobile communication are:

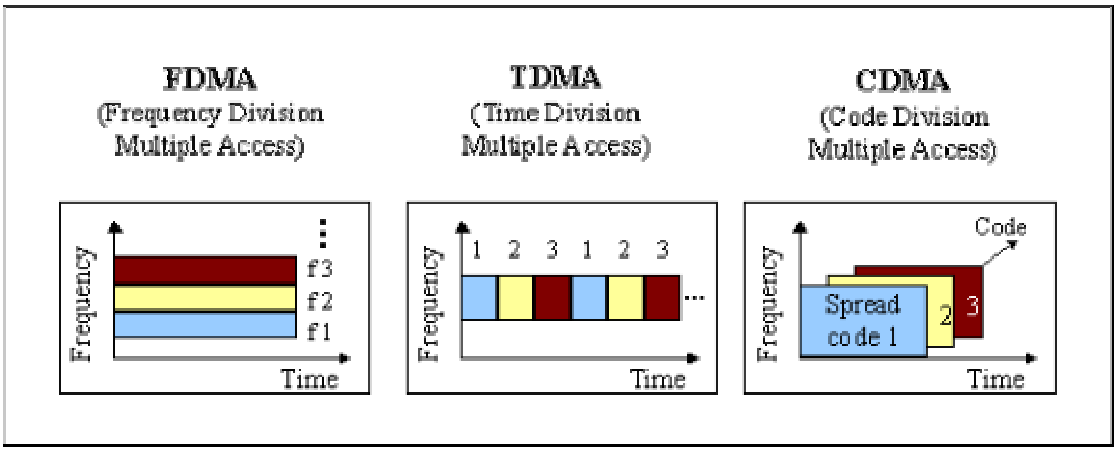


Figure 5: Access Technologies

4.1 FDMA

Frequency Division Multiple Access is the most commonly used analog system. It is a technique whereby spectrum is divided up into frequencies and then assigned to users. With FDMA only one subscriber at any given time is assigned to a channel. The channel is therefore closed to other conversations until the initial call is finished or until it is handed-off to a different channel. A full duplex FDMA transmission requires two channels. One for transmitting and the other for receiving. FDMA has been used for first generation analog systems.

4.2 TDMA

Time Division Multiple Access (TDMA) improves spectrum capacity by splitting each frequency into time slots. TDMA allows each user to access the entire radio frequency channel for the short period of a call. Other users share this same frequency channel at different time slots. The base station continually switches from user to user on the channel. TDMA is the dominant technology for the second generation mobile cellular networks.

4.3 CDMA

Code Division Multiple Access is based on spread spectrum technology. Since it is suitable for encrypted transmissions, it has long been used for military purposes. CDMA increases spectrum capacity by allowing all users to occupy all channels at the same time. Transmissions are spread over the whole radio band and each voice or data call is assigned a unique code to differentiate from the other cells carried over the same spectrum. CDMA allows for a soft hand-off, which means that terminal can communicate with several base stations at the same time

5 GSM technology

In 1992 European Union formed a study group called the 'Groupe Special Mobile' later renamed to Global System for Mobile Communications. The existence of one standard boosted the cellular industry in Europe contrary to the United States where several different 2G systems have been deployed leading to a fragmented market.

Though GSM was standardized in Europe it has been deployed in large number of countries worldwide (approximately 110). There are four versions of GSM depending on the operating frequency. They are :

GSM Variant	Uplink Frequency (MHz)	Downlink Frequency (MHz)
GSM 900	890 ~ 915	935 ~ 960
GSM 1800 (DCN)	1710 ~ 1785	1805 ~ 1880
GSM 1900 (PCS)	1850 ~ 1910	1930 ~ 1990
GSM 450	450.4 ~ 457.6 (or) 478 ~ 486	460.4 ~ 467.6 (or) 488.8 ~ 496

The primary service supported by GSM is voice telephony. Speech is digitally encoded and transmitted through the GSM network as a binary bit stream. For emergency situations, an emergency service is supported by dialing a certain 3 digit number (like 123).

GSM also offers a variety of data services. It allows to users to send and receive data , at rates up to 9.6Kbps. Data can be exchanged using a variety of access methods and protocols such as X.25. A modem is not required between the user and GSM network due

to the fact that GSM network is a digital network. GSM also supports the short message service (SMS) and Cell Broadcast Service (CBS). It supports additional services like call forward, call barring of outgoing or incoming caller, caller identification, call waiting, multiparty conversations etc.

5.1 Network Architecture

The layout of a GSM network is shown below: The mobile station (MS) and base station (BS) communicates across the Um interface or radio link. The BS communicates with the Mobile Switching Center (MSC) across the A interface.

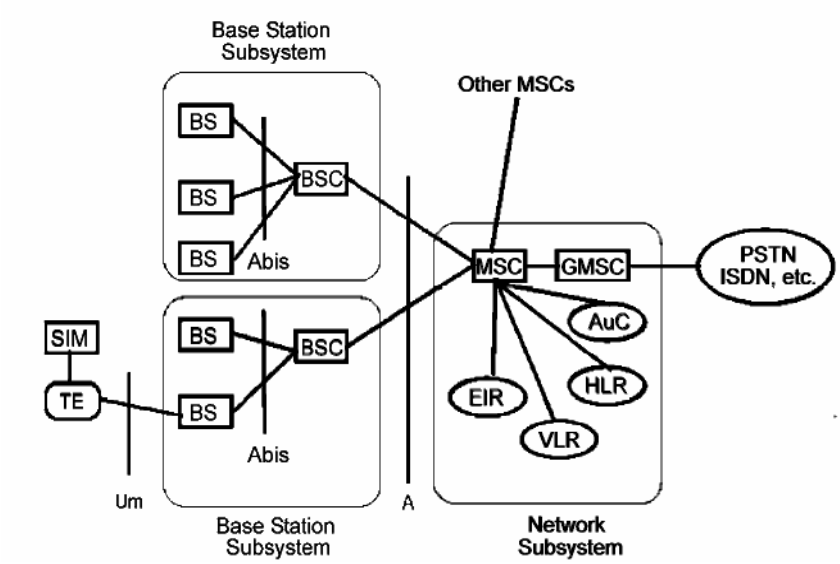


Figure 6 : GSM Network Architecture

5.1.1 Mobile Station (MS)

It consists of the terminal (TE) and a smart card called the subscriber identity module (SIM). The SIM provides personal mobility so that user can have access to subscribed services irrespective of a specific terminal. The SIM card is the actual place where the GSM network finds the telephone number of the user. Thus by inserting the SIM card into another GSM terminal, the user is able to use the new terminal receive, make calls and use other subscribed services while using the same telephone number.

The GSM terminal is uniquely identified by the International Mobile Equipment Identity (IMEI). The SIM card contains the International Mobile Subscriber Identity (IMSI) used to identify the subscriber to the system, a secret key for authentication and other information. The structures of IMEI and IMSI are shown below :

TAC (3 digits)	FAC (1 or 2 digit)	Serial number (upto11 digits)	1 spare digit

IMEI structure

MCC (3 digits)	MNC (2 digits)	MSIC (upto 10 digits)

IMSI structure

The IMEI can be up to 15 digits and comprises:

- 3 digit Type Approval Code (TAC). This is given to the unit after it passes conformance tests.

- 1 or 2 digit Final Assy Code (FAC). This identifies the place of manufacture of assembly of the MS unit
- 11 digit MS unit serial number
- 1 spare digit reserved for future use.

The IMSI is also 15 digits and consists of:

- 3 digit Mobile Country Code (MCC). This identifies the country where the GSM system operates.
- 2 digit Mobile Network Code (MNC). This uniquely each cellular provider.
- The Mobile Subscriber Identification Code (MSIC). This uniquely identifies each customer of the provider.

5.1.2 Base Station Subsystem (BSS)

The BSS contains the necessary hardware and software to enable and control the radio links with the mobile stations. It contains two parts, the base station (BS) and the base station controller (BSC). These communicate across the standardized Abis interface, allowing operation between components made by different suppliers. The BS contains the radio transceivers that define a cell and handles the radio link protocols with the MS. BSs are responsible for frequency administrations and handovers.

5.1.3 Network Subsystem

The central component of the network subsystem is the Mobile Switching Center (MSC). The MSC performs switching of user calls and provides functionality for registration, authentication, location updating, handovers and call routing to a roaming subscriber. The MSC interfacing the GSM network to fixed networks (PSTN) is known as gateway MSC.

5.2 Speech Coding & Radio Transmission Characteristics

The speech coding algorithm used in GSM is called Regular Pulse Excited-Linear Predictive Coder with a long term predictive loop. As with any other wireless networks, GSM encodes data into waves in order to send it over the wireless medium. The actual modulation scheme used is Gaussian Minimum Shift Keying (GMSK), which achieves 270.8 kbps over each of the 200-kHz wide GSM channels. The available bandwidth in GSM is split into 124 carriers, each 200-kHz wide. GSM uses a combination of Time and Frequency Division Multiple Access (TDMA/FDMA) for user separation. One or more carrier frequencies are assigned to each BS of the GSM network and each of those carriers is divided in the time domain. Each time period is a slot and lasts 0.577ms.

6 CDMA Technology (IS-95)

cdmaOne, a 2G system developed by Qualcomm also known as IS-95, has been standardized in 1993 and the first commercial systems were deployed in 1995. cdmaOne utilizes Code Division Multiple Access (CDMA). In cdmaOne, multiple mobiles in a cell, whose signals are distinguished by spreading them with different codes, simultaneously use a frequency channel. Thus, neighboring cells can use the same frequencies, unlike all other standards. cdmaOne operate in the same band with AMPS and it is designed to support dual-mode terminals that can operate either under an cdmaOne network or an AMPS

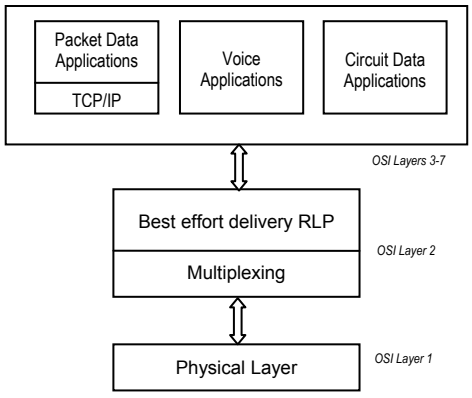


Figure 7: Protocol Architecture of CDMA

network. cdmaOne supports data traffic at rates of 4.8 and 14.4 kbps.

The OSI model protocol architecture of cdmaOne is shown in Figure 8. Layer 1 deals with the actual radio transmission, frequency use, etc. Layer 2 offers a best effort delivery of voice and data packets. The MAC sublayer of this layer also performs channel management. Data originating from different sources are multiplexed and handed for transmission to the physical layer.

6.1 Network Architecture

cdmaOne reuses the AMPS spectrum in the 800 MHz band. cdmaOne uses a channel width of 1.228 MHz both on the uplink and downlink. Therefore, 41 30 kHz AMPS channels are grouped together for cdmaOne operation. In cdmaOne, the same frequency is reused in all cells of the system. This leads to a frequency reuse factor of 1 and is due to the fact that cdmaOne identifies the transmissions of different mobiles via the different spreading codes that identify each mobile. Both cdmaOne BSs and MSs utilize antennas that have more than one element (RAKE receivers) in order to combat the fading wireless medium via space diversity. The use of CDMA for user separation imposes the need for precise synchronization between BSs in order to avoid too much interference. This synchronization problem is solved via the use of the Global Positioning System (GPS) receivers at each BS. Once the BSs are synchronized, which in turn synchronize MSs with them.

cdmaOne utilizes the IS-41 network protocol. The topology of IS-41 shown in Figure 9 is quite similar to that of the network side of GSM. It is defined by a number of functional entities. The way two functional entities communicate and exchange information is defined by the corresponding interface.

The entities shown in this figure are described briefly below:

AC - Access Control, BS - Base Station, CSS - Cellular Subscriber Station, EIR - Equipment identity Register, HLR - Home Location Register, ISDN - Integrated Services Digital Network, MC - Message Center, MSC - Mobile Switching Center, SME - Short Message Entity, VLR - Visitor Location Register

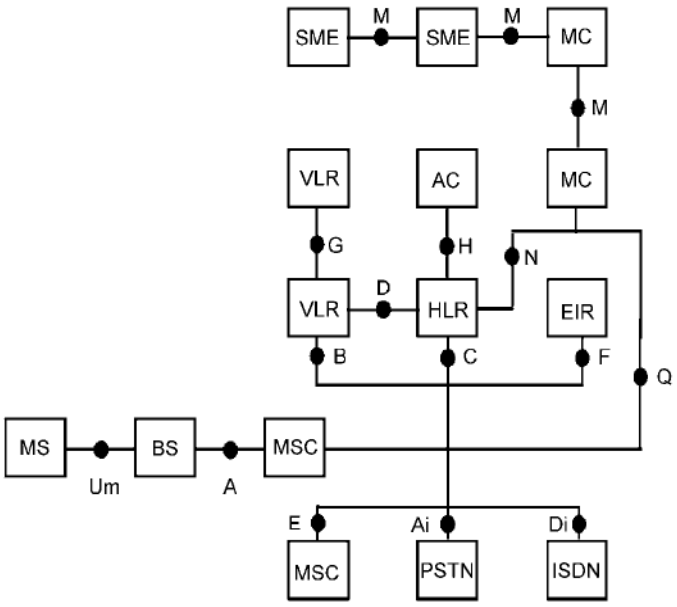


Figure 8: IS-41 network topology used by CDMA

6.2. Channels

Downlink (Forward) channels are those carrying traffic from the BS to the MSs. The cdmaOne downlink is composed of 64 channels distinguished from each other by using different CDMA spreading codes, W0 to W63. The spreading code is an orthogonal code, or called Walsh function. The cdmaOne downlink comprises common control and dedicated traffic channels, most important of which are Pilot, Sync, Paging channels and

Traffic channels. cdmaOne uses Quadrature Phase Shift Keying (QPSK) modulation for transmitting the resulting signal over wireless medium.

Uplink (Reverse) channels are of two types: access and traffic. There can be up to 32 access channels on the uplink, each of which operates at 4800 bps. These channels are used by MSs to initiate calls and respond to paging messages. There can be up to 62 traffic channels on the uplink. These are used to carry user data. The encoded spread data stream is transmitted over the wireless medium via Offset Quadrature Phase Shift Keying (OQPSK) modulation. OQPSK provides more Forward Error Correction (FEC) than QPSK since MSs cannot coordinate their transmissions as efficiently as BSs.

7. Comparison of GSM and CDMA

Table given below shows the comparison of the most popular 2G Mobile standards.

Feature	GSM	CDMA (IS-95)
Technology	Digital Circuit Switched technology based on TDMA	Digital Circuit Switched technology based on Spread Spectrum
Data Rate	9.6 ~ 14.4 Kbps	9.6~14.4 Kbps (IS-95A), 115 Kbps (IS-95B)
Modulation	GMSK	QPSK , OQPSK
Frequency and Carrier	900 MHz or 1.9 GHz 200 kHz	800 MHz or 1.9 GHz 1.25 MHz
Voice Quality	Good	Good
Coverage	35 km (max)	100 km (max)
System Capacity	Good	Excellent
Hand-off	Hard	Soft
International Roaming	Excellent	Limited
Battery Life of handset	Excellent	Good
Handset Selection	Wide choice	Limited choice
Major Countries	Europe, India	USA
Cellular Subscribers	75% (1.36 Billion)	14% (250 Million)

8. Future Mobile Systems

2G systems are limited in terms of maximum data rate, which makes them practically useless for the increased requirements of future mobile applications. Third generation (3G) mobile and wireless networks aim to fulfill the demands of multimedia communication. 3G standardization activities resulted in three main 3G standards with efficient support for both voice and high bit-rate data services.

- EDGE is a TDMA-based system that evolves from GSM and IS-136 and offers data rates up to 473 kbps.
- Cdma2000 is a fully backwards-compatible descendant of IS-95 enabling smooth transition of a 2G IS-95 system to a 3G system and it supports data rates up to 2 Mbps.
- WCDMA is a CDMA-based system that introduces a new 5-MHz wide channel structure. WCDMA is also capable of offering speeds up to 2 Mbps.

The vision of 4G and future mobile systems aims at providing higher data rates and system interoperability.

9. Conclusion

Wireless networks constitute an important part of the telecommunications market. The result of the integration of Internet with mobile system, the wireless Internet, is expected to significantly increase the demand for wireless data services. The use of wireless transmission and the mobility of most wireless systems give rise to a number challenges that must be addressed in order to develop efficient wireless systems. The challenges include wireless medium unreliability, spectrum use, power management, security, and location/routing. Digital cellular standards GSM and CDMA meet the current requirements in voice communications and being upgraded to meet the future demands in mobile multimedia applications. 3G mobile networks represent an evolution in terms of capacity, data speeds and new service capabilities from second generation mobile networks to provide an integrated solution for mobile voice and data with wide area coverage.

References:

1. Principles of Mobile Communication by Gordon L. Stüber, Springer publisher, 2000
2. Handbook of wireless networks and mobile computing, Stojmenovic (Ivan) ED, 2002
3. http://www.gsmworld.com
4. http://www.cdg.org

Web Resources for Chemical Safety Information: An Overview

Nargund, I N[1], Dr.Mallinath Kumbar[2], Phulsundar, R D[1], Dileep, J R[1] and Bhaskar, K[1]

Abstract

Increase in the number of chemical industrial units coupled with complex technologies and process and use of hazardous chemicals have necessitated the formulation and implementation of effective methods for providing safe and health working conditions in work places. The effectiveness of planning and execution of programmes on chemical safety is substantially related to the relevant information. Information is crucial for the safe handling, storage, transporting, personal hygiene, spills and leaks, fire hazards, environment hazards, disposal etc.
Information sources exclusively available to a broad field such as Chemical Safety and sources in other fields, which contain Chemical Safety information, are numerous. In this paper an attempt has been made to present an overview of the Web Resources which depicts credible information on Chemical Safety. Information products and databases developed and provided by the individual organizations are also highlighted.

Keywords**:** Chemical Safety; Web Resources;

1. Introduction

The need for accurate information on Chemical Safety has always been essential for ensuring the safe working conditions. Providing accurate information to Safety professionals, Safety Engineers and other health professionals is an essential role of information specialists. Through the early 1980s most information sources were key textbooks, review articles and monographs. Within the last 10 years, Safety Professionals have had access to new sources of information readily available through the Internet.

A simple search on *'Chemical Safety'* using a search engine such as *Google* returns over

2. Industrial Safety and Health Information:

For any industrial establishment information is required mainly at three levels viz., at Management Level; Production Level; and at Marketing level. Information at production level relates to: production techniques, technical data, maintenance; inventory; quality control etc. Besides these, Industrial Safety & Health (ISH) information is also one of the main components of Industrial Information System.

2.2 Crore related web pages.

It is not possible to detail even a small portion of the sites that provide general information about Chemical Safety. Therefore, the purpose of this paper is not to review all of these sites, but to give an overview and explain how to obtain relevant information to improve and streamline the development of an effective Safety Programme.

---

[1] *Rare Materials Project, BARC, Yelwala, P.O. Mysore (inn@rediffmail.com)*
[2] *Reader, Dept of Studies in Lib & Inf. Science, University of Mysore*

ISH information is supplementary to the basic knowledge of materials and processes and it leads to the scientific and methodological approach towards the control of work and environment. Information required for ISH programme usually depends on the type of work, material and process involved in the production. In chemical industries, information (guidelines/standards) is required for the safe handling; storage; transportation; hygiene; spills & leaks, fire hazards; first aid; waste management etc.

The web resources described in this paper provide key guidance and other information on Chemical Safety. These sites are chosen based on accessibility, quality of information and frequency of updates. This is not a comprehensive listing given the dynamic character of Web, which is expanding rapidly.

3. International Programme on Chemical Safety- INCHEM Database (URL: http://www.inchem.org/)

International Programme on Chemical Safety (IPCS) is a collaborative venture of World Health Organisation; United Nations Environment Programme; and International Labor Organisation. IPCS INCHEM is an invaluable tool for those concerned with chemical safety and the sound management of chemicals. Produced through cooperation between the International Programme on Chemical Safety (IPCS) and the Canadian Centre for Occupational Health and Safety (CCOHS). It provides current and internationally peer-reviewed chemical safety related publications and database records from international bodies for public access. The information provided consists of chemicals affecting the environment and human health. Offering quick and easy electronic access to thousands of searchable full-text documents, IPCS INCHEM is an invaluable tool for those concerned on chemical risks and the sound management of chemicals.

4. IPCS INCHEM contains the following:

- Concise International Chemical Assessment Document (CICADS)
- Environmental Health Criteria (EHC) monographs
- Health and Safety Guides (HSGs)
- International Chemical Safety Cards (ICSCs)
- IPCS/CEC Evaluation of Antidotes Series
- Joint Expert Committee on Food Additives (JECFA) - Monographs and evaluations
- Joint Meeting on Pesticide Residues (JMPR) - Monographs and evaluations
- Pesticide Data Sheets (PDSs)
- Poisons Information Monographs (PIMs)
- Screening Information Data Set (SIDS) for High Production Volume Chemicals

Organisation for Economic Co-operation and Development (http://www.oecd.org/ehs)

The main focus is the dissemination of tools for chemical safety assessment, which are used by OECD member countries, with the aim of promoting harmonisation. The contents include the activities of the programme with special emphasis on dissemination of information via the web site in the form of up-to-date information on activities, publications, directories, databases, contact points and links to other sites of importance. OECD is an inter-governmental organization with 30 member countries. Activities related to chemical safety are carried out as part of the Environment, Health and Safety (EHS) Programme. The programme co-operates closely with other intergovernmental

organisations, through the Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) and the Intergovernmental Forum on Chemical Safety (IFCS).

Some of OECD resources on Chemical Safety which are featured in the website are listed below.

- Chemicals testing /guidelines and guidance documents
- OECD Council Acts related to chemical safety
- Chemicals hazard/risk assessment
- Good laboratory practice
- Chemicals risk management
- Classification and labelling of hazardous chemicals and mixtures
- Existing chemicals
- New chemicals
- Chemical accidents

Global Information Network on Chemicals (www.nihs.go.jp/GINC/other/aboutginc.html)

The GINC, Global Information Network on Chemicals, is a worldwide information network for safe use of chemicals. From this homepage you can explore useful information sources provided by both international organizations and national institutions collaborating for safe control of chemicals.

Objectives:

- To develop a mechanism for providing information on chemicals from international organizations and countries electronically with access through INTERNET to any enquirer; and
- To promote harmonized collection in countries of data on chemicals and make this data available internationally.

Inter-Organization Programme for the Sound Management of Chemicals (IOMC) (www.who.int/entity/iomc/en)

This website provides information on the Inter-Organization Programme for the Sound Management of Chemicals (IOMC), which was established in 1995 to strengthen cooperation and increase coordination in the field of chemical safety.

The IOMC website is divided into sections on the Inter-Organization Coordinating Committee (IOCC) which coordinates and fosters joint planning of the IOMC, the seven Participating Organizations (POs) which contribute to the work of IOMC, Technical Coordinating Groups, and activities undertaken and events held by each PO.

With participating organizations, IOMC has a plethora of activities for promoting Chemical Safety. Some of its activities related to information management are given in the following Table.

Sl No	Activities	Output
1	Asia and the Pacific: Regional Network on Safe Pesticide	Promote safety in production and

	Production and Information	use of pesticides in order to protect farmers, producers, consumers and the environment
2	Chemical Information Exchange Network (CIEN) Project	Developing countries are the major stakeholders in projects, and benefited by: developing their chemical information charter, gained computers and other tools for accessing the Internet, trained staff. Enhanced communications among countries.
3	Clearinghouse on Sound Management of Chemicals	Advisory information (clearing house) service on chemicals. A number of websites and databases, that provides information on legal, technical, and scientific and policy issues relating to sound management of chemicals effectively.
4	Concise International Asessment Document (CICADs)	Concise scientific assessments of all relevant data to make an international assessment of the risks to human health & the environment.
5	Database on use and releases of Industrial Chemicals	This searchable database includes information on emission scenario documents and other information sources for estimating releases of industrial chemicals to the environment.
6	Health & Safety Guides	Health & Safety Guides (HSG) provide concise information in non-technical language for decision-makers on risks from exposure to chemicals together with practical advice on medical & administrative issues.
7	INCHEM CD ROM & Internet	To improve access to information on chemicals & to enable free & rapid access to internationally peer reviewed, authoritative chemical safety information, in accordance with Intergovernmental Forum on Chemical Safety.
8	Integrated Hazardous Waste / Chemicals Management Project	National Information system design specifying parameters for data collection; National Hazardous Waste/ Chemicals Inventory, Assessment of the transboundary movements of hazardous wastes,
9	International Chemical Safety Cards (ICSCs)	International Chemical Safety Cards provide essential health & safety information on chemicals to promote their safe use. They are used at the "shop floor" level by workers in employers in factories, agriculture, construction & other workplaces
10	Screening Information Data Sets (SIDS) for High	Improvement of access to

	Production Volume Chemicals	information and collected data under various projects by international and regional organizations to developing countries. IUCLID database establishment.
11	Web Database of Information on Computerised Tools for Predicting Health and Environmental Effects	This searchable database includes information on models (computerised or capable of being computerised) that are used by governments and industry to predict health or environmental effects exposure and risks.

National Institute for Occupational Safety and Health (NIOSH) - www.cdc.gov/niosh

The NIOSH Web site features many different types of databases and information resources. They are categorized here by Chemical; Injury, Illness & Hazards Data and Information; Publications; Respirators and other Personal Protective Equipment; Agriculture; and Construction. The most popular databases include the International Chemical Safety Cards, NIOSH Pocket Guide to Chemical Hazards, and NIOSHTIC-2. Some of the major databases related to Chemical Safety are given in the below table.

Sl No	Database	Scope
1	Immediately Dangerous to Life and Health (IDLH)	Provides the immediately dangerous to life or health air concentration values (IDLHs) for substances and the criteria and information sources that have been used to determine these values
2	International Chemical Safety Cards (WHO/IPCS/ILO)	ICSC cards summarize essential health and safety information on chemicals for their use at the "shop floor" level by workers and employers in factories, agriculture, construction and other work places.
3	Manual of Analytical Methods (NMAM)	NMAM is a collection of methods for sampling and analysis of contaminants in workplace air, and in the blood and urine of workers who are occupationally exposed.
4	NIOSH Pocket Guide To Chemical Hazards (NPG)	The NPG is intended as a source of general industrial hygiene information on several hundred chemicals/classes for workers, employers, and occupational health professionals
5	Occupational Safety and Health Guidelines for Chemical Hazards	Summarizes information on permissible exposure limits, chemical and physical properties, and health hazards. It provides recommendations for medical surveillance, respiratory protection, and personal protection and sanitation practices for specific chemicals that have Federal occupational safety and health regulations.
6	OSHA 1988 Permissible Exposure Limits (PELs)	PELs are OSHA comments from the January 19, 1989 Final Rule on Air Contaminants Project extracted from 54FR2332 et. seq. This rule was

		remanded by the U.S. Circuit Court of Appeals and the limits are not currently in force
8	Recommendations for Chemical Protective Clothing	Provides assistance in identifying potentially appropriate types of chemical barrier material for protection against skin contact.

Websites sponsored by National Safety Organisations :

Information about the profession of Safety, Health and Environment comes from the web sites of the various professional organizations. Most of the safety organisations functions as national bodies of their respective countries. Given the broad nature of Industrial Safety, there are a large number of such organizations. Some serve broader segments of the profession, while others have a more narrow focus. These web-sites include mission statements, information on joining the organization, and details on their annual meetings. Most of these web-sites provide industry news; Reviews; Publications; Conferences; Meetings and importantly web links to related sites of interest.

Some of the prominent National Safety Organizations are: American Board of Industrial Hygiene **(http://www.abih.org);** British Safety Council **(http://www.britishsafetycouncil.co.uk);** Canada Safety Council **(http://www.safety-council.org)**; Japan Industrial Safety and Health Association (JISHA) **(http://www.jisha.or.jp/english/html)** National Safety Council of India **(http://www.nsc.org,in)** etc.

Other Websites:

Understanding and knowing chemical properties ensures the chemical safety. ChemFinder (**http://www.Chemfinder.com)** a product of Cambridge Soft Corporation is a web resource of General Chemistry information, including empirical, physical/chemical properties, CAS Registry Numbers and Chemical Synonyms etc. This Web site also provides links to other Web pages with information related to the chemical of interest. The National Institute of Standards and Technology (NIST) database **(http://webbook.nist.gov/chemistry/)** provides thermochemical, thermophysical, and ion energetic data. The user interface accommodates searching by molecular formula, structure, CAS Registry number, and chemical name.

The Department of Energy's Chemical Safety Homepage provides a forum for the exchange of best practices, lessons learned, and guidance in the area of Chemical Management. Key topics addressed on this website include: chemical-related best management practices, lessons learned, chemical management tools, such as Material Safety Data Sheets, chemical hazard identification tools, chemical occurrences and chemical compatibility data and information, applicable Federal Regulations, Explosives Safety, and the work of the Chemical Safety Topical Committee (CSTC).

New York Committee for Occupational Safety and Health (NYCOSH) **(http://www.nycosh.org)** provides technical assistance, training and education about workplace health and safety hazards to workers, unions, community members and organizations. NYCOSH's web site provides news and information about on-the-job

safety and health, plus hundreds of links to more information that can be used to enhance occupational safety.

Conclusion:

There is a great treasure of Web-based information available over Internet. It is difficult to find a specific piece of information. A few suggestions for efficient use of the web,

1. use a good search engine and learn the search techniques/shortcuts for that search engine;
2. put only the most valuable sites on your favorites list;
3. take selective advantage of 'alerting' services to receive journal table of contents, meeting announcements etc., ; and,
4. take note of published 'Web Alerts' to find new and relevant sites.
5. Subscribe to discussion forums; listserves; bulletin boards etc

Issues like authenticity, data integrity, original sources of data, and context of the data have to be ascertained by the user community before deriving any conclusions out of collected data using the Web. Information fraternity has to become proficient to take the responsibility of evaluating the reliability of websites. Nevertheless, the Internet is a good starting point for locating state-of-the-art information.

References

1. Chakravorty (B B) : Role of Centralised Safety Department in a Multi-unit Organisation. *Industrial Safety Chronicle*, Vol 35(4) Jan-Mar, 1995 pp. 35-37.

2. Othmer (Kirk) : Encyclopedia of Chemical Technology, 3rd ed., Vol. 18, John-Wiley, N.Y., 1982, p. 60.

3. Raghavan (K V) and Koshy (A) : Chemical Industry in South India: Safety and environmental concerns and counter measures. *Chemical Industry Digest*, 2nd Quarter, June, 1992 pp. 135-142.

4. Ramaswamy (S S): Some Recent Developments in the areas of Safety, Health & Environment. *Chemical Industry News,* Vol 45 (2) February, 2000 pp. 187-188.

5. Gary N. Greenberg : Internet resources for occupational and environmental
6. health professionals., *Toxicology* 178 (2002) 263–269

7. Jacqueline Patterson and others: Human health risk assessment: selected internet and world wide web resources, *Toxicology 173 (2002) 123–143*

8. Christine L. Russom: Mining environmental toxicology information: web

9. Resources: Toxicology 173 (2002) 75–88

Web References are given in the text of the paper.

Information Technology Infrastructure Library - an overview

Vinod.J,

Abstract

ITIL - Information Technology Infrastructure Library - is a set of best practices and guidelines that define an integrated, process-based approach for managing information technology services. ITIL can be applied across almost every type of IT environment. ITIL began in the 1980s as an attempt by the British government to develop an approach for efficient and cost-effective use of its many IT resources. The best practice processes promoted in ITIL are supported by the British Standards Institution's Standard for IT Service Management (BS15000). The main focus of IT Service Management (ITSM) is generally divided into two main areas, Service Support and Service Delivery. Together, these two areas consist of 11 disciplines that are responsible for the provision and management of effective IT services. This paper briefly explains the features and benefits of ITIL.

1. Introduction

ITIL - the Information Technology Infrastructure Library - is a set of best practices and guidelines that define an integrated, process-based approach for managing information technology services. It was developed by the United Kingdom's Office of Government Commerce (OGC). ITIL was conceived in the late 1980s, by the mid of 1990s it had become a standard in service management. ITIL has become so popular as it is a public domain framework which is scaleable. The guidance is documented in a set of books which are defined by related functions like Service support, Service delivery, managerial, software support, computer operations, security management and environmental etc. ITIL can be applied across almost every type of IT environment.

2. Benefits of ITIL

ITIL emphasizes the importance of providing IT services to satisfy business needs in cost effective manner. ITIL approach can be adopted by any organization. The benefits of ITIL are

- Speedy responses to customer enquiries and complaints
- Pro-active approach to service provision
- Improved quality of IT-related information for optimal management and decision-making
- Better management and control over the IT system's infrastructure
- More effective and efficient usage of resources related to service provision and subsequent cost reduction potential

---

Computer Division, Indira Gandhi Centre for Atomic Research, Kalpakkam
E-mail: *vinu@igcar.ernet.in*

- Higher IT system users' productivity due to reduced down times
- Enhanced customer care and higher customer satisfaction
- Discovery and implementation of permanent solutions
- Reduced numbers of incidents
- Improved ROI of IT

3. The Library

IT service to many organizations is usually provided by an internal department and must have IT infrastructure which consists of Software, Hardware, Procedures for communication, Service people, documents etc. These assets and their use must be managed, hence the name IT Infrastructure Management. Together with IT service it becomes IT Service management. ITIL books cover functions of IT Service Management (ITSM). ITIL is organized into a set of books which are defined by related functions like Service Support, Service delivery, Managerial , Software support, Computer Operations, Security management etc. The OGC approached many different organizations for assistance and performed an editorial function. The books were written by one organization and quality assured by others. The ITIL books are of non-proprietary in nature. It emphasizes the best practice because, the books represents the experience of many IT professionals. The books are written to quality standards on par with ISO9001 requirements.

4. Library Coverage

The ITIL books cover functions of IT Service Management and provides cross-references to other books. Each book can be read and functions can be applied in isolation. However, more benefit is derived from viewing IT service Management as an inter related set of functions. The Service Management section of ITIL is made up of eleven different disciplines, split into two sections- Service Support and Service Delivery.

Service Support

- Configuration Management
- Change Management
- Release Management
- Incident Management
- Problem Management
- Service Desk

The **Configuration Management** provides the foundation for successful IT Service Management. The object of Configuration Management is to provide a logical model of the IT Infrastructure by identifying, controlling, maintaining and verifying the version of all configuration items in existence. It is used to account for all IT assets, to provide accurate information to support other Service Management process like Incident, problem, change and release management. The fundamental deliverable is the Configuration Management database (CMDB) which comprises the detailed information of all IT infrastructure components and other important associated assets. These assets that deliver IT services are known as Configuration Items (CIs). Ideally CMDFB also contains details of any Incidents, Problems, Known errors and Changes associated with each CI.

The object of **Change Management** is to ensure that standardized methods and procedures are used for efficient and prompt handling of all Changes, in order to minimize the impact of any related Incidents upon service. A single Change Management process, for the efficient and effective handling of changes, is vital to the successful operation of any IT organization. Changes must be carefully managed throughout their lifecycle from initiation and recording, through filtering, assessment, categorization, authorization, scheduling, building, testing, implementation and eventually their review and closure. Change Management is responsible for controlling Change to all CIs within the live environment. It is not responsible for change within ongoing projects, which are controlled by the project change process.

The object of **Release Management** is to take an holistic view of a Change to an IT service and ensure that all aspects of a Release, both technical and non-technical, are considered together. Release Management is responsible for all legal and contractual obligations for all hardware and software in use within the organization. Release Management establishes secure environments for both hardware and software.

The object of **Incident Management** is to restore normal service to operation as quickly as possible with minimum disruption to the business, thus ensuring that the best achievable levels of availability and service are maintained. It is responsible for the management of all incidents from detection and recording through to restoration and closure. Incident Management should be used to ensure the best use of resources to support the business, to develop and maintain meaningful records relating to Incidents, and to devise and apply a consistent approach to all Incidents reported.

The object of **Problem Management** is to minimize the adverse effect on the business of Incidents and Problems caused by errors in the infrastructure, and to proactively prevent the occurrence of Incidents, Problems and errors. To achieve this Problem Management assists Incident management by managing all major incidents and problems and record all workarounds and rising change to implement permanent structural solutions wherever possible.

The object of the **Service Desk** is to act as the central point of contact between the User and IT Service Management. To handle Incidents and requests, and provide an interface for other Service Management activities such as Change, Problem, Configuration, Release, Service Level and IT service Continuity Management. The Service Desk, unlike the other ten disciplines (processes), is a function essential to effective Service Management. More than just a Help Desk, it is the principal operational interface between IT and their Users. A good first impression by each of its Users is predicted upon its performance and attitude. Often a stressful place for staff to work, underestimating its importance, high profile and the skills required to perform the duties well can severely hinder an organizations ability to deliver high quality IT services.

Service Delivery

- Service Level Management
- Capacity Management
- Financial Management for IT services
- Availability Management
- IT service Community Management

The object of **Service Level Management** is to maintain and gradually improve business aligned IT service quality, through a constant cycle of agreeing, monitoring, reporting and reviewing IT service achievements and through instigating actions to eradicate unacceptable levels of service. SLM is responsible for ensuring that the service targets are documented and agreed in SLAs and monitors and reviews the actual service levels achieved against their SLA targets. SLM should also be trying to proactively improve all service levels within the imposed cost constraints. SLM is the process that manages and improves agreed level of service between two parties, the provider and the receiver of a service.

The object of **Capacity Management** is to understand the future business requirements (the required service delivery), the organisation's operation (the current service delivery), the IT infrastructure (the means of service delivery), and ensure that all current and future capacity and performance aspects of the business requirements are provided cost effectively. Capacity Management is responsible for ensuring that IT processing and storage capacity provision match the evolving demands of the business in a cost effective and timely manner. The process includes monitoring the performance and the throughput of the IT services and supporting IT components, tuning activities to make efficient use of resources, understanding the current demands for IT resources and deriving forecasts for future requirements, influencing the demand for resource in conjunction with other Service Management processes, and producing a capacity plan predicting the IT resources needed to achieve agreed service levels.

The object of **Financial Management** for IT Services is to provide cost effective stewardship of the IT assets and the financial resources used in providing IT services. Financial Management for IT Services is an integral part of Service Management. It provides the essential management information to ensure that services are run efficiently, economically and cost effectively. An effective financial management system will assist in the management and reduction of overall long term costs, identify the actual cost of services and their provision, provide accurate and vital financial information to assist in decision making, identify how IT adds value to the customers business, enable the calculation of TCO and ROI, make customers aware of what services actually cost (if appropriate), support the recovery costs, from customer if appropriate, in a fair and equitable manner, provide measurements of value for money, and provide incentives to produce quality services aligned to business needs, help influence customer behavior for example by providing incentives for using non-critical resources, encourage more efficient use of resources, provide better cost information and control of external contracts and suppliers, and to assist in the assessment and management of changes.

The object of **Availability Management** is to optimize the capability of the IT infrastructure and supporting organization to deliver a cost effective and sustained level of availability that enables the business to satisfy it's objectives. Availability Management ensures services are available when the Customer needs them and is influenced by business demand, the cost required to meet it, the configuration and complexity of the IT infrastructure including the level of redundancy, the reliability of the infrastructure and its components, and the levels of infrastructure maintenance, the processes and procedures used by IT services and the human factors and external events.

The object of **IT Service Continuity Management** is to support the overall Business Continuity Management process by ensuring that the required IT technical and services facilities can be recovered within required and agreed business time-scales. IT Service

Continuity Management is concerned with managing an organization's ability to continue to provide a pre-determined and agreed level of IT services to support the minimum business requirements, following an interruption to the business. This included ensuring business survival by reducing the impact of a disaster or major failure, reducing the vulnerability and risk to the business by effective risk analysis and risk management, preventing the loss of Customer and User confidence, and producing IT recovery plans that are integrated with and fully support the organization's overall Business Continuity plan.

5. Conclusion

The concept of library is changing. The use of IT in library environment has changed the way library functions. Nowadays the digital library and paper-less office etc. are much talked about. All Library and attached departments must follow ITIL frame works and best practices to improve the services.

6. References

1. www.itil.co.uk/
2. www.itil-itsm-world.com/
3. www.itil-toolkit.com/
4. www.pinkelephant.com/

Web Technology as an agent of Web services: Diffusion of Digital Information

Chudamani K.S.

Abstract

Diffusion of information is transfer of information. Web technology is the spreading activation of the usage of knowledge with in a population of users usually within a group characterized by some common element of productive activity such as farming, mining etc.

Web technology is an agent of Innovation and web services provision. This technological innovation will play an increasingly prominent role in organizational and in all business, social and personal relationships. Diffusion theory attempts to describe the process by which innovation is communicated through a channel, That is the Web in this case over a specified time period among members of the social system. This paper examines the diffusion of Web technology as an agent of web service.

Any Web service can interact with any other web service. Web services can run on any platform and can be written in any programming language.

In order that the diffusion is effective the role of Govt. may be found in scholar exchange etc., The role of information is viewed in relation to policy, regulatory changes, Technology diffusion assessment etc.,

1. Introduction:

The Innovation Process begins with the generation of new ideas, which result in a marketable product. The idea emerges from the fusion of the potential need in the market and the technological possibilities. Society is influenced by Web technology. This Technological innovation will continue to play an increasingly prominent role in organizational systems and in all businesses, social and personal relationships as it is the vehicle or agent of web services.

Technological innovation depends not only on the technological feasibility and economic profitability but also on its social acceptability. The idea of a tightly linked social network has been directly related to the diffusion of internal software innovation in the form of user-developed interaction.

Innovations and information are interconnected. Diffusion of innovation happens through the appropriate flow of information. When new products are being planned all producers have to keep up with the current state and future developments of society's aims and needs. For the diffusion of innovation information is needed. Now, let us see what is diffusion of information is.

Web services are applications to send data and instruction to one another, with no data translation or conversion required. Because internet is used the connection problems are minimized.

IISc, Bangalore-12 ; Email: ksc@library.iisc.ernet.in

The web service activity statement explain the W3c's work through

4. XML Protocol working group
5. Web service addressing work group
6. Web services descritpion working group
7. Semantic web services interest group

Innovation in web services is indicator of the impact of web technology on users.
Web methods has had a good early vision and broad support for web services in product . Web methods has substantial involvement in authorings, editing or chairing web services initiatives in the W3c.
Web services are likely to be very important of the next generation of distrubuted services. They are :

Internet : It is a way in which various applications and services are linked
Interoperability : Web services can run on any platform and can be written in an proramming language.
Low barrier to enrty: It is easy to understand, more flexible of linking applications
Ubiquity: Web services relay on HTTP and XML, any device that supports these technologies. Even wireless services can be providing using web services.
Industry support : All major vendors support and are involved with extending the standards around which web services are built.

The greatest benefir for most web services is the improved communication and collaboration that occurs with in an organization

8. Eliminate paperwork
9. Create a best practices
10. Share the lime light
11. Provide relevant news
12. Encourage multiple intranets
13. Provide effective search tools.

Web services are agents of information diffusion through web technology

2. Diffusion of Information

Diffusion of information is an information transfer process of knowledge, know-how procedures and facilities. It is the spreading activation of the usage of knowledge within a population of users , usually within a group characterized by some common element of productive activity such as farming , mining etc.,

The flow of information commences from the collection of innovative ideas which should reflect social aims and needs. The extent of diffusion of a technology at any given time is defined by the degree to which it is being applied at that time.

Technology is believed to be one of the major forces under pinning economic growth. Technology transfer system would lead to an efficient use of resources, information.

We need to develop a simplified data acquisition system for the process of assessment of technology to be transferred to developing countries. Such a system should aim to be both realistic as well as practicable sustainable. Socio-cultural values must be considered first

in the assessment of technology in developing countries. Technology diffusion may be viewed as a complex phenomenon involving the extent of 'use' and extent of 'time'.

It is well known that the successful technology diffusion programs that focus only on thorough, comprehensive and continuous follow up, providing linkages to other Government and private sector partnerships of other countries and are generally effective. Technology diffusion programs suffer from in effective interaction between the public sectors.

Diffusion and adoption of information technology is one that draws heavily on concepts developed in:

1) Marketing –based: Research dealing with diffusion and adoption of innovation.
2) Management –based: Research in innovation and knowledge management.
3) Applied science work in technology transfer.

The terms "Innovation" and 'Technology' are often used as synonymous terms.

According to Rogers diffusion is defined as: "The process by which an innovation is communicated through certain channels over time among the members of a social system. It is a special type of communication, in that the messages are concerned with new ideas. Communication is a process in which participant create and share information with one another in order to reach a mutual understanding.

The classic view of innovation diffusion defines four elements in the process.

4) Innovation: 'An idea, practice or object that is perceived as new by an
5) individual or other unit of adoption'.
6) The five main characteristics of innovation, as perceived by individuals are:

- Relative advantage
- Compatibility
- Complxity
- Trialability
- Observability

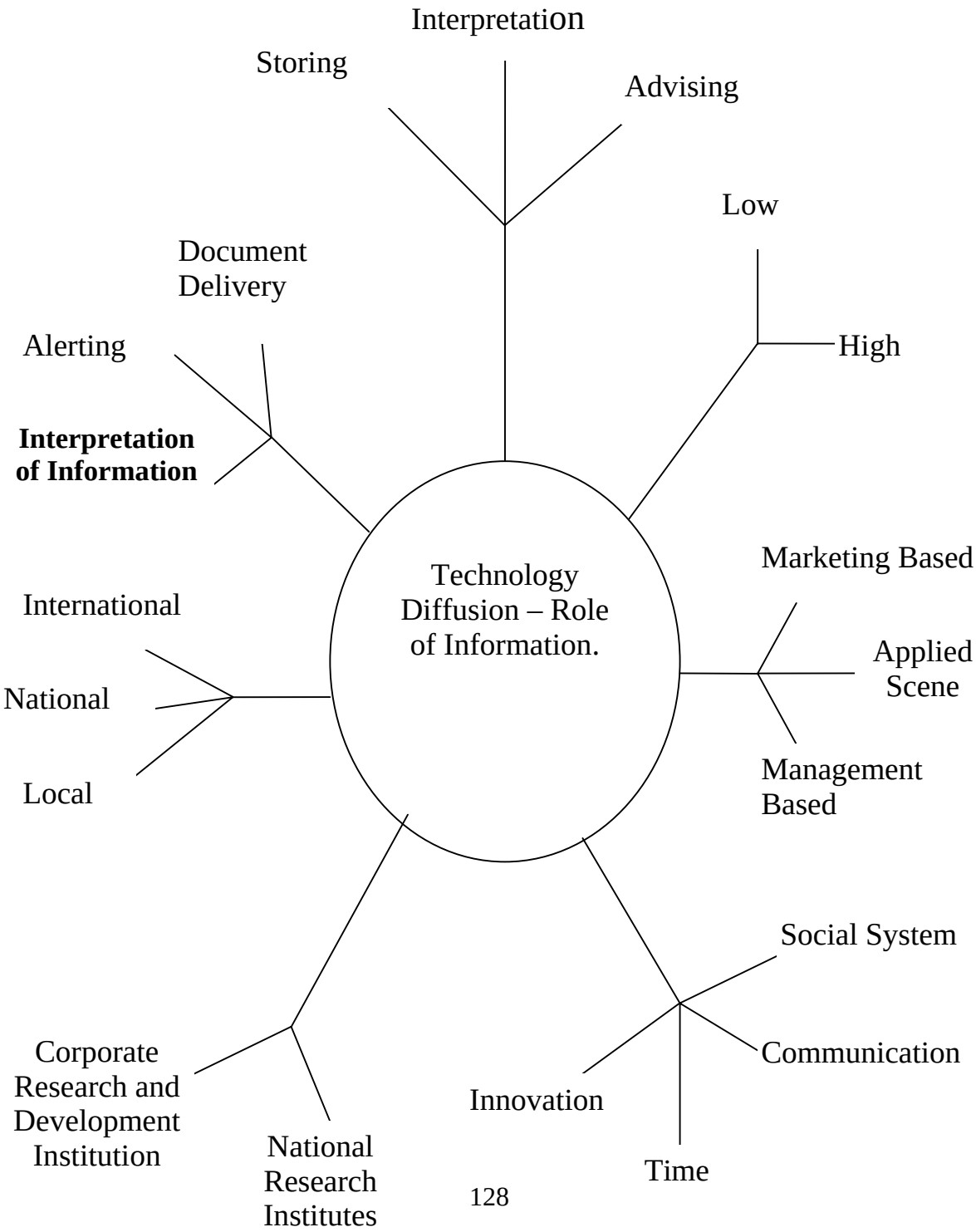
7) Communication channels
'The means by which messages get transferred from one individual to another' is a communication channel.
Two kinds of channels exist:

- Mass media channels: which enable to source to reach a broader
- audience.
- Interpersonal channels: for face-to-face exchanges.

Time There are three dimensions of significance in diffusion research:Time interval that spans knowledge of an innovation through its adoption or rejection.
Relative earliness or lateness with which an innovation is adopted.
Innovation's rate of adoption in a system.
Social system

'A set of interrelated units that are engaged in problem solving to accomplish a common goal'. Rogers defines 'The members or units of a social system may be individuals, informal groups, organizations, and/or subsystems'.
Keeping the above factors in view, this paper looks at the diffusion of web technology as an innovation

Diffusion Of Web Technology As An Innovation



Web is a new channel of communication one that supports mass communications as well as interpersonal communication. The web is also a new structure and space for human interactions. The web as a communications innovation may prove more effective than other communication technologies in helping to promote these new ideas and concepts to a broader audience.

Theories and models have been studies and developed that help us understand other aspects of the innovation diffusion process.

Technology grounded in microelectronics and operating on the basis of digital signals has led to a significant expansion of both telecommunication and computer industries. Now services are major users of information technology, accounting for up to 80% of investment in information technology hardware in countries such as U.K. and the United States and are accordingly changing their levels of capital intensity, sustainability and structure of their capital.

3. Varieties of Diffusion

Mansfield has conceived three varieties of diffusion. They are:

- Inter-firm diffusion: It refers to spread of new process from firm to firm within an industry. It is also known as imitation diffusion.
- Intra-firm diffusion : It refers to the spread or transfer throughout the industry as a whole
- General diffusion: However in general diffusion means inter firm diffusion which reflects the adoption and usage process between firms with an industry.

Co-operative plans are needed to provide national policy makers and businessmen with first hand knowledge of market based information. The potential of market oriented approaches to develop information transfer technologies diffusion projects are necessary in developing countries.

Diffusion model is very appropriate and useful to our broad consideration of how the web is used in developing countries.

4. Web Service Technology Diffusion for Information Transfer in Developing Countries and the Role of Government Institutions

Web technology is the agent of web service. This section looks at the role of govt. in enhancing information diffusion.
---To create national, regional and international networks for training and information exchange, which will involve the sharing of resources and exchange information among institution of the developing country.

---The mutual exchange of scholarly publications, manuals and scientific information.

---To introduce new communication technologies like the internet.

---To advise about possibilities of international technical and scientific contacts exchange programmes and sources of international information.

---The enhancement of inter and multidisciplinary work of teaching, research and social action regarding intra and extra university problems.

---To carry out joint scientific and technical programmers and actions between information centers and developing countries.

Once the role of government is established for facilitating technological innovation diffusion, the next aspect is direct understanding of the role of information.

5. Technology Diffusion in Developing Countries- Role of Information

Before any technology is introduced, an assessment of the technology is made by the govt. This needs objective information. This section points out the role of information in such an endeavor.

--- Policy and regulatory changes to increase information transfer (Local, National levels)

---To overcome the poor linkages from research and development to technology diffusion.

---Technology diffusion assessment to show opportunities being offered by the technology

---Increasing coordination among organizations concerned with information centers.

---To provide first hand knowledge of a particular information or institution and demonstrate its validity to increase market generation of information transfer technology.

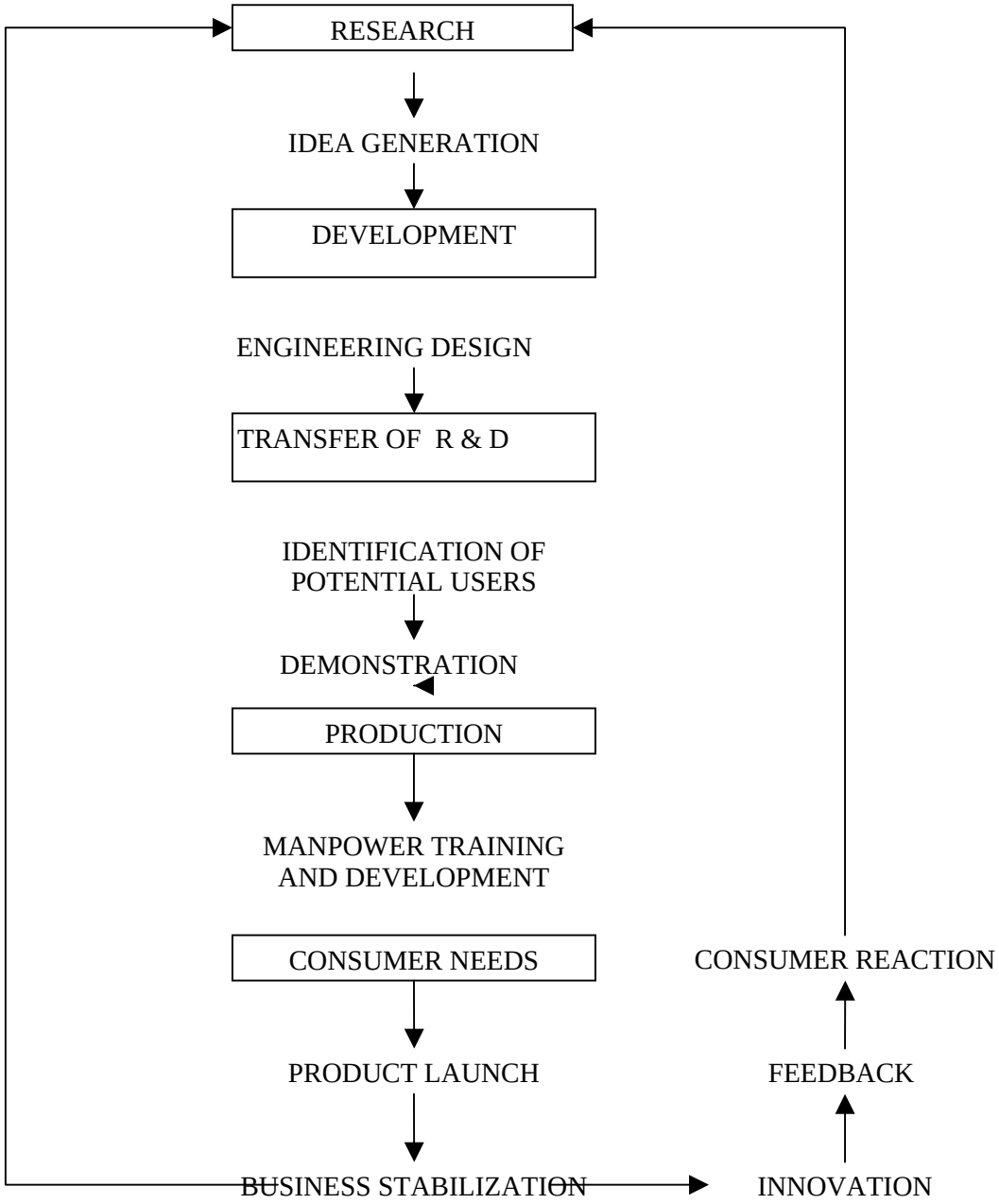
---To develop data collection mechanisms including the establishment of new or the extension of existing data correlation networks.

Because of the decreasing cost of hardware and software required for the web to spread, it has been activated by the developing countries. It involves lowest cost in terms of time and effort. . It is active in fast technology transfer and is therefore located close to the users and R & D work. A flow chart for the technology diffusion life-cycle is provided at the end.

6. Conclusion:

For diffusion of web services web technology is important. Role of government and information are vital, especially in developing countries. From the above exposition it can be noticed that web technology is playing a prominent role in information diffusion. But the information required for the innovation to diffuse is critical as pointed out in the paper.

Flow Chart For Technology Diffusion Transfer Process Cycle:



References

1. Diffusion of innovations : By Rogers, Everetton, (1995). NewYork, The free press. P 5-6

2. Systems approach to appropriate technology transfer.
ed by D.Fleissner. IFAC. 1983
3. Encyclopedia of Library and information science. V.69 P.89-105. 2001
(Supplement : 32)

4. Transfer of technology in Indian industry case studies on utilization of endogenous R & D ed by N.C.B.Nath and L. Misra. 1994 Indus pub.

5. Journal of the American society for information science and technology. V.55 No.7. 2004. A formal knowledge management ontology. By C.W.Holsupple and K.D.Joshi P. 593-612

6. Library Information systems. By Thomas R. Kochtanek and Joseph R. Matthews. 2004

Education through Video Conferencing with Special Reference to Engineering Colleges in Coimbatore District: A Proposal

Mahalakshmi K.[1] and Nirmala J.P.[2]

Abstract

The present-day digital environment has a great impact on the education system especially higher and technical education. It has become a component of education and at the same time facilitates advanced learning techniques with the help of technologies like e-learning packages. The libraries in educational institutions like IITs, IIMs, and NITs are extensively using these advanced learning techniques to enhance teaching, learning, and research. It provides an excellent means of delivering education and enables learning beyond the four walls of a class room. These advanced learning techniques are applicable more to engineering and technical education hence and the present paper makes a proposal of developing and using video conferencing as an educational technology by the Engineering Colleges of Coimbatore District.

Video conferencing uses telecommunications of audio and video to bring people at different sites together for a meeting. This can be like conversation between two people or with more than one person at different sites. Besides audio and visual transmission of people, it can be used to share documents and computer based information. In an Educational institution, video conferencing can provide cross-institutional links, value-added curriculum teaching, value added services such as professional development for staff. Video Conferencing is an excellent media for sharing and effective usage of not only the E-learning packages, but also to share expertise of specialists in their respective fields.

This paper suggests some cost-effective measures for setting up a Video conferencing facility in the colleges and some guidelines to be followed such that the colleges benefit by the advantages of the technique and enrich the learning experience of the students.

1. Introduction:

Video conferencing technology connects people at different locations so that they can communicate, interact, and share information. This connection is accomplished through audio and video links, which enable face-to-face and voice-to-voice communication. Connections are made using specialized equipment and those connections are carried over wiring using the Internet, satellite connections, or microwave signals. Videoconferencing provides students and teachers with the opportunity to expand their teaching and learning possibilities. Students and teachers will no longer have to rely exclusively on the resources available within their institutes. It is used in mass media, Government and Private sector and in health industry. But it is extremely useful for academic institutions in sharing the expertise, knowledge, lab and library resources. Students using videoconferencing will have the following advantages as stated in Wikipedia (1):

[1] *Librarian, Avinashilingam Deemed University, Coimbatore – 641 108*
[2] *HOD, Dept., Printing Tech., Avinashilingam Deemed University, Coimbatore – 641 108*

- Utilizing world-wide resources including experts, professionals, remote institutional resources, and other students
- Broadening the scope of learning resources
- Functioning in a global climate
- Enhanced motivation
- Improved communication and presentation skills
- Increased connection with outside resources
- Effective learning environments
- Expanded teaching capabilities
- Links to people from different social, cultural, and economic backgrounds
- Supports use of diverse media (Reed and Woodruff, 1995). Blackboards, handwritten documents and videos may be incorporated at all sites.

2. Objectives

Video conferencing is considered to play a vital role in technical and higher education. Many educational institutions in India and abroad have adopted this technique. However, in spite of its role to reach the masses, its use is restricted to a few institutions. Hence, a need has been felt to have a Consortium based Videoconferencing among Engineering colleges of Coimbatore District. The present paper is an attempt to study the feasibility and proposes a model of Consortium-based Video conferencing.

3. Application

Videoconferencing in Academic Institutions: According to Wikipedia Videoconferencing is defined as" transmission of synchronized image (video) &speech (audio) back and forth between two or more physically separate locations, simulating an exchange as if the participants were in the same physical conversation." Videoconferencing is a powerful communication tool that has the potential to change the way of delivering information to students. Videoconferencing is one of today's integrative technologies that help educators address the new educational challenges as they offer opportunities that empower students to learn and equip themselves with enhanced knowledge. Owing to its advantage, many academic institutions have adopted it as an integral part of their pedagogic methods. The libraries being centres to supplement classroom teaching and support independent learning are extending this facility to their student community. Video conferencing rooms, equipped with state of the art technology are attached to libraries that attract students for online and interactive courseware offered in their own subjects. To cite a few examples: Detroit College of law; University of North Carolina at Wilmington; University of Buffalo; The Telecommunication centre of Aristotle University, Thessalonica; East view Middle School in White Plains, New York; Jerome School District; Central Michigan University. In India, the libraries of all the IITs, IIMs, IIITs, IISc, Amrita Institute of Technology, Coimbatore etc. are using the Video conferencing facility for the following benefits:

- To access a variety of resources;
- To understand the importance of communicating and collaborating with content providers;
- To understand concepts for effective project management and collaboration;
- To share Expertise knowledge in the respective field;

- To access the Library and laboratory resources of other institutions;
- To add value to lessons and enable teachers to give students the opportunity to interact with experts without leaving the classroom;
- To enrich the Existing programmes;
- To provides additional access to students at remote sites.

4. Types of Video Conferencing

Primarily, four types of videoconferencing techniques are in vogue depending on the particular context (3)

4.1 Desktop videoconferencing

Videoconferencing on a personal computer with a small camera mounted on top of a monitor and microphone attachment. Computer software allows transmission of video, audio, and document sharing. This kind of videoconferencing is designed for one-on-one interaction. This is limited to LANs. It is less expensive, but offers limited resolution.

Components: The components of the Desktop Videoconferencing PC include: a personal computer equipped with an H.323 (Video compression standards for videoconferencing; used over networks that do not guarantee bandwidth, such as the Internet; enables interoperability between different vendor's implementations) based codec (A piece of software that codes and compresses the outgoing and decodes and decompresses the incoming audio and video signals), and video software with a suitably sized monitor, speakers/headphones, microphone, camera, and T.120(A standard for real-time data conferencing) data sharing software for video conferencing(4).

4.2 Broadband

Satellite connection with studio-quality equipment produces an excellent full-motion video connection, but the equipment and transmission expense is huge.

4.3 ISDN (Integrated Services Digital Network)

Many connect via ISDN, because it is an economical solution for high-quality videoconferencing. Services over regular copper telephone lines, working and transmitting at a minimum of 128 Kbps per line, it provides dedicated bandwidth for smooth audio and video (15-30 frames per second). The preferred rate for events is at least 384 Kbps. Recent advances in computer and telecommunications technologies have sparked an interest in 4th generation compressed video systems, which transmit information via today's Internet or telephone network, greatly reducing the cost of videoconferencing.

4.4 IP videoconferencing

In contrast, an Internet-based connection shares or competes for bandwidth with other Internet data, which may cause audio clipping or delays resulting in jerky video on slow networks. Many libraries and other institutions are developing high bandwidth networks and are experiencing better results with IP videoconferencing. Many believe that the Internet will eventually replace ISDN as the medium of choice for videoconferencing.

5. Proposed Model

As has already been stated, the present paper proposes a model video conferencing facility for the engineering colleges in Ciomabatore district. The proposal covers Prerequisites, Participating Members, MOU and Guidelines.

The district of Coimbatore has 21 engineering colleges that function under government as well as private managements.

The ISDN model is considered to be more suitable to establish a video conferencing facility among the engineering colleges at Coimabatore. ISDN uses highly compressed digital video for transmission of motion images. The video compression process decreases the amount of data transmitted over the lines by transmitting only the changes in the picture. By minimizing the bandwidth required to transmit the images, video compression also reduces the transmission cost. It is often transmitted on dedicated T-1 phone lines (A higher speed telephone circuit for dedicated connections capable of 1,544,000 bits/second). These high speed lines are very effective for videoconferencing, but they are typically leased circuits with an expensive monthly cost. The fixed monthly charge is usually based on distance, not usage. Therefore, the cost effectiveness of this system increases with use. Interactive videoconferencing systems can operate at different data rates, at various fractions of T-1 capacity, enabling the transmission of multiple simultaneous videoconferences over the same T-1 circuit. A 4th generation system can also share a T-1 circuit with other digital data uses such as Internet transmissions or file transfers.

5.1 Prerequisites for implementing Video conferencing

5.1.1. Hardware

Basic Equipment-Sender such as

Camera: Videoconferencing systems include at least one camera that will show a presenter. It can be adjusted through the videoconferencing software to capture the scene in a classroom

Document camera: Document cameras are used to project small items, hard copies of documents, artifacts, or other conference relevant materials that the students need to see close up. The document camera gives a close up to each participant and allows students at the near and far end to see, examine, and discuss items at the same time. A simple manipulation of the videoconferencing software menu will switch the view to the document camera.

Microphones: There are a number of microphones available for videoconferencing:

Lapel Microphone: Commonly used, clipped onto a jacket or shirt and worn on a belt. It is wireless and allows movement around the classroom.

Desktop microphones: Flat that sits on the table and picks up the voices as the speakers engage in the videoconferencing session.

Basic Equipment-Receiver such as

Monitor: In a Stationary videoconferencing room, there generally are at least two monitors. These monitors show a view of the distant room and the originating room. It is better to have double hung monitors in the back and this enables the teacher to see the far end room and also view from the broadcasting room.

Whiteboard and Projectors: An interactive whiteboard coupled with a projector helpful in enlarging the image. A large screen gives a greater impression with the far end participants and has a greater presence in the room. An interactive screen gives the added capability to allow the teacher, the ability to work from the touch screen to run the conference.

Speakers: Wall mounted speakers project the audio so that all participants can adequately hear the videoconferencing session.

Basic Equipment: Controllers includes

Keyboards: To manage the videoconferencing software; to adjust the camera angles; to access documents to be shared

Remote Control Devices: A number of devices to control the videoconferencing equipment may be used. Touch screen panel that is operated from the teaching station or can be a remote device that the teacher can use while moving about the room.

Basic Equipment: Processors that include

Codec: coder/decoder - A piece of software that codes and compresses the outgoing and decodes and decompresses the incoming audio and video signals.

Router: A device that connects networks

Bridge: A networking device that connects several video sites into one videoconference

5.1. 2. Software

There are several softwares available for videoconferencing. One of the popular software is Picture Tel, which is user friendly, self explanatory, and makes the use of the video systems an enjoyable experience.

5.2. Participating members

Though Videoconferencing has many advantages; the initial cost of it may be expensive. To share the resources and equipments it is proposed to have consortia based Video conferencing model. Due to Geographical limitations, it is proposed to have the Consortium with the following institutes as members.

1. Avinashilingam Deemed University, Coimbatore-641043.

2. Faculty of Engineering, Avinashilingam Deemed University, Coimbatore-641 108.

3. Kumaruguru College of Technology, Coimbatore-641006.

4. Government College of Technology, Coimbatore-641013.

5. Sri Ramakrishna Engineering College, Vettamalaipalayam, Coimbatore-641-022.

5.3. Memorandum of Understanding and Guidelines

A MoU among participating libraries is essential for effective and efficient utilization of the service and accrue the optimum advantage out of consortium. Hence the MoU needs to include the following vital aspects:

A model is given below.

1. Central agency (proposed that Head of the institute of Avinashilingam Deemed University, Coimbatore, shall be the Chairman) and member institute libraries.

2. An agreement among members should include the terms related to

    - Sharing of Resources: Human Technical &Financial
    - Kind of Resources: Lab and Library Resources
    - Duration of sharing – one/two/more years.

Infrastructure in the member institute: The Member Institute has to maintain the basic equipment such as

a) ISDN line with 128 kbps

b) Camera

c) Codec

d) Micro phone

e) PC-to display PC based presentations and document sharing.

f) Projector –to project the conference on a bigger screen.

g) Document camera for graphics transmission

h) VCR-To record for future reference

i) Slide to video converter-for 35-mm slide input for making presentations.

j) White board: A white board for capturing the images.

k) Knowledge of Expertise: Sharing the Knowledge of Experts in their respective fields.

l) Monthly Recurring Charge: Monthly recurring charge for telephone lines and human Resources.

5.4 Instructional Strategies

Reed and Woodruff, 1995 opined the following instructional strategies while establishing Videoconferencing in academic institutions.

Establish Class Expectations: Some students may adopt the "TV attitude", expecting the lecture to be entertaining, not educational. This can be overcome by well-planned and focused presentations on teacher-student interaction.

Reduce Distractions: Students can also be informed to be minimise extraneous noise and activity.

Encourage Dialogue: It is better to have an interactive session rather than one-way.

Avoid using small fonts and light colours while preparing class visuals. A variety of formatting will also assist in maintaining student interest and attention. When formatting visuals, be sure that they will fit on the TV monitor.

5.5 Role of Librarians

The cost of books and journals are escalating every year, but the annual budget remains the same. It is the primary function of the librarian to serve the needs of their users in the present day scenario. So, the librarian of a Member institute can perform the role of moderator in establishing and delivery of lectures through videoconferencing in their respective institute.

Moreover he can perform the following duties to make use of the videoconference facility to its fullest maximum for their respective institute.

Orienting the students and technicians about the features of all equipments.

To provide the users (refer to students and instructors) with a reference sheet outlining major functions.

It is better to have a technician for assisting in setting up and monitoring exclusively for videoconference.

Training shall be given to the instructors in performing all the necessary functions for the proper operation and utilization of the videoconference facility and observing all the precautions needed to keep the system safe and working.

5.6. Limitations of Videoconferencing

In spite of the advantages, the videoconferencing to be offered by libraries as a pedagogic service has certain limitations in its implementation. The initial cost of the equipment and leasing the lines to transmit conferences may be prohibitive.

Companies which produce codec (Coder/Decoder) have developed unique methods of compression that are incompatible, although protocols have been established to allow communication among brand names. However, this "universal standard" compromises resolution and quality to a certain degree.

If visuals, like handwritten or copied materials, are not properly prepared, students may have a difficult time reading them.

If the "pipe" that carries the transmission among sites is not large enough, the students may observe, "ghost images" when rapid movement occurs in "real time" (Reed and Woodruff, 1995).

If the system is not properly configured, class members may observe an audio "echo" effect (Reed and Woodruff, 1995). The result is audio interference that detracts from the learning environment.

6. Conclusion

The video conferencing plays a vital role in technical education, as a number of courseware is available online that are being imparted through interactive mode. The academic libraries have realized its benefits and introduced this facility in their institutes. However, in India, the service is limited to elite academic institutions and there is a need for other technical institutions to respond to the situation and adopt this advanced pedagogic technology. There are difficulties for individual colleges to implement such a costly programme; therefore consortium-based application may benefit many students with the innovative teaching/learning technique.

References:

1.www.wikipedia.org

2.Reed, J and Woodruff, M (1995). Using Compressed video for Distant Learning <http://www.kn.pacbell.com/wired/vidconf/using.html>

3.Woodruff, M & Mosby, J (1996) A brief description of Videoconference. Videoconference in the classroom and Library. <http://www.kn.pacbell.com/wired/vidconf/description.html#what>

4. http://www.vidcon.nic.in/faq.htm

5.Galbreath, J (1995) Compressed Digital Videoconference. Educational Technology; 35(1), pp35-38.<http://www.edutec.net/textos/Alia/WILLIS/dist11.htm>

Knowledge Management

Knowledge Discovery Tools and Techniques

Soundararajan E,.Joseph J.V.M, Jayakumar C. and Somasekharan M.

Abstract

As digital information is growing in organizations, massive amount of data are generated which exist in database systems. This explosive growth of information requires new analysis techniques that can intelligently transform the useful data into knowledge. Knowledge discovery is defined as the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. It uses machine learning, statistical techniques and visualization techniques to discover and present knowledge in a form that is easily comprehensible. Knowledge discovery takes the raw results from data mining and carefully and accurately transforms them into useful and understandable information. This paper gives a technical overview of knowledge discovery tools, and techniques. It emphasizes the need for a knowledge management system for digital library. Finally the suggestions for improving the performance of knowledge discovery tools are highlighted.

Keywords: Data Mining, Pattern, Knowledge discovery, Knowledge Discovery Tools, Digital Library.

1. Introduction

Knowledge management is a systematic process of acquiring, organizing, sustaining, applying, sharing and renewing both tacit and explicit knowledge to enhance the organizational performance, increase organizational adaptability, increase values of existing products and services and/or create new knowledge intensive products, processes and services. As the digital libraries become more knowledge conscious, knowledge discovery and data mining become essential for knowledge management. The information management system should have data mining capabilities for helping the knowledge discovery process and subsequently to evolve as a knowledge management system.

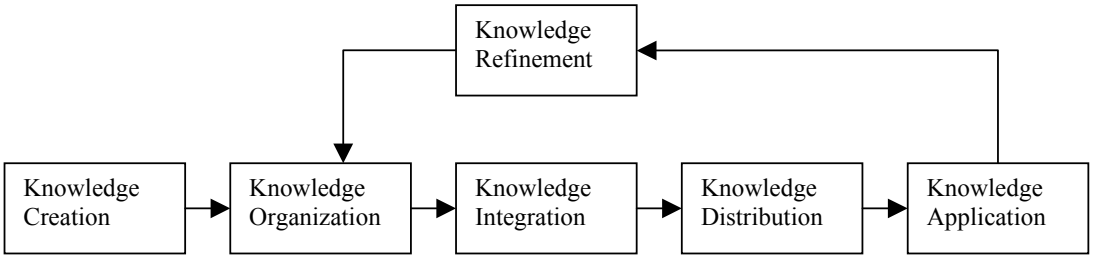


Fig1. Knowledge Management Process

Scientific Information Resource Division, Indira Gandhi Center for Atomic Research, Kalpakkam-603102 , Email:Sound@igcar.ernet.in

2. Knowledge Discovery Process

Knowledge discovery is an activity that produces knowledge by discovering it or deriving it from existing information. Knowledge is then organized by indexing knowledge elements, filtering based on content and establishing linkages and relationship among the elements. Subsequently, this knowledge is made available to users for supporting their decision making process.

Knowledge Discovery refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining involves a collection of tools and techniques for finding useful patterns relating the fields of very large databases. The newest form of data mining is the linguistic summarization of data which aims at a computer generated verbal description of the knowledge implicit in a database often in the form of 'if –then' rules that resemble fuzzy knowledge granules. Text mining is to extract patterns from textual documents. A text mining technique typically involves text parsing and analysis to transform each unstructured document into an appropriate set of features and subsequently applies one or more data mining techniques for extracting patterns.

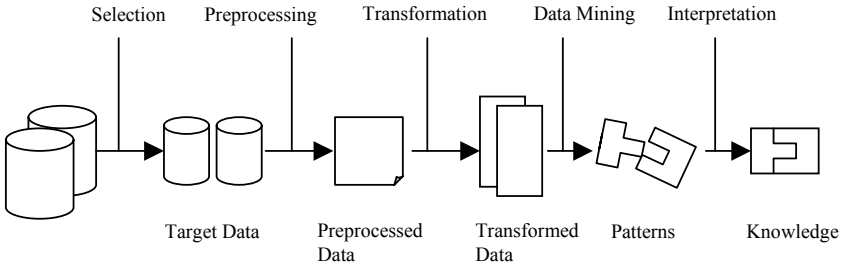


Fig 2: Knowledge Discovery Process

The knowledge discovery process is interactive and iterative and involves the following steps

1. *Learning the application domain:* Includes relevant prior knowledge and goals of the application
2. *Target data set*: Selecting a data set or data samples on which discovery is to be performed
3. *Data Cleaning and Preprocessing:* Removing noise, deciding on strategies for handling missing data fields, accounting for time sequence information and known changes and mapping missing and unknown values
4. *Data Reduction and Projection:* Includes finding useful features to represent the data and using methods to reduce the effective number of variables under consideration
5. *Choosing the function(s) of data mining:* Selecting the data mining function based on data model such as summarization, classification, regression and clustering.
6. *Choosing the data mining algorithm(s):* Includes selecting the method(s) to be used to search for patterns in the data such as statistical algorithms, visualization techniques, deviation trend analysis decision tree analysis etc. Two or more techniques can be combined depending upon the data models.
7. *Data mining:* Concerned with applying computational techniques to find patterns in data in a particular representational form or set of such representations. A Pattern that is interesting and certain enough can be treated as knowledge.

8. *Interpretation:* Includes interpreting the discovered patterns and possibly returning to any of the previous steps as well as possible visualization of the extracted patterns, removing redundant or irrelevant patterns and translating the useful ones into terms understandable by the users.
9. *Using the discovered Knowledge:* Includes incorporating discovered knowledge into the performance system, taking action based on the knowledge or simply documenting it for management/later use.

3. Knowledge Discovery Techniques

The Knowledge discovery can be of two categories: descriptive knowledge discovery and predictive knowledge discovery. The former describes the data set in a concise and summary manner and presents general properties of the data; whereas the later constructs one or a set of models, performs inference on the available sets of data and attempts to predict the behavior of new data sets. The features of knowledge discovery includes

- Large Amount of data
- Efficiency
- Accuracy
- Automated Learning
- High Level Language
- Interesting Results

3.1 Probabilistic Approach

This method utilizes graphical representation models to compare different representations. Visualization tools are a class of advanced graphical presentation tools that facilitate data exploration, hypothesis development, testing and evaluation of new or existing theory of knowledge. Using appropriate computer technology, a skilled and well-motivated human user can directly visualize patterns connection, correlations or lack thereof, among parameters that have been measured or calculated. While this parameter can be codified in numbers, visualization brings the enormous power of human perception to bear in entirely new ways. One of the important uses of visualizing complex relationships among several variables could be to understand the nature of empirical data in order to select the appropriate techniques for further analysis.

3.2 Statistical Approach

The statistical approach uses rule discovery and is based on data relationships. An inductive learning algorithm can be used to generalize patterns in the data and to construct rules from the noted patterns. Online analytical processing (OLAP) is an example of a statistically oriented approach.

3.3 Deviations and Trend Analysis

Pattern detection by filtering important trends is the basic for this approach. This is normally applied to temporal databases, such as analysis of traffic on large telecommunication networks.

3.4 Classification approaches

This approach is based on grouping data according to similarities and classes. Bayesian approach that uses probabilities and a graphical means of representation is considered a type of classification. Bayesian networks are typically used when uncertainty associated with an outcome can be expressed in terms of a probability. This approach relies on encoded domain knowledge and has been used for diagnostics systems. Other examples of classification approaches are decision tree approach and pattern discovery and data cleaning models. Decision trees are hierarchical structures, where each internal node contains a test on an attribute; each branch corresponds to an outcome of the test, and each leaf node gives a prediction for the value of the class variable. Classification approach is useful for organizing the potential metadata of digital library for knowledge generation process.

4. Knowledge Discovery in Digital Libraries

The Library has been the center of the preservation, utilization and distribution of information and knowledge. Digital Library has a much greater capacity for Knowledge Management. The current Digital Library Architecture should include classification and thesaurus – the vocabulary control and knowledge organizing tools, which serves three purposes in a traditional library, the description, organization and retrieval of information. For more effective and efficient exploration, the networked information should be pre-arranged together with vigorous improvement of search techniques. Classification and thesauri that contain condensed intelligence can be used in organizing networked information especially metadata to facilitate the information resources usability and catalyze the Digital Library into Knowledge Management.

Classification and thesaurus can be merged into a concept network and the metadata can be distributed into the nodes of the network according their subjects. The abstract concept node substantiated with the related metadata records becomes a knowledge node. This forms a consistent knowledge network that is not only a framework for resource organization, but also a structure for knowledge navigation, retrieval, and learning.

The bibliographic data is one of the most important resources of library, which will be useful for knowledge discovery process. Based on the subject indexing, the bibliographic data can be combined with the classification and thesaurus to form a knowledge structure, which provides a skeleton for organization of bibliographic data. Corpus knowledge can be formed when new terms can be extracted automatically from the bibliographic data to update the classification and thesaurus. Such a knowledge network provides the user with an opportunity for navigation, searching and learning.

5. Conclusion

Knowledge discovery process has evolved and continues to evolve from the intersection of research fields such as machine learning, pattern recognition, database statistics, Artificial Intelligence etc. While Knowledge Discovery tools hold the promise of an enabling technology that could unlock the knowledge lying dormant in huge databases, they suffer some shortcomings such as problems in representing multiple interrelated relations in databases, incremental rule generation when database is expanded, and finally consistency

and accuracy of the generated rules. Some suggestions for optimizing the performance of data mining systems are

- Reducing the size of dataset considered for discovery
- Avoid discovering inconsistent, redundant, trivial knowledge,
- Defining more efficient data mining algorithm and
- Employing high performance systems to improve the discovery process.

Through the application of Knowledge discovery tools, Digital library will evolve as dynamic landscape for knowledge creation, dissemination and management

References

1. Saso Dzeroski and Nada Lavarac. *Relational Data Mining*. Berlin:Springer, 2001.ISBN 3-540-42289-7.

2. Cornelius T. Leondes. *Knowledge-Based Systems Techniques and Application (Vol-2)*. San Diego: Academic Press, 2000. ISBN 0-12-443877-6.

3. *Knowledge Discovery In Databases: Tools and Techniques*. Preggy Wright .ACM, 1998. URL:< http://www.acm.org/crossroads/xrds5-2/kdd.html>

4. C.W.Holsapple. *Hand Book on Knowledge Management (Vol 1& 2)*. Berlin:Springer, 2003. ISBN 81-8128-038-5.

5. *Analysis of Data Mining Algorithms*. Karuna Pande Joshi. URL< http://userpages.umbc.edu/~kjoshi1/data-mine/proj_rpt.htm>

6. Jun Wang. *A Knowledge Network Constructed by Integrating Classification, Thesaurus, and Metadata in Digital Library*. Intl. Inform. & Libr. Rev. Vol35 Issue 3. 2003, pp383-397.

7. Lorcan Dempsey, Derek Law and Ian Mowat. *Networking and the future of digital libraries*. London: Library Association Publishing, 1997. ISBN 1-85604-241-3.

8. 8.Shu-hsien Liao. *Knowledge Management Technologies and Application Literature review from 1995-2002.* Expert Systems with Applications. Vol.25, Issue 2. August 2003, pp 155-164.

9. Usama Fayyad et.al. *From Data Mining to Knowledge Discovery in Databases.* American Association for Artificial Intelligence. URL:< http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>

Scientific Data Warehouse and Visualization Techniques

Jayanthi T., Ananthanarayanan S. and Rajeswari S.

Abstract

R & D Organisations handling many Research and Development projects produce a very large amount of Scientific and Technical data. The analysis and interpretation of these data is crucial for the proper understanding of Scientific / Technical phenomena and discovery of new concepts. This paper discusses the Main Data Centers available at IGCAR, Design and Development of Scientific & Technical Data Warehouse, its importance and Visualization techniques leading to knowledge discovery.

1. Introduction

All these years the main concentration was given to administrative and accounts data and less thought was given to the content rich data emerging out of our Research and Development areas which can amount to Peta bytes of data. Probably, drilling down approach would through some light on our thinking process leading to new concepts and better understanding among the Scientific & Technical Community. It is all the more necessary for Research and Development Centres like ours to design & Develop a Scientific and Technical Data warehouse that can aid Researchers & Analysts across the country / globe to carry out Collection Based Research programmes.

2. Scientific and Technical Databases

The exponential growth in Science and Technology has lead to generation of larger amount of valuable data through experiments and theoretical research. Scientific & Technical data are growing in volume through basic / applied research in various fields, experiments, design & analysis, computer simulation, plant operations etc. with the complexity of data increasing at a staggering rate. But sufficient technology to handle or extract fully the latent knowledge within the data is still in the evolving stage. To create Large Scientific & Technical Databases (i.e. Data Warehouse) creation and reconciliation of data object and establishing metadata standards are very important. Further defining metadata semantics, creating discipline specific data dictionaries, information models for organizing metadata, and data models for describing data set structure also need to be worked out.

Detailed study on the aspects of distributed / large databases including data clustering and caching; data redundancy, dynamic summarization, and query formulation to allow machine optimization, replication of data among various units, protocols for high-speed, parallel dataflow are to be carried out. While creating Scientific & Technical Databases, the Preservation & Security mechanisms of data are also need to be given serious consideration in order to coexist

---

Computer Division, IGCAR, Kalpakkam; E-mail: jayanthi@igcar.ernet.in

with other security mechanisms already in service.

Collecting vast amount of data and creating useful Scientific and Technical database is a challenging task. Since our centre is engaged in Research & Development activities, the vast data generated out of Plant Operations, experiments and analysis can be well utilized and Large data Repositories can be designed and developed.

The following diagram shows the probable Data Centres within our Organisation from where the Scientific and Technical data can flow towards building up of a Large Scientific & Technical Database.

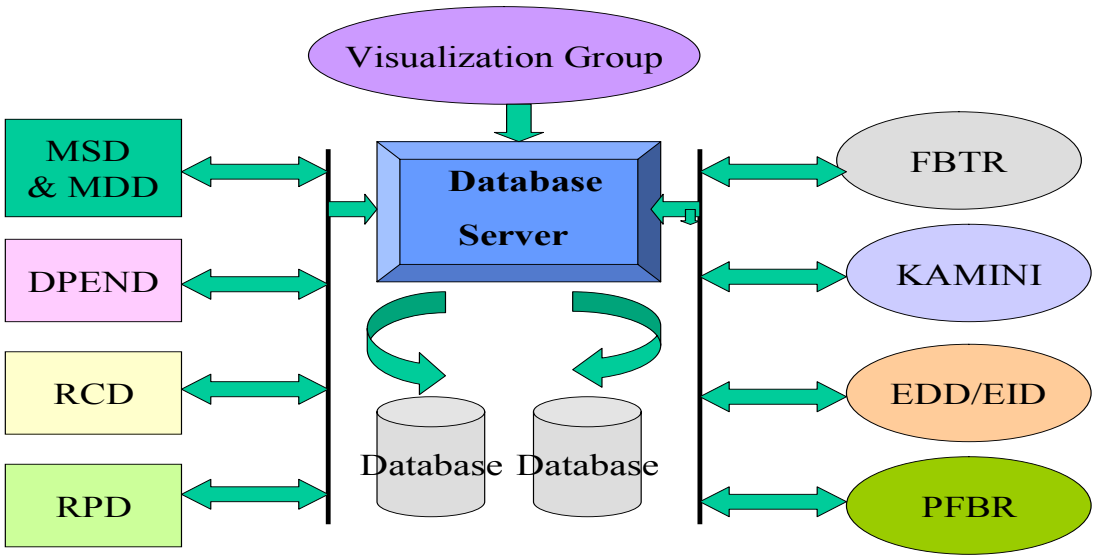


Fig-1. Data Centres within our Organisation for building up of Large Scientific & Technical Database.

3. Important points to be considered

Last decade has witnessed a thousand fold increase in the computer speed, significant increase in the performance, a drastic decrease in the cost of computing and innovations in data storage capabilities, and widespread access to high-speed networks. In order to make use of the available technology to yield maximum benefit to the society, analysis and manipulation of these data through effective ways are very important.

Scientific data is getting not only larger but its complexity is also increasing and extraction of meaningful knowledge requires more and more computing resources. Associated challenges in creation and use of large Scientific & Technical databases are the tools for discovery, searching, filtering and analysis. Most existing tools do not scale to terabyte or petabyte datasets.

More thrust on the area of building strong and intelligent querying system of the datasets is the need of the hour to guide the present Scientific and Technical community. To

efficiently handle terabytes of data, database engines with fast I/O speeds and advanced query engines, that can access geographically distributed data are required.

4. New Concepts and New Ideas

The typical way to proceed with analyzing large datasets is to first generate some abstraction of the dataset in the form of features or summarization. Scientists & Technocrats then use this information to guide them to the regions of interest. In this exploration phase, they often wish to "drill down" in the dataset. This is done by specifying a subset of the dataset, either directly, or by the summarization features, repeatedly refining the focus. Once something interesting is found, there may be a long run of the computer for pattern-matching or other data-mining to find other places in the dataset which "look like that". Finally, the results of the search can be used to create new knowledge. Following diagram indicates the approach to new concept:

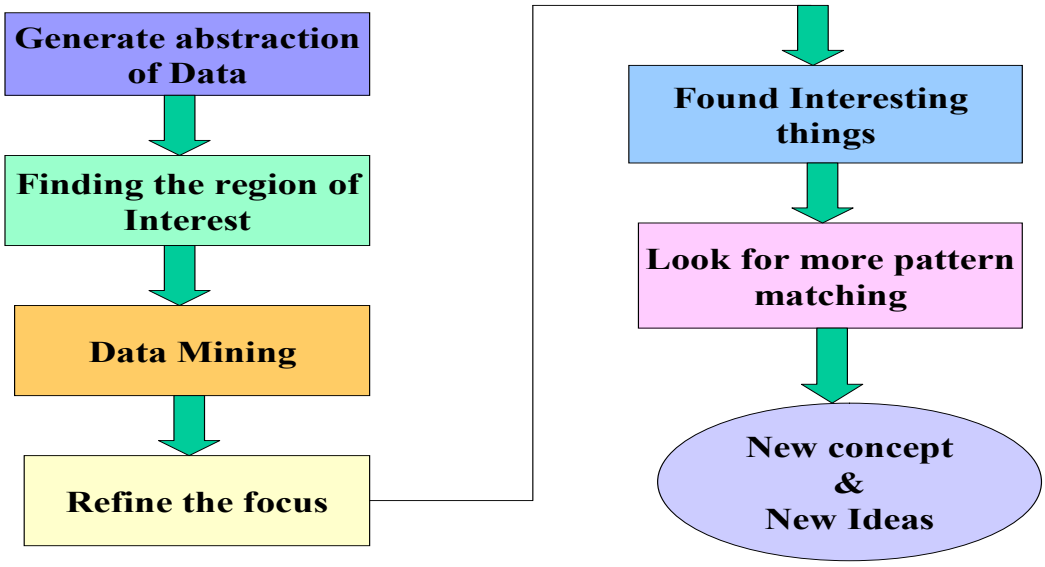


Fig – 2. Generation of new concepts & new Ideas

5. Collection Based Research

Collection Based Research (based on Large Scientific and Technical Database) is a promising and emerging field wherein Scientists and Technologists from remote corners can participate in the research programme across the country / globe, making use of the available infrastructure and internet facilities to their full advantage. These remote scientists and technologists with a mixture of computing equipment on their desks, need catalogues and indexes of the data archive, the ability to select data objects, to define complex processing to be done, then choose how the results are to be returned to them. In addition, adequate authentication systems are needed to control access to the data across multiple security domains.

As simulation was added to the traditional scientific & engineering cultures, collection-based research will also play an important role in acquiring new concepts. The experimental results can be catalogued and made available to others on the Internet. The collection-based scientist will reduce, mine, and sift that data to make or break a hypothesis from the theorist. In this new paradigm, the raw or nearly-raw data is published in a way that would be impossible with only paper journals, not just tables and graphs, but rather ramified palaces of logic. These would be subsequently reduced and interpreted by other researchers. Thus the person who takes the data may be different from the person who reduces it to small, palatable representations that can be printed on paper.

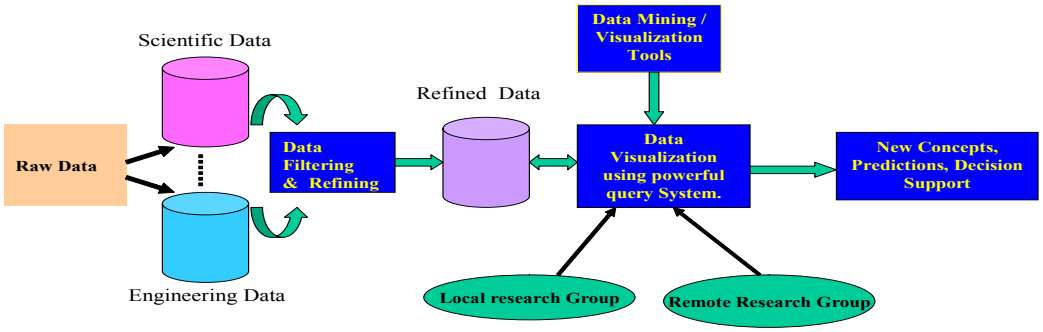


Fig. 3. Utilisation of refined data by Local / Remote Researchers to discover the hidden truth/ideas/concepts

In this way the large Scientific & Technical repositories can have more access by more scientists and open more of the scientific process to examination by the scientific community. This concept is particularly important where the basic infrastructure and experimental facilities are not available.

6. Database Interoperability

There is an urgent need to advance Science & Technology not only by using, but also by fusing information from multiple sources, from multiple digital archives. The greatest leverage will come if the integration is not only within the lines of established disciplines, but across these lines. Interoperability of scientific databases will drive global collaboration in the scientific and Technical worlds.

Scientific & Technical Databases of various DAE Units can be fused to form a single Virtual Database to aid collection based research across the units of DAE.

7. Virtual Data Centres

Virtual Data centers can be created using the huge collection of Scientific & Technical data across the units. Collection based research can be done using the strong and intelligent query models suiting the research of scientific community for proper visualization of data to unearth the hidden facts.

Data-handling systems for scientific data must be very flexible, since it is often not known how the data will be used when the system is being designed. Scientific users often want to write their own programs and run these on a server; besides security problems, designers must try to accommodate a variety of languages and a variety of levels of expertise.

The scientific community, with its focus on research, has pioneered many concepts that were eventually accepted by the commercial world and the same results can be expected out of research carried out using the Virtual Data Centres.

8. Preservation, Security and Authentication of Databases

The period for which the scientific and technical data be preserved depends on the value of the data into the future, and also on the cost of maintaining the data in a usable state. Sometimes, the content of a database can be reproduced at reasonable cost, in which case there is no need to worry about long-term preservation.

It may also be that the scientific value of the data can be completely extracted within a year or so of the data being created, in which case it can, in principle, be deleted. But the reality is that we cannot always be sure that there is nothing more. Creators of scientific databases should be encouraged to consider in advance how the database will be used in the future. To facilitate preservation, there is a need for a clear designation of what constitutes the boundaries of a given information object, and a need to resolve pointers within the data object to other data, such as local file names, directory paths, web hyperlinks etc.

Collection based research would call for internet access for which there must be sufficient access control to assure the safety of proprietary data. Establishing policies and mechanisms that provide the access control are equally important and have to coexist with other security mechanisms. Access and control policies should be clear and unambiguous to those who create and use scientific data. The data may be public, or it may be restricted to a group of users listed explicitly or by domain name.

One should encourage projects that demonstrate easy to use, yet strong, authentication schemes, including ways to issue usage-permission to valid users, ways to log usage, ways to provide different levels of authentication. A most important facet of this problem is authentication in distributed systems, so that a user only needs to log-in once, yet multiple, heterogeneous services can be enabled as a result.

9. Conclusion

Our centre being one of the leading organizations in the country, carrying out large scale Research & Development projects produce a very large amount of Scientific & Technical

data. The plant data generated at different stages of reactor operations carry lot of meaningful information known as well as unknown. Similarly content rich data generated out of basic and applied research, particularly in the nuclear fields and other related areas, are amounting to Peta bytes of data. Hence, the need of the hour is, design and development of Scientific and Technical Data Warehouse and appropriate Visualisation Techniques to allow more researchers, analysts and technologists across the country / globe to carry out Collection Based Research effectively leading to new concepts and ideas.

References:-

1. http://www.llnl.gov/CASC/sapphire/overview.html
2. http://www.siggraph.org/education/materials/HyperVis/percept/visrules.htm
3. http://seamonkey.ed.asu.edu/~alex/pub/multi-vis/multi-vis.html
4. http://www.gris.uni-tuebingen.de/gris/jb96/node11.html

Knowledge Management in E-environment

Vinitha K[1] Kanthimathi S[2] and Tharani Devi K[3]

Abstract

Knowledge Management helps us to share, learn and regenerate the new knowledge. Recent advancements in Information Technology (IT) and knowledge sectors have opened up important Knowledge Management opportunities. This paper presents the concept and definitions of Knowledge Management (KM), its importance, process, various modules of KM. The application of KM library & information centres and an overview of its emerging role, barriers, the challenges for the Library and information professionals, and the basic requisites for successful KM are highlighted.

Keywords: Knowledge management, Information Technology, Information professionals

1. Introduction

Librarians had till recently been the "back room" staff- invisible to clients, unknown to top management, irrelevant to business development / marketing groups and insignificant to projects/ business groups in typical corporate organization. Categorized as 'support staff' by personnel departments, the librarians were underused, undervalued and of course, underpaid.: The emerging business environments coupled with the explosive technological capabilities of the information era have brought the library profession to an increasingly visible & a more team oriented platform in progressive corporate organization

Knowledge Management (KM) is now an emerging trend of leading organization with transformation from manufacturing- oriented economy to service-oriented economy, the role of information and knowledge has become more prominent for all organizations. A knowledge resource is no longer just a factor of corporate status or image-It is now an issue for survival itself. What does this mean to library and information service professionals? Whether librarians not organizing and providing access to knowledge earlier?

Is it a shift from the data-oriented approach? Now it is KM, a discipline initiating the metamorphosis in the librarian's role in corporate organizations and knowledge-based companies call for an integration of librarian's role with business strategies. Hence, library and information professionals need to be proactive team player in the emerging area of KM.

KM needs to be taken seriously as an issue for Library and Information professionals because it is perceived to offer a substantial enhancement to their role. The leading organizations today are getting aware of the value of knowledge in various forms.

[1] Research Scholar & Librarian, National Engineering College, Kovilaptti, TN,
[2] Librarian, Rani Anna Government college for women , Tirunelvelli, TN,
[3] M.Phil. student

Information professionals can play an important role in KM that will definitely enhance the professional image and their role.

2. Concept and Definitions

The term 'Knowledge Management' has been borrowed from the corporate world, which has used it as a strategy to seek as well as capture the knowledge residing in people's heads and consciousness to help the business remain on the progressive path standing in a keen competitive environment. The main target of application of KM in profit-seeking organizations is to gain competitive advantage and to increase turnover to make a profit by enhancing and improving operations systems.

It is defined that KM is 'about putting information and people's competence to work together'. KM refers to 'transfer of knowledge into capabilities for effective action."

The systematic process of finding, selecting, organizing, distilling and presenting information in a way that improves an employee's comprehension in a specific area of interest. KM helps an organization to gain insight and understanding from its own experience that helps to focus the organization. Specific KM activities help focus the organization on acquiring, storing and utilizing knowledge from such things as problems solving dynamic learning, strategic planning and decision-making. It also protects intellectual assets from decay, adds to firm intelligence and provides increased flexibility.

3. Significance

The strategically perspective, today's practice in KM still has a few demerits. The reasons for this are manifold. Firstly the existing knowledge is captured and capitalized only to a low degree because knowledge is messy in character. Secondly, time factor- many employees are willing to document and use existing knowledge but pressure of work in enterprise should be allowed to participate in management but not in real practice. As a result, the uncultured knowledge can't be utilized for KM purposes.

Knowledge is the business asset of any progressive organizations that get ahead. The implementation of KM helps the information flow in the organization and in implementing organization-learning practices. KM is not just managing or organizing books or journals, searching the Internet for users or arranging for the circulation of materials. Each of these can be an isolated part of KM.

KM is enhancing the use of organizational knowledge through information management and organizational learning. The purpose is to deliver direct value to the business. The knowledge is embedded in the processes and experiences, skills, wisdom and capabilities of people. KM rests on two foundations, i.e., utilizing the organised information and application of people's competencies, skills, talents, thoughts, ideas, imaginations, etc.,

KM aims to draw out the tacit knowledge people have, what they carry around with them, what they observe and learn from experience, rather than what is usually explicitly stated. Managing knowledge goes much further than capturing data and manipulating it to obtain information. The aim of KM is for business to become more competitive through the capacities of their people.

4. Process

KM is a conscious strategy of getting the right knowledge to the right people at the right time and helping people to share and put information into action in ways that strive to improve organizational performance. KM is used in Library and Information Centre's (LIC) for communicating knowledge between levels of management and who are directly involved to work processes as well as in service sectors.

Though KM is viewed as a 'process is about acquisition, creation packaging and applications or reuse of knowledge', it is also said to 'consist of identification acquisition, developing, sharing and distributing, using and preservation of knowledge'. KM process includes all the terms and terminology described below:

Tacit Knowledge - includes the individual employee's expertise, memories, values and beliefs, viewpoints and values.

Explicit knowledge - is the process of communication from one place to another in a systematic way through documents and is more formal and codified.

Corporate memory - is the connection of know-how of an organization. This know-how relates to problems –solving, project experiences, and human resources management.

Intellectual assets/knowledge assets/capital*:* similar terms, which comprises knowledge assets regarding products, technologies, and market that a business owns.

Information economics - a study of the clear value of information to an organization.

Data mining - the exploration and analysis of automatic and semiautomatic means of large quantities of data in order to discover meaningful patterns and rules.

Knowledge representation - the process of describing and presenting usable way of knowledge known by a person.

Knowledge mapping - finding existing knowledge in an organization and creating a detailed picture of skills. Maps can be simple directories of names to search online databases of human expertise, research materials and pre-recorded information.

Concept mapping - the visual summary of ideas or topics and these ideas or topics are related to each other.

Knowledge engineer - a person responsible of acquiring knowledge and developing data and rules for expert systems.

4.1. Steps in Knowledge Management

1. *Knowledge capture*: A systematic procedure for organizing, structuring knowledge to make it accessible and usable to people.

2. *Knowledge organization*: An organization that values and uses its own knowledge in reflective ways that lead to profound shifts in directions, values, beliefs and operating assumptions.
3. *Knowledge preservation:* Once the knowledge is collected, and codified it has to be stored in a suitable form in the organization's knowledge base. The knowledge can be stored in forms such as individual employees, and by computer knowledge base. The advantage with the computer is its unlimited memory and instant access. Intranets and the knowledge bases are the tools, which store the organizational knowledge.

5. MODULES

- A KM system is based on eight vital modules, which are as follows:

- INFORMATION: The most important bezel acts as an instant access to update and customize information
- EXPERTISE :Connects in real-time experts in an organizations to members who earn assistance and even the tacit knowledge can made explicit

- COLLABORATION: Plays an important role to facilitate on-line brain storming sessions and preserves information
- TEAM: Ensures efficient and systematic management among share skills
- LEARNING: Abridges skill gap with the help of on-line sessions

- INTELLIGENCE: Deals mainly with the explicit knowledge among shares skills
- KNOWLEDGE TRANSFER: Relates to (a) Machine-readable data files (b)various on-line databases and CD-Rom information resides in the shape of optical disks, juke boxes or magnetic tape autoloaders and (d) Computerized networked systems
- KNOWLEDGE MAPPING: Identifies the body of knowledge within the organizations, which is primarily concerned with mute knowledge base and makes a repository of all skills and expertise in the organization.

6.Human Resources Management In Libraries

Human resources management takes it as its basic starting point to train high quality specialized talents and to revitalize the library undertaking. In practice, we should pay full attention to diversity and variation of library staff's requirements, strengthened management of different library staff by applying contingency management approach i.e. to some people, rigid management method is applied, rigorous supervision and control imposed, and quality requirements of work according to regulations and procedures are made clear, and to the rest of the people, more flexible management method is applied to let them participate in decision–making and consultation and undertake more jobs so as to bring their management abilities into full play and realize organizational and personal objectives.

Human resource management is the core of KM in libraries. In the knowledge economy era, the libraries will attach importance to vocational training and lifelong education of library staff to raise their scientific knowledge level and ability of acquiring and innovating knowledge. An all-round improvement of library staff's quality and positioning of human value will become important objectives of KM in libraries. The objectives of KM in libraries are to promote knowledge innovation, to promote relationships in and between

libraries, between library and user, to strengthen knowledge internetworking and quicken knowledge flow. In the knowledge economy era, libraries will carry out researches on development and application of information resources, construction of virtual libraries, protection of intellectual property rights in the electronic era, etc. thus founding the base for knowledge innovation

7. Application of KM in LICs

LICs are the best by fractionalized systems and services (e.g., integrated library systems, disparate commercial online databases and CD-ROM products) that "…. create … problems for knowledge workers… faced with integrating information delivery systems in order to do their work". K M systems comprising commercial news feeds, numbers of generic, open ended discussion databases and passive document libraries threaten to simply compound our problems. At great cost, companies and organizations build these systems only to become disenchanted when they fail to deliver the productivity gains tout in the literature. They fail because they do not provide information in a work-centered context that magnifies its utility several folds.

Successful KM systems in a LIC are those that"package knowledge" in the process of doing real work. Packaging knowledge involves the "… filtering, editing and organizing pieces of knowledge…in such a way that it's insightful, relevant and useful" such actionable knowledge. This needs to be supported by standard platforms and infrastructure. GroupWare is a clear choice. Within a GroupWare infrastructure, the value of case histories as a form of KM is markedly increased through the capability to add integrated access to directory and messaging services, external information resources and services, corporate applications and local file systems. Accordingly, a mature GroupWare infrastructure with significant end-user development capability has to be chosen.. Given the innovative nature of KM, it is essential that librarians be capable of actively participating in the application development process. In general, KM in libraries should include such aspects as follows:

- **Knowledge innovation management -** Refers to the management of the production, diffusion and transfer of knowledge as well as the network systems constructed by related institutions and organizations. It includes aspects, namely, theoretical innovation management and organizational innovation management.
- **Knowledge dissemination management** - Knowledge creators do not have much time and energy to look for knowledge users, Though there are a multitude of knowledge users, it is very difficult to acquire knowledge that already exists in the minds of knowledge creators as restricted by various objective and subjective conditions. Therefore, libraries may play the part of knowledge leader diverse media and channels to disseminate new knowledge. In the 21st century, the internet ,with its mass information and extensive contents, will provide people with the main approach to searching knowledge and acquiring information
- **Knowledge application management**: Libraries should also attach importance to provision of services for people to acquire knowledge and achieve maximum functions and efficiency of knowledge information. Therefore, knowledge services based on high-speed information networks should be carried out by.
- **Virtual libraries** - Setting up virtual libraries or information centres for enterprise, governments, public organizations and scientific research institutions. It is difficult for an enterprise or a social organization to put sufficient manpower, material and

financial resources on information gathering, organizing and developing. It is also impossible and unnecessary to spend a large amount of funds on information resources for their own use. Libraries can create virtual libraries or information centres for these organs separately according to their respective information requirements by using abundant information resources on the high-speed information networks.

- **Digitizing library resources** -The electronic libraries or digitized libraries are the technical modes and development trends of libraries in the knowledge economy era. The knowledge services of Libraries in the future will start with creation of databases comprising electronic journals and books in different languages that have discipline features and can operate on high-speed information networks. Great efforts should be made to transform all existing large non-electronic information resources into electronic information and integrate them into electronic libraries.
- **Digitized knowledge services** - This presupposes creating steps for users-oriented information. Service systems such as information dissemination, information search and special supply of information; quickening the creation of digitized libraries; studying the methods, means and techniques of information distribution and search with the internet as the base and web techniques as the core.

8. Information Professionals

Traditionally, library professionals are assigned to take care of direct roles such as catalogue cards, classified tools help to find right books needed. The tools for thesaurus construction and controlled vocabulary are already helping us manage knowledge. Document management such as content management and accessibility of images play role in KM activities. However with the emergence of KM in almost all services organizations, librarians and information professionals have to take on additional participative roles in expert systems, artificial intelligence and Knowledge Based Management System (KBMS) lessons learnt through these technologies are directly applicable to KM. By helping knowledge workers in the performance of cognitive tasks. Decision support systems are like the knowledge management systems. Information professional need to broad base their expertise in the library to include communication, facilitation, training and management skills. They also require a basic understanding of computing and network architecture. All knowledge –based organizations today survive and grow on the strength of their knowledge workers, rather than floor workers or administrative staff. Where the librarians are placed in theirs segments? Are they being treated as-till recently- just support staff in administrative workers category, or do they too qualify as knowledge workers? The answer depends on who is seen as a knowledge worker? Almost everyone is, because either you are using the knowledge or contributing in gathering knowledge in the organization's business. There is hardly any hierarchy among the knowledge workers. The knowledge work is about the acquisition, creation, packaging or application or reuse of knowledge.

KM system environment requires professionals with skills and strategies to handle the emerging situation. Information professionals should view this emerging phenomenon through focused attention on the process of KM; should keep an eye on how other interactive communities perceive KM; should stimulate thinking about the role of library and information service professionals in KM; and should visualize the role of library management knowledge

9. Barriers

KM is in reality a difficult process to implement particularly for an organization whose employees have been conditioned to the idea that knowledge equates power. In such an organizational culture little or not at all sharing of knowledge and information takes place between departments or even among colleagues within the same section. The staff in this organization would tend to hoard information, practice, and hence the "superiority" of their enlightened position, form being taken away from them. It is clear that the employees of different organizations require adjusting in values, culture and behaviours. These adjustments take time-often years and require commitment from the top, clear focus, and persistent application within library units. However, there are some barriers to the implementation of KM in organization. They are:

- Ignorance- not knowing who has the right information required for the job
- Lack of time to find out and absorb the best practices recommended
- Lack of a relationship between the sources and recipient of knowledge
- Time lag taken to implement best practices recommended across departments.

To overcome the above barriers in implementing KM in organizations certain measures are required: well defined organizational structure, co-operation with each other for sharing knowledge, elimination of bossism, mutual understanding among colleagues, eradication of superiority minded attitude, expertise to new information technology and ultimately involve the employees in participative management. Some other important factors in favours of the implementation factors of KM are:

- Set up the right people in the right job at the right time without any biases.
- Tackling personnel to solve the complex as and when they arise in the organization by cultivating their tacit knowledge.
- Authority and responsibility should be delegated as per level of position to professionals by which they can react immediately according to the situation with their professional experiences and efficiencies.
- Developing professionals by training, introduction of new technologies and know-how who try to keep up with them as a knowledge person.
- Develop sharing of knowledge system between units as well as among organizations as a whole through cooperation of the professionals.
- Support from top management, who recognizes the value of information or knowledge as a resource.
- Identification of the link between knowledge and the new measures of performance.
- A knowledge leader who can actively drive the knowledge agenda forward with commitment.
- Polices that drive the agenda forward.
- Creation of an environment or work culture that supports innovation, learning and knowledge-sharing

10. Conclusion

Economic environment and information environment are changing quickly today. KM has become a powerful tool for promoting innovation realizing and reengineering the various walks of life. It occupies an outstanding position in the creation of the Knowledge

innovation systems of a country. How far the library circles meet the challenge of knowledge economy and build the KM systems of libraries is a subject that demands our urgent study and solution. It is impossible to accomplish such important tasks by using man's brains only. In the modern society the knowledge changes with each passing day, it will be possible to link closely knowledge sources and knowledge workers by computer networks, thus constructing knowledge networks in libraries based on realization of single-point information

References:

1. Cao Yi. The Reorientation of libraries in the knowledge economy era. Library Work & Research, ol.3, 1999, pp.24-26.
2. J N Rao. Knowledge management by Library and Information centers. University News vol.40, no. 44; 2002, pp. 4-10.
3. K Kalseth.Knowledge management from A business strategy perspective FID Review, vol.1, no.1,1999, pp.37-51.
4. K Tarapanoff. The Knowledge Society & the Information Professional, Challenges & opportunities. FID Review, vol. 1, no.1, 1999, pp. 28-36
5. Sheng Xiaoping, Knowledge management of libraries in the 21st century. Library Magazine, vol.8, 1999, pp. 29-32.
6. Thomas H Davenport. The Future of Knowledge Management. URL: http://www.cio.com/archive 010196-dovencpor-content.html
7. Y. Malhotra. Knowledge management for the New World of Business. Asian Strategy leadership. Institute Review, vol.6; 1998, pp. 23-34.
8. Wang Delu. The collection and processing of Knowledge February 4,1999 http:// www/bsti.ac.cn/bsti_kmchina/gei/048_001.htm
9. www.csu.edu.au/special/online99 /proceeding99/brown-williams.htm
10. www.stevedenning.com/whatis-Knowledge management.html

Information and Knowledge Management Using GNOWSYS

Meena Kharatmal, Sandhya R., and Nagarjuna G.

Abstract

The objective of this paper is to introduce GNOWSYS (Gnowledge Networking and Organizing System) to the community and demonstrate its use for managing library, knowledge management, information gateways (portals) and distributed semantic grid for publishing knowledge.

GNOWSYS, a generic system for semantic computing, is a free-software developed at Homi Bhabha Centre for Science Education (Tata Institute of Fundamental Research), and is released as an official GNU project. It is specially made for publishing vocabularies, propositions, ontologies, complex systems, web services, semantic computing and such informative resources including library resources on the web. It is designed keeping in mind the recent developments in semantic computing and knowledge representation. It is a hybrid database system with distributed, hierarchical, object-oriented, relational database support. It can serve as an information grid, it can export data in regular semantic web standards such as OWL, XTM, and its native GNOWML. It can be used for creating digital encyclopedia, thesauri, dictionaries, glossaries, multi-lingual databases, for building e-learning applications, etc. A unique feature of GNOWSYS is its potential to publish knowledge at the ultimate granular level (terms and predicates of propositions).

Keywords: Knowledge Management, Information Management, Digital Library, Knowledge Organization, Semantic Web

1. Introduction

One of the ways of information communication is through the World Wide Web. The success of World Wide Web produced lots of information. Though World Wide Web is well-known today, often it is not utilized at its fullest. Usually, accessing the specific information is a trenous effort on the user as there are problems to manage the information. In order to overcome this problem, and to harvest the World Wide Web, new technologies need to be developed and GNOWSYS is one such tool. The paper introduces GNOWSYS (Gnowledge Networking and Organizing System) to the community and demonstrates its use for managing library, information and knowledge management. We shall present a case study on using GNOWSYS for creating the web version of the library OPAC and knowledge management

2. Semantic Web

The World Wide Web is being used today typically for seeking information, searching for and getting in touch with other people, establishing peer groups, reviewing of research, expanding the business, reaching out globally, working on-line etc. In the Web technology, the main valuable and indispensable tools are search engines without which the Web would not have been a huge success. However, there exists serious problems associated with their use. The problems being high recall, low precision; sensitivity of

Homi Bhabha Centre for Science Education, Tata Institute of Fundamental Research, Mumbai

results to vocabulary which occurs due to usage of different terminology, retrieval in the form of single Web page and to extract the information and collate it together calls for manual work[9].

The obstacles can be overcome by making the Web content into a machine-accessible information. The solution seems to lie on developing increasingly sophisticated techniques based on knowledge representation (KR) and computational linguistics, in order to represent Web content in a form that is easily machine-processable and to make use of intelligent techniques. This plan of revolutionizing the Web is referred to as the Semantic Web, evolving gradually out of the existing Web.

In order to create Semantic Web, advanced knowledge management systems are required. The most important aim of the Semantic Web is to organize knowledge in conceptual spaces according to its meaning. The inconsistencies will be checked and new knowledge will be extracted which can be managed by the automated tools. A keyword-based search will be replaced by query answering which can be handled over several documents i.e. a queried or a requested knowledge will be retrieved, extracted and presented in a human-friendly way[9]. The vision of the Semantic Web was articulated by Tim Berners-Lee in the following way:

> The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. — Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001.

The Semantic Web provides a common framework that allows data to be shared and reused across applications. It is an approach to integrate data across the Web and intends to create a universal medium for information exchange by using the semantics (meaning) in a machine-understandable language to the Web contents. GNOWSYS is one such tool developed applying the semantic principles.

3. GNOWSYS: Gnowledge Networking and Organizing SYStem

GNOWSYS is an object oriented semantic engine developed using ZOPE and Python. In GNOWSYS each object is provided by an unique URL. Since it is built using a web-application server, the data can remain anywhere on the Internet, keeping only the metadata in the systems. It is possible to manage, publish and distribute knowledge by applying the semantic web principles. GNOWSYS can be used for developing ontologies, databases, electronic encyclopedias, portal sites, concept bases (knowledge bases), expert systems, etc., and in general can be used for any knowledge and information management and knowledge representation. It is envisaged that GNOWSYS will help to build a semantic grid for making multi-lingual distributed concept base encompassing all knowledge of human beings, including that of science. A portal, www.gnowledge.org[1], is being developed demonstrating the said potentials of GNOWSYS.

3.1 A Hybrid Knowledge Base

GNOWSYS, is a hybrid database system with hierarchical, relational, distributed, object-

oriented, database support. In a hierarchical database, records are grouped in a logical hierarchy, connected in a branching structure. A relational database is a database constituting of set of relations in the form of tables in which the records are in the form of attribute values. In the object-oriented database, the contents are stored in the form of objects, with features of inheritance (the values which are assigned for parent objects are inherited to the child objects assigned), polymorphism (it is the ability of objects belonging to different types to respond to methods of the same name). A distributed database is a database that is under the control of a central database management system in which storage devices are not attached to a common CPU. It may be stored in multiple computers located in the same physical location, or may be dispersed over a network of interconnected computers[6].

3.2 Architecture

The architecture of GNOWSYS, is designed keeping in mind the recent techniques involved in knowledge representation (KR) area. KR, in artificial intelligence (AI), is used to study formalisms to model human knowledge and for problem solving. GNOWSYS can be applied for drawing concept graphs, semantic nets, concept maps, SemNets which are currently used by several researchers and educationists[20,19,11] to enhance conceptual learning in the context of science education. The architecture of GNOWSYS is structured to accommodate different dimensions of KR such as—*generality, semantics, complexity, inference*[12] as shown in Figure 1.

We come across a wide variety of concepts in the form of particulars or generals or even abstract. In order to understand the wide variety of concepts in our discourse it is essential to organize the concepts based on their order of generality. Along the *generality dimension* of GNOWSYS, the three different levels of generality such as tokens (particulars), types (generals), and metatypes (types of types), can be organized. For example, it is possible to organize knowledge of the taxonomical classification system along the generality dimension. In the taxonomical knowledge base, there exists concepts such as—mammals, dogs, cats, Fido, Tom, humans, James Watson, etc. The generality dimension is applied when organizing the above knowledge base—Fido, Tom, James Watson are the particulars and hence organized in the token layer; dogs, cats, humans are generals and hence are organized in the type layer; and taxonomical concept, person concept are organised in the metatype layer.

The semantics and consistency check is carried out along the *semantic dimension* of GNOWSYS. To start with, in the first layer all kinds of propositions are allowed to store without any semantic constraints. This is observed in the case of novices when they start with representing their knowledge in idiosyncratic way without following any constraints or principles. The propositions are stored in the form of *well formed formulae* (WFF). In the second layer, the WFFs can be combined with the semantic constraints, logical connectives, modalities, propositional attitudes, quantifiers etc. The consistency is implicit at this layer and hence it is referred as *implicit structured system* (ISS). In the third layer, the validity constraints can be imposed explicitly and therefore gives rise to *explicit consistent system* (ECS) which is quite similar to the experts' knowledge system. The semantic dimension allows to represent novice's and expert's knowledge (alongwith the transformation), with the semantic constraints and validity. The semantic dimension is implicit throughout the entire knowledge base built using GNOWSYS, and hence it is not

represented in the figure.

The basic components of a knowledge base are ObjectType (OT), Object (O), RelationType (RT), Relation (R), MetaType (MT), EventType (ET), Event (E), FlowType (FT), Flow (F). These components help to store the terms, propositions and procedures. The *complexity dimension* helps to generate complex compositions using the structure groups consisting of ProcessType (PT), Process (P), StructureType (ST), Structure (S), Encapsulated Class, Programs and ProgramType.

Using the *inference dimension* of GNOWSYS, the epistemic values such as validity and truth can be checked. At present, there does not exist any in-built module supporting for inferences but with developing an interface between the GNOWSYS and any existing inference engines, it is possible to deduce consequences using deductive inference.

3.3 Underlying Technology

GNOWSYS is developed as a product of ZOPE, by extending ZOPE classes and using Python programming language. ZOPE (Zee Object Publishing Environment)[7] is a free (as in freedom) web application server for building content management systems, intranets, portals, and custom applications. ZOPE allows to store content and custom data, dynamic HTML templates, scripts, catalogue, and connections to relational databases. It features a strong through-the-web development model, allowing to update the website from anywhere in the world alongwith powerful integrated security model. It is possible for connecting not only to ZOPE's object database, but also with relational database allowing for strong data integrity. It runs on the GNU/LINUX, UNIX-based platforms as well as Windows NT. It can be used with most popular web servers or its own built in web server. ZOPE is designed for Web object as well as Web development model. It is written using Python, an interpreted, interactive, object-oriented programming language with clear syntax and dynamic semantics. Python[4] emphasizes readability due to its clear syntax and therefore is easy to maintain. GNOWSYS is interoperable, i.e. it works on all the known operating systems.

3.4 Semantic Computing

As presented above, GNOWSYS has procedural objects, alongwith declarative objects, such as function, program, and class. GNOWSYS is unique in this feature that it can store and activate (executable) objects. Using this feature, it is possible to design applications without writing programs in any programming language i.e., specifying the semantics of a program and mapping the elements of the program to the surrogates of procedures is sufficient for GNOWSYS to test the application design. This forms the foundation for semantic computing using GNOWSYS. This feature enables anyone, even those who do not have computer science or programming background, to develop a knowledge base provided they can grasp the logic of the domain.

4. Dynamic Information Management Using GNOWSYS

In the present scenario, with the help of available library software packages, it is possible to organize the library resources. But this is not enough, because the resources are not

always books and journals, there is also more information available on the internet. If we want to fully exploit or utilize the information available globally, we need to go beyond just organizing the library collection, to information and knowledge management. The existing softwares do not meet the above requirements, and hence GNOWSYS plays a major role in managing all kinds of digital resources alongwith traditional resources. Using GNOWSYS it is possible to integrate the functionalities of information management such as managing the documents and also for planning and coordination.

Information management involves anything from library management, to office management, or any types of record management. Due to budget constraints, most organizations find difficult to implement information technology (IT) for their services including libraries. It is important for managing databases, office resources, providing online access over the internet. To cater to all these requirements, usually there are special applications for each purpose. A generic information management system like GNOWSYS can serve all the requirements in the office such as accounting, creating inventories, managing leave records, etc., and also for replacing the specialized software.

The reader would think that all the above requirements can be achieved using any database, and would also like to know what is so special about GNOWSYS. We would like to emphasize that GNOWSYS is especially developed for managing dynamic knowledge base systems. It is made to model the cognitive development (the development of human knowledge). For example, in a human's ontogeny (lifetime), knowledge is acquired by changes in conceptual schemes, and it gets developed through various processes. It is very essential to capture these changing conceptual schemes as they get developed. KR is one such alternative for modeling human knowledge, and GNOWSYS which is based on KR principles, allows for representing various conceptual schemes in one's ontogeny or learning course. GNOWSYS allows one to begin representing the knowledge starting from loosely formed structures which can at the same time have contradictory beliefs, semantic inconsistencies in an implicit form and over a period of time transforming the knowledge into an explicit and consistent form. This is what we assume in the context of learning. In this way it can capture the development of human knowledge because it always preserves the earlier knowledge structure, a feature which is not found in the conventional database management system (DBMS) due to their rigid structures or schemes.

In the conventional DBMS, in order to create database the scheme should be frozen before adding the data. However this does not allow to incorporate changes in the existing schemes, and one ends up in creating different schemes as and when required. In this way the schemes in DBMS are rigid. On the contrary, GNOWSYS is flexible because its schemes can change according to the input and the requirements. The schemes in GNOWSYS are generated as they get developed while we input the data. GNOWSYS helps to represent developing systems or changes.

5. Library Management Using GNOWSYS

The reader may have realized that GNOWSYS can be used to create wide variety of applications related to information management. In this section our focus is to present a case study on using GNOWSYS for efficient library management system. Based on the three layer architecture of the GNOWSYS, we organized the metadata of the library.

Metadata is data that describes data or content objects. The metadata was categorized in the form of ObjectType, Object, AttributeType, Attribute, RelationType, Relation. Objects are classified according to subclass/superclass relation, and also instantiation relation (class-instance). Relations are assigned between objects and their classes (called ObjectTypes or Metatypes). The library's metadata includes ISBN, author, editor, publisher, author mark, pages, subject, class number, accession number, etc. The ObjectType is used for classification. The subject is classified as ObjectType since it is a class. The unique feature of a book or a journal for example, accession number or periodical number are assigned as ID, and those such as ISBN, ISSN, are classified as AttributeType which can be assigned values. Every physically existing material i.e. each book or each journal is classified as an Object. The organization of the library metadata is shown in figure 2.

When creating entries for books, one has to select its parent ObjectType which is nothing but the subject to which it belongs. Once the object is created, then values are to be assigned for each attribute of the book.

5.1 Library on the World Wide Web

The Centre's library database is currently managing the database of around seventeen thousand resources on the existing proprietary software. Using GNOWSYS, this library database was launched on the World Wide Web. Since GNOWSYS is a web-application server, incidentally also developed in-house, the library database can be uploaded on the internet. Inorder to create a web version of the library collection, we worked on the data exchange model. First the metadata of library information science (LIS) was translated into GNOWSYS model. The database (.dbf) file of the existing software was first opened in a spreadsheet and saved as tab delimited text file. Using Python, a script was written to read each and every field of the database tables which was created using previous database management system. After the successful generation of the script, each and every field of the database tables was fully converted into objects in GNOWSYS. Once the database was successfully imported the display on GNOWSYS was designed in the user-friendly way alongwith the two standard formats viz., MARC and CARD. One of the most widely used feature in the library is query for the collection. The search script enables one to generate results for the query based on author, editor, title, publisher, ISBN and keywords. Since both data and metadata of objects can be stored, indexed in a catalogue it enables for a faster query.

5.2 Data exchange using GNOWSYS

GNOWSYS supports data exchange in several forms. It can accept any import filter as long as the database exports into text or XML (eXtensible Mark-up Language). XML, a sub-set of earlier mark-up language called the Standard Generalized Markup Language (SGML), is an open standard used for defining, validating as storing structured data objects by expressing them as tagged text in Unicode. The data-exchange modules of GNOWSYS will support various standard knowledge representation schemes such as CL (common logic), PetriNets, KIF (knowledge interchange format), CG (concept graphs), OWL (web ontology language), XTM (XML Topic Maps), etc.

5.3 Managing User Profile

A staff user profile is created by importing the user's database. This facilitates the circulation as well providing selective dissemination of information (SDI), current awareness services (CAS). This allows each user (with secured login ID and password) to recommend or reserve the books on-line, or request for any other information. It is also possible to view the table of contents, generate bibliographical entries for citations.

5.4 Selective Dissemination of Information

Since an user profile already exists, it is much easy to create relations between the areas of interest of the users with that of the collection (i.e. keywords). Using GNOWSYS, it is possible to establish linkages or relations between the users and new arrivals in the library. Whenever there is a new arrival in the library, users are alerted through email. It can generate online request forms regarding abstract, reviews, etc. Current awareness in the form of new arrivals, contents of current journals, clippings of newspaper articles can be disemminated to the users using the user profile.

5.5 Digital libraries

In the IT age, there is a growing demand for the latest information. The digitization of the library involves "series of activities that brings together collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge"[18]. The main advantages of a digital library are that the digitized information can be accessed from anywhere in the world, it has fast access, the storage, management and retrieval of information is precise and easy, and the digitized information can be easily converted into a hard copy. The data of the digitized information is transmitted in its exact form (images, video, audio, text, etc.). GNOWSYS can be used for digital document management system (DMS), for any large organization or for personal use. It allows for creation, storage, retrieval and dissemination of documents in electronic format.

World Wide Web made possible several digital libraries around the world, for example, Wikipedia, Wordnet, OpenCyc, Netscape's Open Directory which has a storage of digital data. Using the metadata of GNOWSYS, it is possible to create relations with these digital libraries and produce a structured and organized knowledge base.

6. Content Management Using GNOWSYS

Content Management (CM) is a set of processes and technologies which can create, update, publish, translate, archive, and retire. To start with, an instance of digital content is created by authors, and later the content is edited (by editors), and on approval of the content it is ready for publication. The next generation of content management is best referred to as semantic content management, where the content carries meaning that is expressed by metadata according to a semantic structure based on open standards[22]. The semantic content management is about managing content objects based on their properties.

7. Knowledge Management Using GNOWSYS

GNOWSYS is being used as a tool for knowledge management in science education. At present, there exists wide variety of tools for knowledge representation (KR), of which the most widely used are Concept Maps[20,19], SemNet[11], semantic network, concept circle diagrams[19], concept graphs[23]. The graphs generated from these KR tools are usually stored separately thus making the knowledge base inaccessible for the users across the network. However, in the design of GNOWSYS, the graphs generated by various applications can be shared and published by the system through XML based representations schemes.

GNOWSYS is being used in research on "Knowledge Organization in Biology Education". A knowledge base of concepts and semantic relations is being developed using GNOWSYS[15]. Our research study is to characterize and organize knowledge based on KR using the grammar of scientific knowledge. The research methodology of organizing knowledge is based on KR approach which involves to apply logical principles. Using GNOWSYS, we have created a knowledge base of some biological terms wherein the knowledge is classified and organized into MetaType, ObjectType, RelationType, Relation, Objects, Attributes. Some of the objectives of our research are—to organize concepts based on their cognitive function (role); assign valid and authentic semantic relations (knowledge organizers) to the concepts; to compare and restructure (reorganize) the novice's knowledge structure with that of an expert's knowledge structure. Using GNOWSYS, graphical representations such as concept maps, concept graphs can be generated based on the knowledge base. The knowledge organizers help eliminate ambiguity, maintain parsimony and apply precision to the scientific body of knowledge[16].

The figure 4 shows a principled concept map of organic molecules generated from our knowledge base. It depicts the knowledge organizers i.e. relation types which help to represent some of the knowledge about organic molecules.

8. Conclusion

GNOWSYS, a web application server, is developed in-house and is released as an official GNU project. It is a free (as in freedom) software i.e. it gives freedom to run the program, study the program and adapt or improve/change the program, and also to redistribute the program to the community.

The objective of this paper is limited to introducing GNOWSYS to information scientists. We have shown that GNOWSYS can be used in the area of information and knowledge management, for content management, creating digital documents, etc.

Acknowledgements
We wish to thank Bipin Apandkar and Prasanta Baruah for providing the programming support for library management and Shrikant Gaikwad for providing the technical support.

References:

1.GNOWLEDGE portal site URL: <http://www.gnowledge.org/>

2.HBCSE library website using GNOWSYS URL: <http://www.hbcse.tifr.res.in: 9673/library/hbcse-library/>

3. A multilingual thesaurus URL: <http://www.gnowledge.org/wordnetbth/>

4.Python URL: <http://www.python.org>

5.Website of Homi Bhabha Centre for Science Education URL: <http://www.hbcse.tifr.res.in/>

6.Wikipedia, the free encyclopedia URL: < http://en.wikipedia.org/wiki/Database_systems>

7.ZOPE (Zee Object Publishing Environment) URL: <http://www.zope.org>

8.GNOWSYS: Gnowledge Networking and Organizing SYStem, 2005 URL: <http://www.gnu.org/software/gnowsys/>

9.Grigoris Antoniou and Frank van Harmelen. The semantic web vision. In: A semantic web primer. Cambridge: The MIT press, 2004, pp. 1-21. ISBN 0-262-01210-3

10.D. Kamalavijayan, Harsha Parekh, K.S. Raghavan, S.B. Ghosh, N.R. Satyanarayana, S.K. Savanur, Muttayya Koganuramath, T.J. Mathew (eds.) International conference on information management in a knowledge society. February 2005. New Delhi: Allied Publishers Pvt. Ltd., 2005, vol.I & II. ISBN 81-7764-770-9 & 81-7764-771-7

11.Kathleen Fisher and James Wandersee. Mapping biology knowledge. The Netherlands: Kluwer Academic Publishers, 2000. ISBN 0-7923-6575-5

12.Nagarjuna G. GNOWSYS: a system for semantic computing, 2005 URL: <http://www.gnu.org/software/gnowsys/gnowsys_files/cpv2/

13.Ramesh C. Gaur. Re-engineering library and informational services: process, people & technology. Mumbai: Allied Publishers Pvt. Ltd., 2003. ISBN 81-7764-487-4

14.D. Kamalavijayan. Information and knowledge management. Delhi: Macmillan India Ltd., 2005. ISBN 1-403-92652-2

15.Meena Kharatmal and Nagarjuna G. Knowledge organization of biology education, 2004 URL: <http://zope.hbcse.tifr.res.in: 9673/Meena/Biology/>

16.Meena Kharatmal and Nagarjuna G. Understanding science through knowledge organizers: an introduction. Episteme-1: An International Conference to review Research on Science, Technology and Mathematics Education. An international conference held at Goa, 13-17 December 2004. Mumbai: HBCSE, 2005. <http://www.hbcse.tifr.res.in/episteme/themes/meena-knowledge-organizers>

17.Doug Lenat. The CYC knowledge server, 2005 URL: <http://www.cyc.com/cyc/technology/>

18.Gail McMillan. Digital libraries support distributed education, 1999 URL: <http://www.ala.org/ala/acrl/acrlevents/mcmillan99.pdf>

19. Joel Mintzes, James Wandersee and Joseph Novak (eds.) Teaching science for understanding—a human constructivist view. USA: Academic Press, 1998. ISBN 0-12-498360-X

20. Joseph Novak and D. Bob Gowin. Learning how to learn. UK: Cambridge University Press, 1994. ISBN 0-521-31926-9

21. National library of medicine. Unified medical language system (UMLS), 2005 URL: <http://www.nlm.nih.gov/research/umls/>

22. Profium White Paper. Technical impact of semantic on content management. October, 2004. <http://www.profium.com/library/Profium_technical_white_paper.pdf>

23. John Sowa. Knowledge representation: logical, philosophical and computational foundations. USA: Brookes/Cole, 2003. ISBN 0-534-94965-7

24. Morton Winston, Roger Chaffin and Douglas Herrman. A taxonomy of part-whole relations. Cognitive science, vol.11, 1987, pp.417-444.

Knowledge Management in a Research Organization

Joseph JVM., Soundararajan E.,Jayakumar C. and Somasekharan M.

Abstract

Realization of knowledge as the core competence, coupled with recent advances in information technology, has increased organizational interest in the subject area - Knowledge Management. Knowledge Management System (KMS) is directed towards enhancement of the productivity and performance of an organization. Such a system would empower employees to access a repository of "best-of-breed" documents, information resources and external knowledge sources. KMS could provide 'Expert Knowledge' on areas that could best benefit from the skill sets and provide them with the collaborative tools needed for effective project management. This paper explains the need and impact of KMS in a research organization and how it would help for researcher's work/output and institutional values.

Keywords: Knowledge, Externalization, Knowledge Repository, Knowledge Management, Knowledge Sharing .

1.Introduction

Information is a key resource in the present day knowledge driven economy. Knowledge management is necessary for organizations in particular for research organizations because what worked yesterday may or may not work tomorrow. There is a lot of knowledge, which is meant to be diffused but mostly either unrecognized or withheld form wider distribution even within an organization. Thus there is a need for KMS in every organization, to facilitate a free flow to identify and access intellectual skills, resources and the expertise available within one's own orgasnisation.

"Knowledge Management caters to the critical issues of organizational adoption, survival and competence in face of increasingly discontinuous environmental change. Essentially, it embodies organizational processes that seek synergistic combination of data and information processing capacity of information technologies, and the creative and innovative capacity of human beings"(Malhotra, 1997).

The key factors that have an impact on research organizations are high level of uncertainty leading to inability to predict the future. Defining goals and objectives, best practices and use of information technology may not necessarily help the organizations to achieve long-term organizational competence and survival. The organizations need the capability to understand the problems in this world of changing environmental conditions. KMS focuses on doing the right thing more than setting the things right. Hence, it can said that KMS ensures creation, dissemination, renewal, and application of knowledge towards organizational sustenance and survival.

Scientific Information Resource Division, Indira Gandhi Centre for Atomic Research, Kalpakkam .
E-mail joe@igcar.ernet.in

2. Merits

The successful implementation of KMS helps the organizations to deliver creative products and services. In today's information-driven economy, organizations uncover most opportunities and ultimately derive most value from intellectual rather than physical assets. To get maximum value from an organization's intellectual assets, knowledge must be shared and utilised as the foundation for intra- and inter-organisational integration. Consequently, an effective KM program should help an organization do one or more of the following:

- Foster innovation by encouraging the free flow of ideas
- Improve customer service by streamlining response time
- Enhance employee retention rates by recognizing the value of employees' knowledge and rewarding them for it
- Streamline operations and reduce costs by eliminating redundant or unnecessary processes

A creative approach to KMS can result in improved efficiency, higher productivity and increased revenues in practically any business function. KMS connects the people with the knowledge for helping their decision making process. The best efforts of KMS are as transparent to employees' workflow as possible. In the R&D sector, managing knowledge is considered as a key to achieve organizational goals.

3. Structure and Process

Organizations that are knowledge-oriented for the realization of its strategic goals are called knowledge intensive organizations. To move an organization from a knowledge-intensive structure to a knowledge-based structure, it is necessary to view the organization structure from a knowledge perspective. The level of knowledge orientation of an organization is based on seven characteristics: strategy, organizational structure, technology, performance measurement, HRM (Human Resource Management), culture and level of explicitness of knowledge

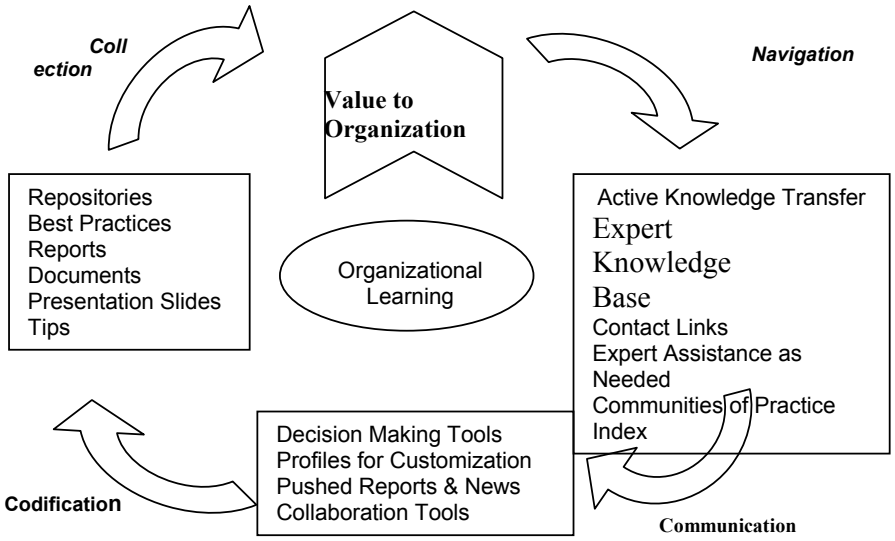
On identification of the knowledge intensive situation of the organization, the next step is to evaluate the specific knowledge problems in the organization, processes or activities. The most usual knowledge problems are unbalanced distribution, fragmentation, unavailability and inaccessibility of knowledge. In bringing knowledge management into an organization, it is required to select and implement a number of processes that will help the organization to be better at creating, finding, acquiring, organizing, sharing and using the knowledge it needs to meet organizational goals. There are many such processes some are highlighted below:

- Conducting knowledge audits to identify knowledge needs, knowledge resources and knowledge flows
- Creating knowledge strategies to guide the overall approach
- Connecting people with people to share tacit knowledge using approaches such as communities of practice or learning events

- Connecting people with information to share explicit knowledge using approaches such as best practices databases, and using content management processes to ensure that explicit knowledge is current, relevant and easily accessible
- Creating opportunities for people to generate new knowledge, for example through collaborative working and learning
- Introducing processes to help people seek and use the knowledge of others such as peer assists
- Teaching people to share knowledge in ways that inspire people by using storytelling techniques
- Encouraging people to prioritize learning as part of their day-to-day work, by learning before, during and after the tasks and projects they have performed

These stages of the KMS functional process can be clearly identified and understood in Figure 1.

Figure 1- Knowledge Management Cycle



4.Challenges

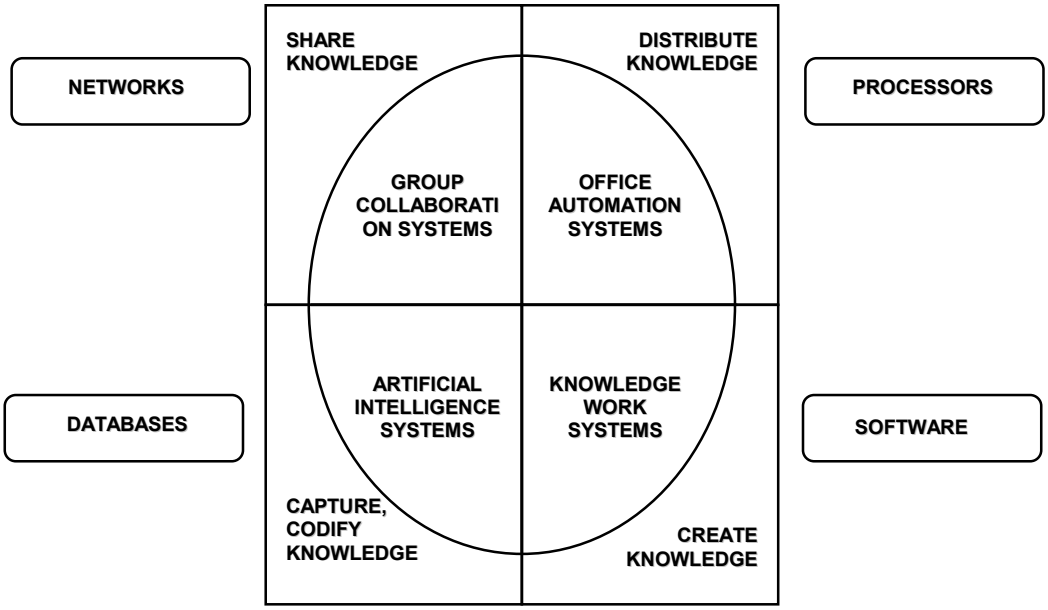
In a research environment, where an individual's knowledge is valued and rewarded, establishing a culture that recognizes tacit knowledge and encourages employees to share it is critical. One way organizations motivate employees to participate in KMS is by creating an incentive program. However, then there's the danger that employees will participate solely to earn incentives, without regard to the quality or relevance of the information they contribute. KMS is not a technology-based concept. While technology can support KMS, it's not the starting point of a KMS program. The KMS is mainly based on people, knowledge, and business objectives. Technology is the last step in the KMS process that says how it will be carried out. As with many physical assets, the value of knowledge can erode over time. Since knowledge can get stale fast, the content in a KM program should be constantly updated, amended and deleted. The relevance of knowledge at any given time changes, as do the skills of employees. Therefore, there is no endpoint to

a KMS program. Like product development, marketing and R&D, KMS is a constantly evolving business practice.

KMS program should take only the required and relevant information from the available information resources. Since KM is not a technology-based concept but a business practice, enterprise wide KM efforts should have dedicated KM staff headed by a chief knowledge officer or other high-profile executive.

5. Information Technology and KMS

Information and Communications Technology (ICT) is an important ingredient of virtually every successful KMS program. They are the enablers and supporters of knowledge management. ICT has intensified the interaction among the processes previously isolated from each other across space and/or time. The information networks are the main infrastructures of the new knowledge based organization. It offers the opportunity for a virtual information & knowledge support system, which will connect data, information, knowledge and people through virtual communities, knowledge repositories and knowledge portals.



ICT has offered a variety and a wider range of KM tools. These include: a new generation of artificial intelligence solutions, new flavors of document management systems and various collaborative technologies such as the internet, intranet, GroupWare, computer mediated collaboration, data warehouses, knowledge discovery in databases (data mining), computer-based yellow pages, simulation tools, intelligent agents and so forth.

6. Conclusion

With increasing uncertainty, complexity, risks and challenges, abundant data, lack of organizational coordination, there have been more pressures for the organizations to survive. KMS has aroused a new interest and looked as an effective solution for a growing number of research organizations, where the perception of knowledge and

learning are key success factors. Individual, team and organizational learning, coupled with a well planned, designed and developed KMS offer the best capability an organization can have to change, adapt, and influence its environment in a way that optimizes its performance over time. It is important that KMS also consider the important issue of privacy to personal information and knowledge.

References

1. C.W. Holsapple. Handbook On Knowledge Management. New Delhi: Springer, 2003. ISBN 81-8128-038-5.
2. Malhotra, Yogesh.(1998). Knowledge Management, Knowledge Organizations & Knowledge Workers: A view from the Front Line. URL:<http://www.brint.com/interview/maeil.htm>
3. Malhotra, Yohesh. Integrating Knowledge Management technologies in organizational business processes: getting real time enterprises to deliver real business performance. Journal of Knowledge Management, Vol.9 no.1 2005 pp.7-28.
4. Kavindra Mathi, *Key Success Factors of Knowledge Management. URL:*< www.dmreview.com/whitepaper/ KavindraMathi.KnowledgeManagement.pdf>
5. Javier Ruiz Fernandez and Eneko Alvarez *Management of Internal and External Knowledge in Research and Technology Organisations* 3rd European Knowledge Management Summer School 7-12 Sept, 2003 San Sebastian, Spain.
6. Managing Knowledge : Knowledge Work and Artificial Intelligence. URL:< http://www.cofc.edu/~marlind/MISchap12.ppt>

Advances in Knowledge Management

Kuriakose K.K., Satya Murty S.A.V., Athinarayanan S. and Swaminathan P.

Abstract

Knowledge Management has become a core management discipline that enables knowledge to grow , flow and create value in an organization. In other words it is the process by which the organization generates wealth form its knowledge-based assets. It involves people, information, workflows, 'best practices', alliances, 'Communities of Practice' etc. "Knowledge" in this context includes both the experience, emotions, values, hunches and understanding of the people in the organization, and the information artifacts, such as documents and reports, available within the organization and in the world outside. Value is created in the organization by "productivity" and "innovation", through the application of knowledge to work. Knowledge is basically classified into explicit knowledge and tacit knowledge. Explicit knowledge is available in paper, computers etc., where as tacit knowledge resides in the heads of people.

Knowledge Management is concerned with creating organizational environments for people to share, create and leverage knowledge for innovation and competitive advantage. Knowledge Management has created a paradigm shift in the thinking of people from "knowledge is power" to "knowledge *sharing* is power". 'First Generation Knowledge Management Systems' deal more with explicit knowledge. It follows the hierarchy of 'Data', 'Information', 'Knowledge' and 'Wisdom'. Here the organization is treated as an 'Information Processing Machine'. The 'Second Generation Knowledge Management Systems' deal more with tacit knowledge. The central tenet here is that ' knowledge resides in the user and not in the collection; what matters is how the user reacts to a collection of information'. Here organization is treated in a more humanistic & organic way.

In this paper the concept, benefits, and implementation of Knowledge Management Systems is described. The concept of 'Second Generation Knowledge Management Systems' necessitated by new organizational environments is introduced. The concept of 'Communities of Practice' and 'knowledge intrapreneur' are explained. It is emphasised that successful Knowledge Management implementation requires the synergistic combination of innovation and creativity of humans and advanced capabilities of new information technologies and more importantly the initiative and involvement of top management. It is concluded that 'Second Generation Knowledge Management Systems' can achieve simultaneous 'freezing' and 'unfreezing' of 'organizational best practices' to ensure that the *effectiveness* of the decision making is not sacrificed at the *altar* of increased *efficiency.*

1. Introduction

Ever growing uncertainty, unpredictability, increasing complexity and globalisation seem to have unleashed a whole series of rapid changes of unforeseen magnitude, affecting every sphere of life. Survival and success in the 21st century depend entirely on our ability

Electronics & Instrumentation Group, IGCAR, Kalpakkam – 603 102
Email- kuriakose@igcar.ernet.in

to respond to these changes with new coping mechanisms, with dynamism and effectiveness. This is true for individuals, for communities and for organizations. While the enormous changes in information technology has thrown up immense possibilities and opportunities, there are enormous gaps in the access, control and utilisation of technology, information and knowledge. And these gaps divide countries, regions, different sections of society, and even people within an organization.

While it is easier to minimize the technological gap, bigger challenge lies in bringing about changes in attitudes and values, changes in the ways of working and communicating, and changes in the entire framework of relationships and processes -- among people, within organizations, and in the interface with the socio-economic environment, which is necessary to cope with the rapid changes. More and more organizations are turning to Knowledge Management, to address these challenges.

Today, a range of technologies like computers, internet, groupware, information warehouse, video-conferencing etc. offer unprecedented opportunities to disseminate information, know-how and insights rapidly and cheaply to a worldwide audience. The reach of information, know-how and experience possessed by individuals can be greatly extended once it is captured and explicated so that others can easily find it and understand and use it. Knowledge may be explicitly available in the form of reports of activities, minutes of meetings, memoranda, proceedings of conferences, other documents or databases, audio and video recordings, multimedia presentations or implicitly as the experience, emotions, values, hunches and understanding of the people in the organization. Many factors have transformed the way in which organizations now view knowledge, but perhaps the pivotal development has been the dramatically extended reach of know-how through new information technology. Rapidly falling costs of communications and computing and the extraordinary growth and accessibility of the World Wide Web present new opportunities for knowledge-based organizations, to share knowledge more widely and cheaply than ever before. Nevertheless, even with modern tools, the process of knowledge transfer is inherently difficult, since those who have knowledge may not be conscious of what they know or how significant it is. Thus the know-how is sticky and tends to stay in people's heads. The diverse efforts of organizations around the world to share knowledge are being pursued under various labels, including Knowledge Management , Knowledge Sharing, Knowledge Exchange, Knowledge Focus, Knowledge Creation, Organizational Learning, Intellectual Capital Management, Intellectual Asset Management etc. There is no agreed definition of Knowledge Management , even among practitioners. The term is used loosely to refer to a broad collection of organizational practices and approaches related to generating, capturing, disseminating know-how and other content relevant to the organization's activities, and the enabling organizational culture.

Knowledge is basically classified in to 'tacit' and 'explicit'. Tacit knowledge resides in the heads of people and explicit knowledge is available in paper, computer etc. Tacit knowledge is what the knower knows, which is derived from experience and embodies beliefs and values. Tacit knowledge is actionable knowledge, and therefore the most valuable. Furthermore, tacit knowledge is the most important basis for the generation of new knowledge. The key to knowledge creation lies in the mobilization and conversion of tacit knowledge. Explicit knowledge is represented by some artifact, such as a document or a video, which has typically been created with the goal of communicating with another person. Explicit knowledge is information which can be used for problem solving. It is

applied information. Both forms of knowledge are important for organizational effectiveness. It is found that the strongest contribution to current Knowledge Management solutions is made by technologies that deal largely with explicit knowledge. Contribution to the formation and communication of tacit knowledge and support for making it explicit are currently weak, though some encouraging developments like communities of practice(CoP) are discussed. Hence the initial focus will be on KM solutions based on explicit knowledge and subsequently on tacit knowledge

2. Knowledge Management

In the strict sense the term 'Knowledge Management' is a contradiction, since knowledge is not a tangible thing which can be managed. However Knowledge Management is increasingly seen, as signaling the development of a more organic and holistic way of understanding and exploiting the role of knowledge in the processes of managing and doing work, and an authentic guide for individuals and organizations in coping with the increasingly complex and shifting environment of the modern economy. The essence of Knowledge Management is that it creates and nurtures the *environment* that *enables* knowledge to grow, flow and create value in the organization. Hence Knowledge Management *manages the enabling environment* and *not* the knowledge.

For the purpose of this paper, Knowledge Management (KM) can be considered as an umbrella term for a variety of loosely related practices, programs, and technologies associated with leveraging the "knowledge" of organizations for greater performance or competitive advantage. Knowledge Management has to be implemented as an organizational discipline, where the process of capturing an Organization's collective expertise, wherever it resides – in databases, in paper or in people's heads – and distributing it to wherever it can produce maximum pay of, is institutionalized. Knowledge Management creates the necessary environment in the organization for people to create, leverage & share knowledge so that the organization gets the greatest value from the collective knowledge available to it. "Knowledge" in this context includes both the experience, emotions, values, hunches and understanding of the people in the organization, and the information artifacts, such as documents and reports, available within the organization and in the world outside. According to Peter Drucker, knowledge has become *the* resource, rather than *a* resource. "The central wealth-creating activities will be neither the allocation of capital to productive uses nor "labour". Value is created in the organization by "productivity" and "innovation", through applications of knowledge to work. Effective Knowledge Management typically requires an appropriate combination of organizational, social, and managerial initiatives along with, in many cases, deployment of appropriate technology; a synergistic combination of creativity, innovation of human beings with advanced capabilities of information technology. Successful KM implementation demands deep rooted behavioral & strategic change and hence requires the top management initiative and involvement. It also needs the involvement of everyone in the organization. Only then the organization can be viewed as a living organism capable of creating continuous innovation in a self organizing manner.

The central tenet of Knowledge Management is to raise the effectiveness, speed and quality of learning, decision making and customer service at the level of the organization and individual. By institutionalizing best practices existing in organizational pockets,

facilitating greater reuse and better virtual team work, Knowledge Management raises the organization's ability to deliver higher quality output at higher speed.

Knowledge Management is literally changing the way many organizations do business, treat people, effect change, and look toward the future. We do live in a knowledge age, where what you know and with whom you share it not only has value, but also is a competitive advantage. Knowledge Management basically addresses issues related to knowledge sharing/knowledge exchange/knowledge creation within an organization and occasionally beyond organizational boundaries to suppliers, customers and retirees.

Knowledge Management as a distinct management concept, is more than a decade old now, and has finally emerged as a serious candidate for an academic discipline of core management competency.

3. Benefits of KM

Knowledge Management , when successfully implemented provides many tangible and intangible benefits for the organization. A few tangible benefits are given below.

- **Raising the quality of Service**

  The primary mechanism for raising the quality of services delivered to the customer is the institutionalization of best practices residing in organizational pockets, a process which needs the sharing and adoption of these practices across departmental interfaces.

- **Reducing the cost**

  Achieving greater productivity requires higher level of reuse. The cost of redoing something that has been done earlier and relearning something that has been learnt earlier, anywhere in the organization, should be less. Successful implementation of KM leads to reduction in the cost of product/service.

- **Managing risk**

  The de-risking measures in diversifying into new technologies, domains, services and geographical areas, require that the organization must learn new ways of doing things. Managing changes in team compositions resulting from attrition and personal movements require that as much knowledge as possible should be documented

- **Meeting Growth Expectations**

  Maintaining a consistently high pace of growth needs the definition and dissemination of scalable process that support the delivery of high quality customer solutions and an ability to rapidly enable new recruits on technology, process and cultural issues.

- **Managing Virtual Teams**

  Increasingly globalized operations and rising customer expectations have meant a more complex execution model, often requiring teams that are spread across continents to collaborate in delivering single customer solution. Such virtual team work requires

a mind set of working with co-workers who may be situated in different time zones and may belong to different cultures, and good technologies to support communication and collaboration.

The intangible benefits include increased goodwill of customers, cultural change that results in more synergistic organization etc.

4. KM Success Stories

Many organizations have successfully implemented Knowledge Management. Following is a partial list of such organizations .

- Infosys Technologies Limited
- Ashok Leyland
- Oracle Corporation
- IBM Corporation
- Tata Consultancy Services
- Tata Steel
- Maruti Udyog Ltd
- Patni Computer Systems
- McKinsey& Company
- Xerox Corporation
- World Bank
- US Government

5.0 Global KM Initiatives

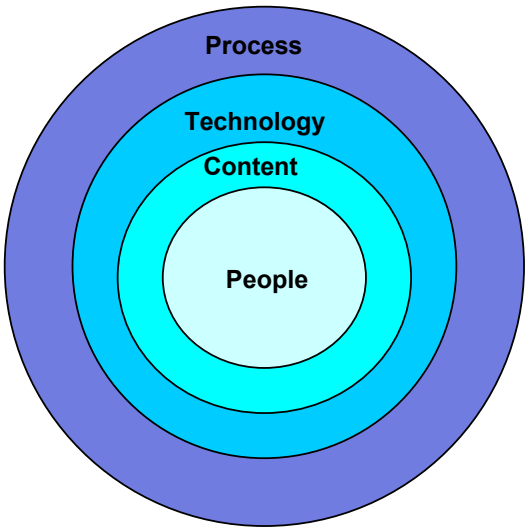
There are many global initiatives in the field of Knowledge Management. Ark Group is organizing annual conferences KM Asia, KM Europe, KM Australia. They also publish KM Magazine. American Productivity and Quality Centre(APQC) holds regular conferences in KM. A large number of organizations have joined APQC to benchmark their KM practices. 'Most Admired Knowledge Enterprise' (MAKE) awards , in the categories of MAKE Global, MAKE Asia, MAKE Europe, MAKE North America & Make Japan are given annually. IBM Institute for Knowledge-Based Organizations is a global consortium of member organizations engaged in understanding how organizations derive value from Knowledge. KMPro(Knowledge Management Professional Society) conducts certification programs CKM(Certified Knowledge Manager) and CKA(Certified Knowledge Agent). The Gurteen Knowledge Community is a global learning community of over 12,000 people in 138 countries across the world, who are committed to making a difference; people who wish to share and learn from each other and who strive to see the world differently, think differently and act differently. They also publish 'Global Knowledge Review'. KMSI(Knowledge Management Society of India) , is a community of KM professionals in India to share learnings in the emerging field of KM. Other global forums for knowledge Management are KM World, KM University GKEC(Global Knowledge Economics Council) etc.

6.0 Implementation of KM

There are two approaches in implementing KM across the organization. The first one is an evolutionary approach in which the first set of systems are used to popularize the idioms, demonstrate the utility and obtain the acceptance of significant sections of people within the organization. These systems are subsequently expected to evolve quickly to cater to the user feed back and scale up for future requirements through a responsive, networked resources that would help, create and evangelize the new systems on a continuous basis. The second approach is to implement KM across the entire organization at a single stroke. The evolutionary approach, though more time consuming, is preferred because of its high probability of success.

It is to be noted that KM is a discipline and it is in support of achieving organizational goals. The major dimensions of KM are people, process, technology and content as shown below.

KM - Dimensions



6.1 People

The principal component of a successful KM practice is its people dimension, which can be divided into KM Organizational roles & incentive scheme.

6.1.1 KM Organizational Roles

The KM organizational roles in smaller organizations can be a centralized one, while in larger organizations it needs to be facilitated decentralized one. In facilitated decentralized approach, the central KM Group will provide over all directions and will be responsible for centralized functions such as KM policy, KM technology, KM research, KM publicity, Content management, sourcing of external KM contents etc, while field practitioners at unit level will be responsible for decentralized functions like internal KM contents, content review, content certification etc.

6.1.2 Incentive Scheme

Rewards and recognition address the universal question "what is in it for me"? Measures for incentivizing the employees' participation in the acquisition, submission, reviewing and reuse of knowledge assets are very important in spreading the KM culture in the organization. In the early stages of implementation, providing the right quantum and kind of incentives is considered as an important success factor for the KM movement. Incentives are provided to all the important stake holders in the KM process – authors, reviewers and users. Authors are to be rewarded when their knowledge asset is accepted for publication. Subsequent to the publication, authors can also be rewarded by employees who find the knowledge asset useful by assigning suitable grades to the knowledge assets.

Additionally users can also provide textual feed back on the utility of the document or suggestions for improvement. Subject matter experts are to be rewarded for their efforts in assessing documents submitted for publication. Users of knowledge artifacts are rewarded for their contribution in the spread KM culture and resultant organizational benefits. The quantum of rewards which employees get is an index of their contribution and involvement in the knowledge sharing across the organization.

The material rewards provide visibility to the KM initiative and shapes beneficent participatory behavior in the short term. However, a successful KM program needs to go beyond material rewards in the long term. An integration of KM as an activity and a concern into the reflexive aspects of working in communities and public recognition as a tool to further this can be used as a major motivational factor for knowledge sharing.

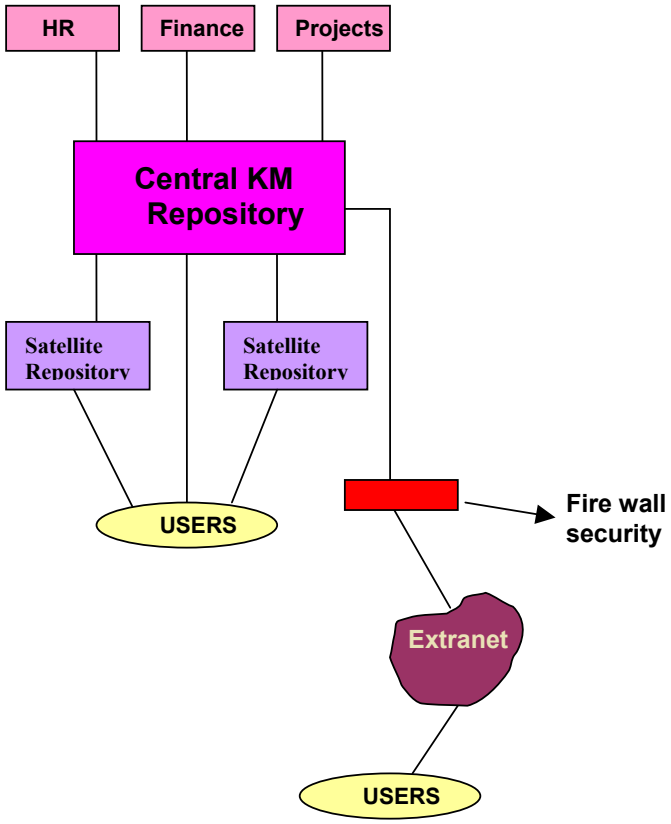
6.2 Content

Providing context for every piece of content in the knowledge repositories of the organization is a basic concern in the definition of the content architecture. The organization of assets along with different dimensions as specified by the content architecture helps the provision of the context and facilitates easy retrieval of the right type of knowledge asset corresponding to a user requirement. Different content types are to be identified based on the requirements of the organization. Typical content types can be project experiences, product reviews, frequently asked questions etc. in the relevant areas of the organization. Knowledge assets are also to be tagged based on the roles for which they are considered most suitable and useful and by security parameters that allow access only to relevant roles. Each knowledge asset can also be tagged by a single grade number which can be arrived by the weighted aggregation of the grades given by the subject matter experts, multiple users of the asset and also by the frequency and recency of its use. This grade number is a market determined indicator of the quality of the knowledge asset.

6.3 Technology

The technology architecture should provide all the basic functionality and features associated with robust, scalable and secure enterprise knowledge portal. The suite of applications available through the portal has to be customized to suit the business process of the organization and cater to the knowledge needs of the employees (within intranet and extranet) which can vary from self learning to online collaboration.

A typical knowledge portal contains a central KM repository and several satellite repositories which can be accessed by both intranet users and extranet users as shown in the following figure:



The rationale behind the satellite repository system is to permit specialized Departments in the organization to own content relevant to their areas. The knowledge portal needs to have integration with various corporate databases such as HR, Finance, Projects etc. The central components of the knowledge portal are the easy to use knowledge navigator utility with contextual and powerful browse / search features, personalization, subscription , user specified knowledge hierarchy, keyword-based search criteria, expert locator, collaboration tools, online learning tools etc. The portal shall also include features such as online document submission facility, review and publication work flow, online chat facility and innovative ways of showcasing new and relevant content.

6.4 Process

The process architecture covers those processes that are necessary to manage various KM functions and that will facilitate the KM movement. The fundamental KM processes are acquire, disseminate and utilize. One important guiding principle is to integrate and leverage to the extent possible, existing reporting processes to extract information for potential reuse across the organization and thus limit the creation of additional process exclusively for KM. The technology architecture facilitates online submission, review and publication of various knowledge assets and integration with various organizational applications such as HR, Finance, Projects etc. The content architecture facilitates the utilization of knowledge assets by various users.

7. Assessment of Benefits of KM

The benefits of KM can be assessed based on various parameters such as the number of knowledge artifacts available in KM repository, the number of knowledge artifacts used in various projects, the number of Knowledge Workers, the quantum of rewards etc. KM benefits can also be assessed based on the increase in productivity, quality, customer satisfaction etc. Internal survey also can be conducted to assess the benefits of KM.

8. Challenges in Implementing KM

The main challenge in implementing successful KM is convincing employees to the benefits of KM and motivating them towards the sharing culture. Those people who do not like to share knowledge can be divided into two groups . One group do not like to share because they think "knowledge is power" and by sharing knowledge they loose their power. The second group think that by using someone else's knowledge they become inferior. Both these groups can be motivated by the incentive scheme suggested.

9. Second Generation Knowledge Management Systems

Hitherto the focus was on explicit knowledge. The following part of this paper focuses on tacit knowledge , the interaction between tacit and explicit knowledge and the resultant knowledge creation.

Organizational knowledge creation is defined as the capability of an organization as a whole to create new knowledge, disseminate it through the organization, and embody it in products, services and systems. Organizational knowledge creation is the key to innovation. The 'Second Generation Knowledge Management Systems' are basically 'Knowledge Creation' systems. The distinction between explicit knowledge and tacit knowledge is the key to understand the difference between the 'First Generation Knowledge Management Systems' and 'Second Generation Knowledge Management Systems' . The former places emphasis on explicit knowledge and the latter on tacit knowledge. The approach taken in a typical organization which implements 'First Generation Knowledge Management System' is "*What is important is to find useful knowledge, bottle it, and pass it around"*. On the contrary the approach taken in 'Second Generation Knowledge Management Systems' is "*There is a great big river of data out there; rather than building dams to try and bottle it all up into discrete little entities, people are given canoes and compasses"*. These systems address the knowledge creation and knowledge dissemination process that are participative and anticipative. Instead of a formal step-by-step rational guide they favour a set of guiding principles for people to understand "not how it should be done" but "how to understand what might fit the

situation they are in". It is assumed that "only a few rules, some specific information and a lot of freedom". The policy statement of a typical organization which implements 'Second Generation Knowledge Management Systems' is *" Use your good judgment in all situations; there will be no additional rules."*

9.1 Knowledge Conversion

Explicit and tacit knowledge are not totally separate. They are mutually complementary. They interact with each other in the creative activities of human beings. Knowledge is created and expanded through social interaction between tacit knowledge and explicit knowledge. This interaction gives rise to the following four modes of knowledge conversion.

- Socialization (tacit to tacit): Socialization includes the shared formation and communication of tacit knowledge between people, e.g., in meetings. Knowledge sharing is often done without ever producing explicit knowledge and, to be most effective, should take place between people who have a common culture and can work together effectively. Thus tacit knowledge sharing is connected to ideas of communities and collaboration. A typical activity in which tacit knowledge sharing can take place is a team meeting during which experiences are described and discussed.
- Externalization (tacit to explicit): By its nature, tacit knowledge is difficult to convert into explicit knowledge. Through conceptualization, elicitation, and ultimately articulation, typically in collaboration with others, some proportion of a person's tacit knowledge may be captured in explicit form. Typical activities in which the conversion takes place are in dialog among team members, in responding to questions, or through the elicitation of stories.
- Combination: (explicit to explicit): Explicit knowledge can be shared in meetings, via documents, e-mails, etc., or through education and training. The use of technology to manage and search collections of explicit knowledge is well established.
- Internalization (explicit to tacit): In order to act on information, individuals have to understand and internalize it, which involves creating their own tacit knowledge. By reading documents, they can to some extent re-experience what others previously learnt. By reading documents from many sources, they have the opportunity to create new knowledge by combining their existing tacit knowledge with the knowledge of others.

Knowledge creation is a "social" process between individuals as well as between individuals and an Organization. But in a strict sense, knowledge is created only by individuals. An Organization cannot create knowledge by itself. What the Organization can do is to support creative individuals or provide the contexts for them to create knowledge. Organizational knowledge creation, therefore, should be understood as a process that "organizationally" amplifies the knowledge created by individuals and crystallises it as part of the knowledge network of the Organization.

9.2 New Organizational Environments and Changing Knowledge Needs

Institutionalization of 'best practices' may facilitate efficient handling of routine, 'linear,' and predictable situations during stable or incrementally changing environments. However, when this change is discontinuous, there is a persistent need for continuous examination and renewal of the basic premises underlying the 'best practices' stored in organizational memory. The 'First Generation Knowledge Management Systems' are largely devoid of such capabilities needed for continuous learning and unlearning processes mandated by an increasing pace of discontinuous and radical change. Such processes of ongoing knowledge creation are needed for organizational survival and competence in the new environment. The radical and discontinuous change of the new environment upsets the traditional organizational response of predicting and reacting based on pre-programmed heuristics. Instead, it demands what may be characterized as 'anticipation of surprise.' As observed by Steve Kerr;"the future is moving so quickly that you can't anticipate it...We have put a tremendous emphasis on quick response instead of planning. We will continue to be surprised , but we won't be surprised that we are surprised. We will anticipate the surprise." How can one move beyond the 'First Generation Knowledge Management Systems' based primarily on predictive models, to systems that can facilitate anticipation of surprise?

9.3 Towards Knowledge Creation Systems

One possible solution may be knowledge creation systems. These systems do not reject the notion of 'best practices' per se but consider the continuous construction and reconstruction of such practices as a live process. Such systems need to contain both learning and unlearning processes. Continuously challenging the current 'organizational way,' such systems are expected to prevent the core capabilities of yesterday from becoming core rigidities of tomorrow, by unlearning, ineffective best practices . These systems add the context, synergy and trust necessary for creating actionable knowledge. The role of such systems seems relevant in ensuring that the organization is doing the *right thing*, in contrast to the optimization based predictive models that focus on doing *things right*

9.4 Towards Actionable Knowledge

The traditional view of knowledge management has treated knowledge in terms of prepackaged or taken-for-granted interpretation of information. However, this static knowledge works against the generation of multiple & contradictory viewpoints and interpretation of the problem based on multiple views , that are necessary for meeting the challenge posed by the continuously changing environments. A more proactive involvement of human imagination and creativity can perhaps facilitate greater internal diversity of the organization that can match the variety and complexity of the changing environment.

9.5 Constructive and Dynamic Nature of Knowledge

Knowledge resides in the user and not in the collection of information. It is to be emphasized that only human beings can take the central role in knowledge creation. The business managers need to realize that unlike information, knowledge is embedded in people and knowledge creation occurs in the process of social interaction.

The above argument suggests that the role of human sense making processes in organizational knowledge management is crucial for sustaining organizational

effectiveness. What are the implications for the organizations and their members given the changing organizational environment that demands increasingly faster cycle of new knowledge creation?

9.6 Towards Communities of Practice

Communities of Practice(CoPs) offer an effective solution for new knowledge creation. At the simplest level, CoPs are small groups of people who have worked together over a period of time. Not a team, not a task force, probably not even an authorized or identified group. People in CoPs can perform the same job or collaborate on a shared task or work together on a product . They are peers in the execution of "real work." What holds them together is a common sense of purpose and a real need to know what each other knows. Through the active process of the community, tacit knowledge is shared—thus ensuring that it is not 'locked up' in one individual. Such communities may typically evolve through stages like preparation, warm-up, operation, and eventual consolidation. Exchange of tips, stories, case studies, and jokes form important knowledge transfer mechanisms and bonds within the community. There may be many communities of practice within a single organization and most people may belong to more than one of them. The challenges for those responsible for harnessing the organization's knowledge creation activity are to identify and nurture these groups and create conditions that facilitate the emergence of new CoPs.

CoPs which is an outgrowth of the quality circles and the evolution of networked/ learning organizations, represents a way of making explicit an organization's untapped brainpower, creativity and potential. These communities can become centers of expertise which can be highly leveraged in an organization prepared to capitalize upon economic change and unexpected business opportunities.

In the traditional view, CoPs may be perceived as professional societies or they may represent collections of people who transcend any individual discipline, society or profession. If the future belongs to those who are able to transform boundaries, then participants in these CoPs may represent the leadership of tomorrow.

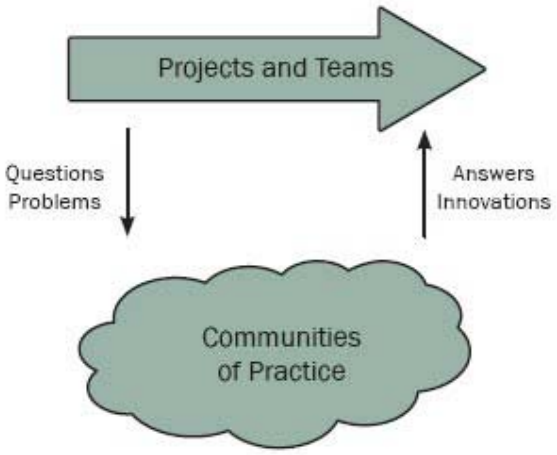
But there needs to be one compelling force which binds the entire organization together—one that creates a common language and shared purpose. Leadership must come from the top-down, bottom-up and middle-out. In other words, everyone has ownership.

Innovation strategy may be the one way to create a simple solution amidst all the complexity. Every function has undergone dramatic change over the past few years—as a function, as a profession and as a collection of expertise across traditional boundaries. Each has learnt from the other and contributed to their body of knowledge. Individual CoPs have evolved on virtually every topic imaginable. It is this cross-boundary sharing and learning which is accelerating the momentum of progress. Today, we see the emergence of the collaborative enterprise as well as cross industry collaboration, marshalling the talents of people throughout the organization to create and move ideas. They have discovered the value of their complementary competencies and how they contribute to the whole. They are all active participants in this rapidly maturing community of knowledge practice. This is precisely the reason that the movement towards Knowledge Management and Knowledge Creation are so fundamental and CoPs are fertile grounds for knowledge creation. However it is common to hear people lament : "But my

people are so busy just getting the job done. We don't have time to sit around and chat about these things". This attitude makes it difficult to get a community, of the ground from scratch. However if middle managers can be convinced of the benefits the current projects will get from CoPs , then their commitments for CoPs will significantly increase.

9.6.1 Link between CoPs and Project work

Middle managers are under great pressure to 'get the job done', and if they believe that an initiative for a CoP is merely a distraction, they will mount a great deal of passive resistance. The concept of 'the arrow and the cloud' described below illustrates how teams provide direction for communities and how communities can support team



The 'arrow' represents projects that must be completed by a specific date and that must deliver a specific outcome. A successful project has a clear direction and definite deadlines to meet. Most organizations focus their resources on the arrow. The 'cloud' represents a CoP. It is more interested in the learning journey than the destination. Outcomes are less clear. Although the arrow and the cloud are quite different organizational entities, one informs the other.

For example, every project invariably faces challenges and problems that must be overcome. If the project team knows that a CoP exists within the organization, the team can pose questions to this community—which represents a network of experts and expertise. Moreover, by receiving 'real-world' problems from a 'real-world' project team, the CoP can focus its efforts on a subject that is valued by the organization. CoP can inform project teams by briefing them on new thinking, models, and tools relevant to the project.

However, a balance must be struck—too much direction from project teams or other management teams can result in the community resisting such direction, and, at worst, disbanding its activities altogether. The key to the 'arrow' and the 'cloud' working together is that each must be aware of the existence of the other. Ideally, project-team members should also be community members, and vice versa.

9.7 Toward Communities of Knowledge Intrapreneurs

Regardless of the industry or organization an individual is working in, he or she is expected to act more and more as an internal entrepreneur, or *intrapreneur*. Given the increasing relevance of the knowledge value chain in the organizational business processes, one can anticipate that most individuals in knowledge-based organizations would be acting as knowledge *intrapreneurs.* The term *'knowledge intrapreneur'* seems more appropriate in this context than 'knowledge worker' given the changing nature of organizations and work roles. The new work roles demand that every worker, act to an extent as a manager as well as an entrepreneur in the organizational knowledge-creation process. Such *knowledge intrapreneurs* are expected to contribute to the organizational knowledge-creation processes based on developing knowledge relationships and knowledge exchanges within and outside the formal boundaries of the organizations.

10. Conclusion

Metcalfe's law states that the usefulness, or utility, of a network equals approximately the square of the number of users of the system (n^2). Applying the same law to knowledge sharing communities, it can be stated that the value people perceive in joining a community grows very rapidly as the number of people who already are members of that community increases. Though the initial efforts required in convincing people on the benefits of knowledge-sharing is extremely high, as more and more people get convincied , it becomes easier to convince still more. It may be kept in mind that knowledge sharing cannot be supervised, nor forced out of people, but happen only when people cooperate voluntarily. However with a judicious mix of awareness, motivation and facilitation steady progress can be made in knowledge sharing. As the knowledge-sharing community grows and reaches the '*critical mass'* , the movement becomes self sustaining. KM becomes part of the organizational fabric, and sharing becomes an integral part of every role across the organization, and thus reducing the effort required for dedicated KM roles. It creates a paradigm shift in the thinking of people from "knowledge is power" to "knowledge *sharing* is power".

One can anticipate that the new paradigm of knowledge creation and dissemination would have implications for most types of knowledge work. The paradigm shift is anticipated to have implications for traditional channels of knowledge creation and dissemination. It is also anticipated to facilitate the democratization of policy-making processes that influence specific groups and communities. It is anticipated that a balance between the technological and human elements of future knowledge management systems would facilitate both learning and unlearning processes. This balance is anticipated to result in systems that facilitate 'anticipation of surprise' demanded by the continuously changing organizational environments. The resulting knowledge creation systems would balance the emphasis on *efficiency* and *effectiveness*. Such systems are anticipated to take into consideration context, synergy and trust necessary for translating information into actionable knowledge. Such systems would also address the long-term and ongoing knowledge creation needs of the organizations served by knowledge *intrapreneurs,* who can practice the policy *"use the good judgment always"*.

As organizations are facing increasing complexity and unpredictable environment, coupled with higher attrition rate of trained personnel, more and more organizations are realizing the importance of Knowledge Management. Already many organizations are implementing Knowledge Management, to effectively meet the challenges of 21st century

and it is matter of time that other organizations embrace the discipline. However organizations need to ensure that the *effectiveness* of decision making is not sacrificed at the *altar* of increased *efficiency*.

References

1. Madanmohan Rao., *Leading with knowledge*.,Tata McGraw-Hill,2003.
2. Liebowitz ,J., *Knowledge management handbook* .1999. CRC Press.
3. Malhotra, Y., *From Information Management to Knowledge Management: Beyond the 'Hi-Tech Hidebound' Systems.* Knowledge Management for the Information professional,2000.
4. Marwick,A.D., *Knowledge Management Technology*. IBM Systems Journal,2001
5. Shawn Callahan., *Want to manage tacit knowledge?* Leader Values.,2005.
6. Sveiby KE., *Knowledge Management – Lessons from the Pioneers*.,2001.
7. Hirotaka Takeuchi., *Beyond Knowledge Management: Lessons from Japan.,*1998
8. Malhotra Y.*, Knowledge Management for Organizational White-Waters: An Ecological Framework*. Knowledge Management(UK), 1999

Design of Knowledge Management System for higher education : Study

Arumugam Rathinavelu

Abstract

Educationists adopt new technologies to fulfill the various requirements of the learners. Knowledge management (KM) is one of newly developed theme to improve continuous teaching-learning practices in an educational institution. The serious concern is one of improving the quality of teaching-learning process to retain teachers in the college and to produce students who are employable. Extreme shortage of qualified faculty is the major reason for the lack of quality in education system. Education is closely associated with KM to identify, create and share updated information and existing knowledge with every one.

Recent development in information and communication Technology (ICT) helps faculty to create and share high-quality multimedia contents through web based knowledge sharing system. I propose a new system to support collaborative learning among teachers and students. In every higher educational institution, exchange of ideas and opinions between professors and students would definitely increase the possibilities for successful results. In addition to this, there is necessity for higher educational institutions to take initiatives to share knowledge and experiences as resource material among teaching faculties and students.

In this proposed learning environment, students get necessary information about Curriculum, Assignments, Test, Lecture notes, References, URL links, Latest Technology information, White papers, Discussion forum, Desktop Video Conferencing and so on. In every higher educational institution, Teaching faculties don't get an effective knowledge share tool to create and share their knowledge base among themselves and with student communities. This paper will assess the need and demand for using Information and Communication Techniques (ICT) to develop a knowledge share tool to improve effective Teaching – Learning Process effectively.

Introduction

Globally, educational institutions are increasingly moving towards the delivery of courses using computers to provide students with the opportunities to learn at their own pace, together with a reduction in traditional lectures. There is also a trend to provide access to the courses via the Internet or intranet (Kemm, R.E., 2001). According to Johnson and Johnson (1992), cited in Kemm, R.E.(2001) collaborative learning is recognized as a potential factor in supporting the development of higher order cognitive abilities. K-Share is an e-learning model developed to extend collaborative learning opportunities to students at our college.

The idea behind K-Share is to create the lecture content materials by faculty members to deliver to students throw web during regular course work. Colleges may use many teaching-learning tools such as web CT, blackboard from leading web –based technologies. But a customized knowledge-sharing tool is developed to fulfill the need of any higher education institution. The efforts are put in to develop and integrate

Dr.Mahalingam College of Engg and Technology,Pollachi, TamilNadu: E-mail:Starvee@yahoo.com

knowledge management practices along with e-Learning concept for subsequent improvements in their (university exam) results.

The global aim is in developing online learning content by lectures and sharing it among faculty members and then deliver to students during the time of learning. Our concept of knowledge-sharing system consists of

1. Simple user friendly workplace in web.
2. Easy tools to include lecture notes in the format of word doc, PPT, PDF and so on
3. Provide link to required web resources
4. Access to respective subjects each semester by faculty and students.
5. Student will provide feedback about usage of k-share
6. Web OPAC to find out resources
7. Discussion forum
8. Email system

This paper will discuss about the e-learning approach for students of Indian institution, Knowledge management practices for faculty members of higher educational institution, benefits of knowledge sharing concept in education sector, Components of K-Share system, improvements by using K-Share system.

Results so far suggest that the usage of K-Share system motivates self-learning of student and their academic performance. At the same time, the level of confidence is found high with the faculty members who involved in content development and sharing.

Education System

Over the past decade there has been a growing concern about the role of higher educational institutions and how they are meeting the needs of employers. Increasingly, higher education institutions are being asked by industry, government and higher education funding bodies to produce graduates with versatile workplace skills, as well as subject-specific skills (Luca et al, 2001). In general, the education sector could be described as one where there is slow creation and diffusion of knowledge. The general question is when KM is a driving force for change else where, how is the education system responding and how should it respond? (Walshe,J., 2002). In India, colleges are generally affiliated with reputed universities in every region. Universities do frame the syllabus and then the affiliated colleges prepare lecture materials for classroom teaching in respective places. The university will conduct the final exam. In this set up, every college prepares their own lecture materials and delivers the same to students. The preparation of lecture material and content may vary from place to place and teachers teach the same syllabus differently. Because teaching is based on the skill of teacher, background of student, method of teaching and teaching aids. Dr.Mahalingam College of engg and technology is established with Internet and intranet facilities for about 1000 computers for various branches of study in Engineering, Science and Technology. Campus is wired with fiber optic cable with 4 Mbps Internet bandwidth connectivity with redundant capacity. Faculties are trained to use of Internet tools and Interactive multimedia tools for classroom experience since 2001. Video conferencing facility is extended with international institutions for collaborative project work. Since the management has implemented facilities in information and communication technology(ICT),our students are doing better learning activities at our institution. Our role is to ensure the same level of input to all

students equally and also to provide quality – learning materials to the students . The lecture content is generally screened and advised by the appointed Subject Matter Experts (SMEs) before delivering to students. Presently, Our K-Share system is made available through intranet for students to access the learning content within our institution.

E-Learning Approach

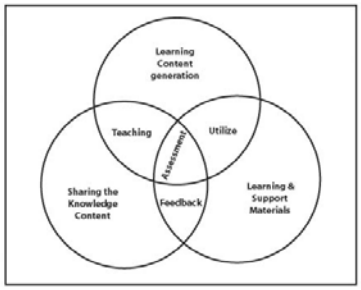
Web-based learning approach is proposed to improve the self-paced learning of students. In general, K-Share tool is aimed to improve, effective teaching-learning process between teachers and students.Students were involved to give necessary and feedback about 'K-Share' during their course of learning . K-Share is developed to share learning support materials for the courses of computer science and engineering & information technology dept (CSE and IT). Most of the subjects are well suitable in nature to prepare electronic learning materials by the faculty of CSE & IT under the guidance of Subject Matter Experts (SMEs).
K-Share is an integrating knowledge management suit to produce innovative and effective knowledge sharing environment for both teachers and students. The role of teaching faculty is to lead students to necessary information and knowledge resources. Students can get necessary assistance from any experts to solve their queries or problems in the case of virtual learning workplace (Kiili,M., 2001). Since most of the colleges are having very limited trained and qualified teaching faculty, our k-share system do focus the benefit of learners. According to Kiili (2001), the traditional idea of young student attending lectures and seminars on a daily basis is not the usual solution for getting high education any more. This means offering wide-range knowledge sharing platform reduce their anxiety and improves their confidence in learning things much faster than traditional teaching environment.

Traditional Classroom and ICT

In general, the learning process is depends on the involvement level of students and teacher in a traditional classroom. Students may come from different background and at the same time, teachers also need to update their knowledge (or) information perfectly. Much of a community's knowledge lies within its documents, discussions, decisions, conceptual models, formal educational modules, processes, and the awareness by members of other members' expertise. For an academic professional society, community documents include books, journal articles, conference papers, audio tapes, videos, still pictures, course syllabi, and tutorials (Bieber et al., 2002). WWW Techniques are implemented here to develop effective k-share tool by linking and navigating those documents for better knowledge sharing platform among people. According to Newman,B. (1999), the general knowledge model organizes knowledge flows into four primary activity areas: Knowledge creation, retention, transfer and utilization.

A Frame Work for K-Share Tool

A framework is proposed for developing a knowledge-sharing tool for higher educational institution. It consists of learning content generation, sharing of created knowledge content and identifying the learning and support materials. All the three elements are connected here by assessment. The element learning and Support material will be utilized to generate lecture content. The lecture content will be delivered to students after SME's evaluation. Lecturer may need to do some changes or modifications based on feedback from SMEs.



(Fig.1 A frame work for Our Knowledge sharing system)

According to Stephen Denning (1998) as cited in Carayannis,E.G. (2002) defines KM as the process for knowledge sharing. He also states the key dimensions to a knowledge program,

KM Programs have dimensions of collecting and connecting
KM Programs include social processes by which communities of practice enable knowledge sharing to take place
KM Programs support the maintenance of beneficial external partnerships.
KM Programs use IT to assist in knowledge creation and knowledge use (in addition to the most common aspect of supporting the dissemination of know-how)

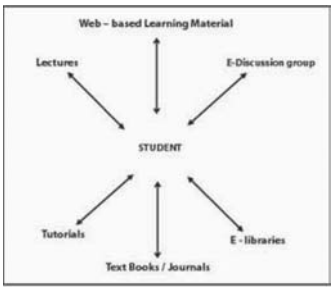
Internet is the powerful technology for sharing the knowledge in the case of electronic learning. Computer software programs are being used in timetabling and school management to improve the use of staff time, student time and space, thus reducing costs significantly (Hildreth et al., 1999).

Components of K-Share

The purpose of K-Share is to improve student learning and outcomes in the field of higher education and to motivate faculty members to share the available knowledge within them. Faculty members develop learning material for students by using various IMM tools. SMEs and KM co-ordinator will formulate the guidelines to evaluate the lecture notes during preparation and then the lecture notes will be delivered to students through 'K-Share'. Our specific objectives are

To develop a collaborative learning system
To explore the knowledge expertise of faculty within institution
To propose a conceptual model for enabling web-based learning

We refer the framework model of Smith,D. (2000),



(fig.2 , frame work by Simth,D.,2000)

It is aimed to be a low-cost KM tool for every person. Individual faculty, students, staff, and other college/university citizens need access to simple, low-cost, interoperable KM tools. So that they can create and manage content/context for personal use and sharing (EDUCause, Sept/Oct, 2003). To provide better teaching-learning environment, educational institutions must initiate knowledge management practices in their capacity to acquire, utilize and share knowledge by using ICT.

The questionnaire was used to ask our faculty members to give feed back about implementation of KM practices.

(Fig 3 : Screen shots of K-Share tool – an intranet site)

Faculties are motivated to give response for the given questionnaire and based on their feedback, K-share is designed as a web-based knowledge-sharing tool. Faculty members are motivated to prepare and host their lecture notes along with support materials (in the form of PPT, PDF, DOC and so on) and URLs for further references. Students do login to 'K-Share' and to get lecture notes and learning support materials during their course work in every semester. In addition to this,E-Mail, Discussion forum, WebOPAC and e-Books are add-ons in K-Share.

Conclusions

The main objective is to integrate the knowledge management practices and e learning together for developing an effective knowledge-sharing platform between teachers and students. This is to see improvements in their self-learning attitude and academic performance. We aimed the following results by introducing this concept,

Consistency in Content delivery
Quality Content in Learning materials
Students self-learning concept
Performance improvement in the university results.

We found remarkable improvement in all the above according to our best knowledge and data gathering from teachers and students. Data is collected from faculty members, SMEs and students of CSE and IT Dept using questionnaires, online feedback form, observation and interview. This project is also aimed to extend to all other institutions thro internet.

Acknowledgements

Appreciation to the active participation of the faculty members of CSE and IT dept, to develop K-Share suite. The project is fully funded by the management of institution.

References:

1. Bieber et al. (2002). Towards knowledge – sharing and learning in virtual professional communities, Proceedings of the 35th Hawaii International Conference on system Sciences – 2002, IEEE – Computer Society.
2. Carayannis,E.G. Laporte,B. (2002). By Decree (or) by choice? – a case study – implementing knowledge management and sharing at the education sector of the World Bank group, world bank institute working papers.
3. Kemm et al (2001). Collaborative Learning: On – Campus in a Technology based Environment (PP 7.1) , AUC Academic and Developers Conference 2001 Proceedings.
4. Kiili,M. and Kiili,J. (2001). Knowledge management as part of modern education concept in engineering education retrieved from http://www.sefi2001.dk/papers/pdf/6.pdf on 20-08-2003.
5. Luca et al (2001). Designing an On-line Learning environment to support the development of generic skills: a case study (PP11.1) , AUC Academic and Developers Conference 2001 Proceedings.
6. Newman, B., Conrod, K.W. (1999). A frame work for characterizing knowledge management methods, practices and Technologies retrieved from http://www.3-cities.com/~bonewman/newman_conrad%20PAKM2000-16.pdf , knowledge – management forum, on 24-09-2003.
7. Norris et al .(2003). A revolution in knowledge sharing , EDUCAUSE Review, Sept/October 2003.
8. Rathinavelu,A. (2002). Preparation of teaching instructional material by using presentation method of IMM tools ,Lecture given to faculty memebers of engg colleges, AICTE Sponsored FDP, Dr.MCET, Pollachi, INDIA
9. Rathinavelu,A. (2003). Need of e learning and knowledge sharing system for higher educational institution in India , Lecture given in short term training program on Interactive Multimedia sponsored by AICTE and ISTE, Dr.MCET, Pollachi, INDIA.

10. Rathinavelu et al(2004), Knowledge sharing system using ICT to find out improvements in Teaching – learning cycle A case study in an Indian higher education system, EDU-COM 2004,24-26 November,2004, INTERNATIONAL CONFERENCE on New Challenges for Sustainability & Growth in Higher Education, Thailand

11. Smith,D., and Hardaker,G. (2000), E-learning Innovation through the Implementation of an Internet supported learning environment, retrieved from http://ifets.ieee.org/periodical/vol_3_2000/e04.pdf on 06.08.2003.

12. Walshe,J.(2002), Knowledge management in education and learning, forum – oxford summary report.

Knowledge Sharing Techniques

Machine Translation as an Invaluable Technique in Knowledge Sharing

Tessie George

Abstract

In this age of globalisation and mass scale digitisation and computerisation of libraries, there are two types of digital divides which hinder knowledge sharing. The first is between the 'computer haves' and the 'computer have-nots' and the second is between those who do and those who do not understand a particular language, mainly English, in which the text appears.

The paper deals with Machine Translation or E-Translation as the ideal and unique tool to bridge the second type of digital divide i.e. between persons who do not understand the same languages. Traditional translation is touched upon and Machine Translation or E-translation is discussed in detail and as an improvement over traditional translation. The usefulness of E-Translation in different types of 'translation demands' is discussed and future trends are reviewed.

The conclusion is drawn that Machine Translation or E-translation is a necessary tool to enhance free flow of knowledge and to achieve optimum knowledge sharing and thus E-excellence in digital asset management.

1. Introduction

Advances in digital communication have made the world into a global village. But a great digital divide separates people who do not understand and speak the same language. For example, if an anglophile tries to read the annual report of the CEA at Paris or the Japan Atomic Energy Research institute on their Website, he may be disappointed to find that he is not able to understand much of it as it may be in French or Japanese. Likewise scientists from these institutes who understand only Japanese or French cannot read the IGCAR or BARC websites, which are in English. We all face this problem, especially when surfing the net. India is in the throes of an IT revolution and talking about broadband and introducing computers in villages, but only 5% of the people of the country can speak and write in English and the rest communicate in the local language. *

This language divide hinders the free flow of digital information and hampers knowledge sharing and optimum digital asset management. The answer to bridging this digital divide and aiding digital asset management lies in Machine Translation or E-translation.

According to the Oxford Dictionary, translation is the act of rendering the sense of a word , sentence , book etc. in another language. In E-translation, the involvement of the human translator will be minimal and the major part of the work will be done electronically.

* Union Communications Minister Mr. Dayanidhi Maran , The Hindu June 21st. 2005.

Scientific Information Resource Division (SIRD), Bhabha Atomic Research Centre, Mumbai- 400 094
E-mail:tglisd@apsara.barc.ernet.in

2. Traditional Translation

In traditional translation, the translator who is given a translation job reads over it , looks up unknown words in a dictionary , translates the text paragraph by paragraph either writing down the translation or dictating it to a stenographer who takes it down in short hand and types it later.

This is a cumbersome process where heavy, often dusty dictionaries have to be handled . In the editing and proof reading stage, if there are any mistakes to be corrected or changes to be made, untidy marks are made on the written translation or if the translation is typed , it has to be retyped or unsightly white marks made.

The storage and retrieval of these translated texts is also difficult and unreliable. The fragile typed or handwritten texts will be stored in cupboards where they are prone to damage and destruction of various types. They have to be catalogued and retrieved through a cumbersome manual process.

If the translation has to be further disseminated to other institutes or countries, again the slow and sometimes rather unreliable postal system has to be depended upon.

Nowadays translators use the more convenient method of reading the text in the source language and using the computer to obtain the target language text.

Traditional translation may prove an expensive process as translators normally charge rather high fees. In addition translators may not always be available in the required language.

3. Electronic Translation or (E-Translation)

E-translation seems the ideal solution to several of the problems of traditional translation. Here the translation job is read over and the foreign language text can be input in one of several ways.

3.1 Input

The text could be keyed in, it could be scanned and fed in or it could be input using Voice Recognition Software

3.1.1 OCR Scanning Software

This is used to obtain a softcopy of a document , that can if required be later fed into the E-translation software.

The document is scanned page by page As a page is scanned and read over by the OCR software, the mistakes and unrecoganised words are highlighted and if the software is a good one, suggestions are made for corrections of these words.

The user can pick one of the words suggested or make his own corrections and then save the page. This softcopy of the document can now be input into the E-Translation software. It is evident that in order to judge the accuracy of the scanned text, one needs to know the

source language fairly well , which most often is possible only for the human translator, if it is in a foreign language.

3.1.2 Voice Recognition Software

The text to be translated can be read out loud into the Voice Recognition Software which then converts it into digital information which can be fed into the E-translation software.

Of course the Voice Recognition Software has to be 'trained' to recognize a particular person's voice . It also has the facility for improving the voice recognition faculty and for making corrections in the softcopy of the text, so that a near perfect softcopy can be imported into the E-translation software.

Here too, however, it is evident that the text in the source language can be read aloud with some degree of accuracy only by a human translator, if it as a foreign language.

3.1.3 Keying-in

The source text can be keyed in directly into the computer and fed into the E-translation software. Here again the source language can be keyed in with any degree of accuracy only if one knows the foreign language to some extent and even then it may prove to be a rather tiresome task.

3.2 Translation

The translator can then import this softcopy of the foreign language text into the E-translation software or into the translation workstation (or translators workbench) a concept introduced by Trados, which combines features like multilingual word processing, terminology management and translation memories.

3.2.1 Multilingual Word Processing Features

Text can be input and edited in various languages different from English, French or German to even Japanese, Russian or Korean. Foreign language keyboards are available for inputting frequently used languages and virtual keyboards for editing and correcting the target language or less frequently used languages. For example, if French and Japanese are frequently used , one could have French and Japanese keyboards and if one translates rather infrequently into Russian , one can use the Russian virtual keyboard of the translation workstation to edit the translated text. Thesauruses, grammar and spell checks are also available for languages on request.

Of course, it is only a human translator who knows the foreign language, who will be able to edit the foreign language text or use the Thesauruses, grammar and spell check features.

3.2.2 Terminology Management Software

Terminology databases can be created according to the specific needs of the user. This terminology database can later even be updated, changed or accessed as the need arises. The developer of the software can create the terminology database for the translator or user. The facility for on-line updating of this terminology database should be provided for.

This is particularly useful for company specific terminology. For example, BARC could ask the software developer to create a terminology database of nuclear related terms.

3.2.3 Translation Memories

These are unique to Translation Workstations. Once a translator is satisfied with the accuracy of a translation, the source and the target texts can be aligned and stored in the translation memory. This can later be accessed in somewhat the same manner as a dictionary. Every time the same text or a similar text comes up for translation very accurate translation will be rendered because entire sentences or paragraphs from the target language are imported from the translation memory into the new translation.

The translation workstation could also enable the translator to build links to the Internet to access on-line dictionaries or other documentation to aid in the translation.

In addition, many translation workstations are offering fully automatic translation, using E-translation software.

The advantage of these translation workstations is that the translator can use any or all these facilities as he feels necessary for a particular translation.

4. E-Translation Vs Traditional Translation

4.1 Input

The E-translator who has suitable and convenient input methods like OCR scanning or voice recognition software may not need a stenographer at all. Once the text is input, one or many of the various tools in the translation workstation can be used , alone or in various combinations to render the translation.

4.2 Dictionaries

Instead of bulky dictionaries , the terminology databases, E-dictionaries or translation memories can be used to search for words, phrases or even complete sentences. Alternately the text could be directly put through the translation software to carry out the translation and the terminology databases and E-dictionaries can be used to rectify incorrectly translated words or phrases.

4.3 Editing

The editing process on the computer is much more refined – as we all know – than in manual translation where erasers and whiting ink is used. The grammar and spell check tools and Thesaurus can also be used for editing when required.

4.4 Storage and Retrieval

The output of E-translation can be stored on CD ROMs or on the institutes Intranet, to be retrieved in a sophisticated and swift manner with key words and a search engine.

4.5 Dissemination

The output of E-translation can be disseminated to practically hundreds of destinations at the click of a mouse , through Email, unlike in traditional translation.

5. Usefulness of E-Translation in Different Domains of Translation Demand

E-translation has different degrees of usefulness depending on whether the translated data is going to be used for assimilation, dissemination, interchange or access of information.

For **assimilation** of information i.e. when the approximate meaning of the text is required, E-translation is very useful. Though most E-translations do not produce very high quality translation, the user can extract the basic information required from the E-translated text in the target language. After all 'Some translation is better than no translation'.

However when used for the **dissemination** of information, the translated target language text must be very accurate. Here the E-translated text may not be much more than a draft text that may require a lot of post-editing by human translators.

E-translation is most successful for quick **interchange** of information . This is especially true on the internet , when electronic pages have to be translated quickly as in Email, chat rooms and web pages, without much importance being given to the quality of the translation.

There are several E-mail sites that offer instant translation of your Email into several languages. Here the digital divide is bridged with fairly efficient E-translation software that favours speed over accuracy. However the sentences need to be short and with a simple construction, rendering E- translation simple and more accurate.

Of course E- translation proves the greatest digital asset when one can **access** information in any language on the internet , using E- translation software.

6. Future Trends: E-Interpretation

If one can speak of E- translation, why not encourage E-Interpretation where the software recognizes speech in the source language , irrespective of the speaker and inputs it into the E-translation software. It is translated and reappears as text on the computer screen in the target language or is converted back again into speech in the target language.

Visualise this scene in a large international meeting: a conference hall equipped for simultaneous interpretation with a computer screen in front of each delegate – on which she can select to view the language of her choice. Voice recognition software that can recognize any voice without prior training should be available. As the speaker speaks, his voice will be digitized by the Voice Recognition Software and input into the E-translation software and the translated text will appear on the screen in front of each delegate in her chosen language. This digital source language to target language, speech to speech interpretation would seem an ideal way to bridge the digital divide between delegates of the seminar who do not speak and understand the same language.

7. Conclusion

In this age of digital information , the fastest way to bridge the digital divide which separates people who do not understand the same language is E-translation . Hard copies of information ranging from scientific and technical texts, faxes and commercial documents can be digitized, E-translated fairly accurately and disseminated at the click of a mouse. E-translation is of course indispensable in the electronic age of Email, E-commerce and web surfing , when everything depends on the electronic medium.

Distributed Database Federation: An experience with IGCAR Library Network

Karthik Viswanathan,V Soundararajan,E Joseph,JVM .and Somasekharan, M..

Abstract

In the current information era, where data are spread across diverse platforms, the user would find it tedious to access each source of data. This complexity can be simplified with a unique tool that could search the diverse databases and retrieve the desired results to the user on a single window. Federated Search technique would facilitate the above-mentioned task. One such distributed computing technology is the Common Object Request Broker Architecture (CORBA). CORBA is a set of specifications provided by the Object Management Group (OMG) to facilitate communication between distributed systems. The federated search system implemented at the Scientific Information Resource Division of the Indira Gandhi Center for Atomic Research (IGCAR) enables the users of the library to search for data from across the databases on different platforms. The success story of the implementation of this feature is presented as a case study in this paper.

Keywords: CORBA, Z39.50, Interface Definition Language, Linux, Distributed Databases, Federated Search technique**,** Object Management Group**.**

1. Introduction

Information centres, particularly, R&D organisaitons, handle a colossal and complex set of data. Data organization and management has posed a challenge to the information professionals. In this process, varied problems are encountered. One such is that data are stored in adhoc modes. Data may be scattered across different databases and on heterogeneous platforms. A user who searches data need not be aware of the variances in the platforms. The task of retrieving data from those divergent databases cannot be thrust upon the end-user. Hence a federated search technique is vital to simplify the task of the user and to shield him from the complexities of the multifarious platforms. In favour of this argument, IGKAR's experiences in this attempt has been presented in this paper as a case study report.

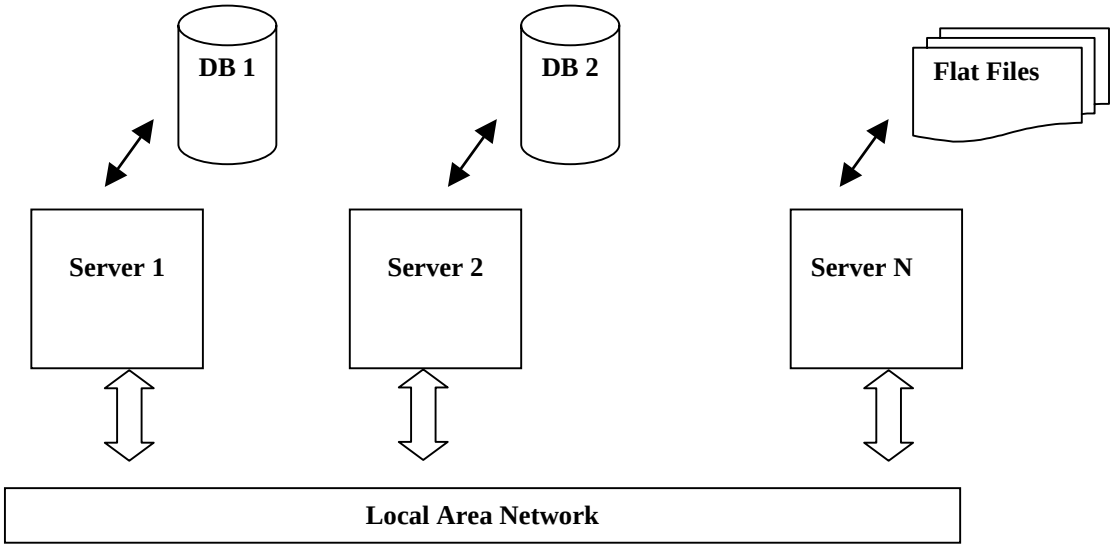
2. Distributed Setup

The Scientific Information Resource Division at the Indira Gandhi Center for Atomic Research (IGCAR) maintains a huge repository of intellectual resources in the form of Catalogues, Journal Articles, Abstracts, Publications, E-Books, Reports, Thesis, Magazines, etc. These resources are stored in different databases and some in flat files, spread across different platforms. Each one has a separate interface and retrieval rules for access. This makes the patrons to remember each system, which will in turn decrease the information availability.

To improve the information access and for retrieval of information distributed across the IGC campus, a system was designed and developed as a pilot study for unifying access. (Figure 1)

SIRD,IGCAR, Kalpakkam-603102. E-mail:carthik_cit@yahoo.com

Figure - 1 SIRD SET UP



3. Federation Techniques

3.1 Z39.50

ANSI/NISO Z39.50 is the American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. ANSI/NISO Z39.50 can be implemented on any platform enabling different computer systems with different operating systems, hardware and database management systems to interoperate and work together seamlessly. It is a client/server-based protocol for information retrieval, which is state and connection-oriented. The Z39.50 standard was developed to overcome the problems associated with multiple databases search such as having to know the unique menus, command language, and search procedures of each system accessed. The Z39.50 simplifies the search process by making it possible to use the familiar user interface of the local system to search both the local library catalogue as well as any remote database system that support Z39.50 standard.

3.2 RMI (Remote Method Invocation)

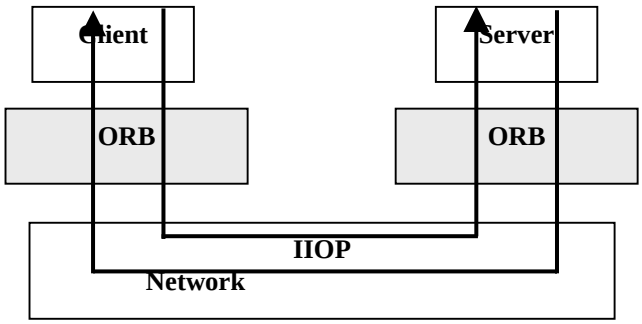
RMI provides a way to access remote objects over the network. RMI is basically an inter-object communication protocol. It is primarily used for communicating different platforms that involves only java. RMI cannot be used with non-java languages. The key point behind RMI is that it is easier to implement, but scaling up the system is difficult. Java RMI is a mechanism that allows one to invoke a method on an object that exists in another address space. The other address space could be on the same machine or a different one.

3.3 CORBA (Common Object Request Broker Architecture)

CORBA is a generic framework for distributed object applications. CORBA delivers language independence through its use of the Interface Definition Language (IDL). IDL and its associated constructs can be translated, or mapped into any language, thus freeing developers and architects to choose the best language for implementation. The crux of CORBA is the ORB (Object Request Broker), which acts as the bridge between two different systems.

The figure 2 depicts the architecture of CORBA. Each system that wishes to access a remote system should have ORB running on it. The actual communication occurs between the ORBs. For the process of ORB communication between different platforms, the Internet Inter-ORB Protocol (IIOP) proves useful. The IIOP translates the ORB requests into TCP/IP calls and sends them over the network to the system at the other end. CORBA offers a better solution when compared to RMI. RMI is language specific; but CORBA can be used irrespective of the programming language and platform.

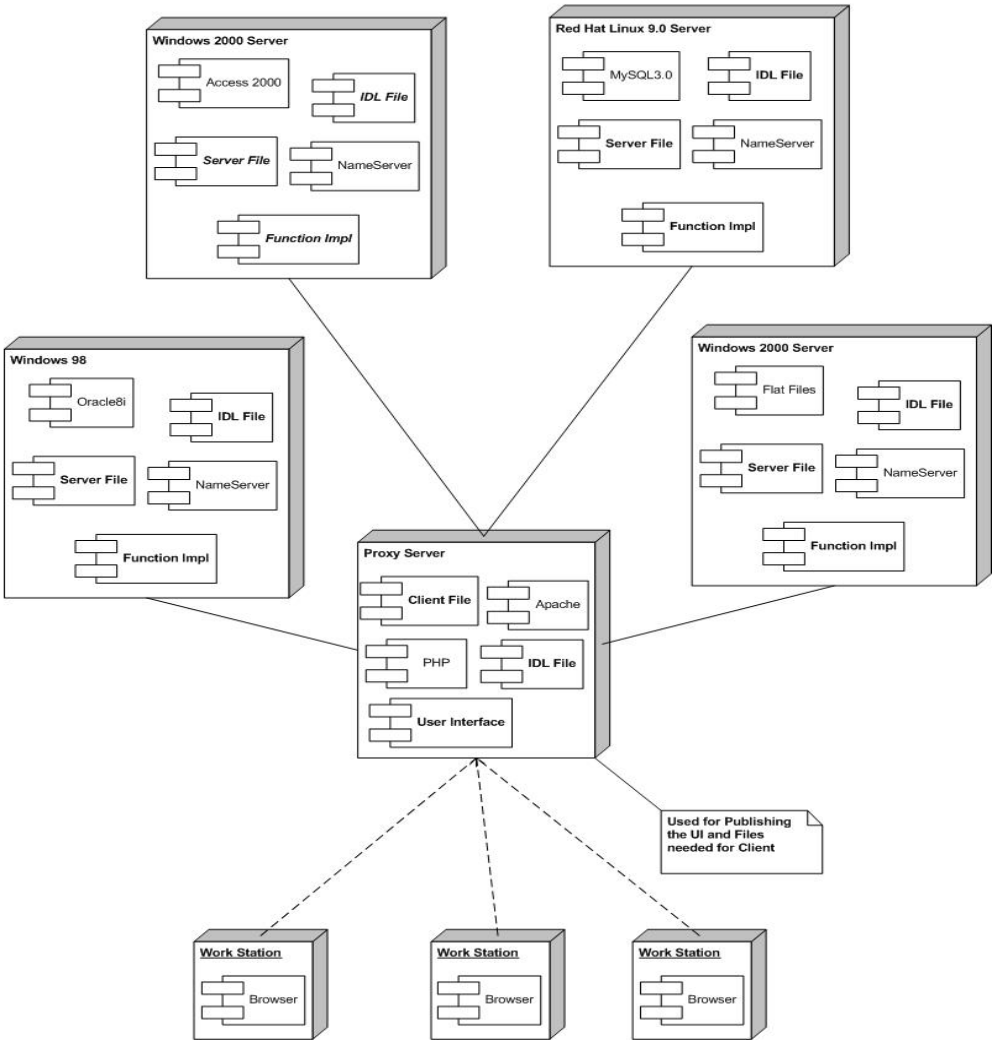
Figure 2 CORBA Architecture



4.Architectural Design

The following figure 3 is the designed structure of the deployment of the CORBA system. The module that acts as the server runs on the systems that contain the databases. It provides the function that services the client request; execute the task of fetching data from the databases and pass the results to the client. The server also runs the naming service. The server registers any object implementation that it provides. Also the client searches for object references in the naming service. The module in the client contains the files that are needed to contact the server, pass the parameters and retrieve the search result from the server. It also contains the necessary user interfaces needed to operate the system. The client module and the user interfaces are published on a web server so that it can be accessed from anywhere.

Figure 3 Deployment Diagram



5. Implementation

The SIRD maintains resources on different databases such as MySQL database on Linux server that holds details of books, Access Database running on a Windows 2000 server that contains details of journals, flat files on Linux server that contains details of publications, etc. The user interfaces and the files needed by the client are published on an Apache web server. The user can login to the system by providing a user id and password. Once the user logs in, the search page is displayed, where the user keys-in the search string and selects the resource to be searched. Then based on the resource, the request is routed to the server that contains the resource.

Java is the language used for developing the client and the server. Functions are defined in an interface definition language and mapped to java using the *idlj* compiler that comes with the jdk1.5. Once the request is passed to the server, the server queries the database, and passes the results back to the client. The client parses the result into a HTML file and this is displayed in the browser window. Figures 4 & 5). For the convenience of the patrons, the system also maintains a profile that lists the keywords that a particular patron

has searched previously. The system also generates reports containing details of the time of search and the resources that was scanned for a search request.

Figure.4 Search Request Screen

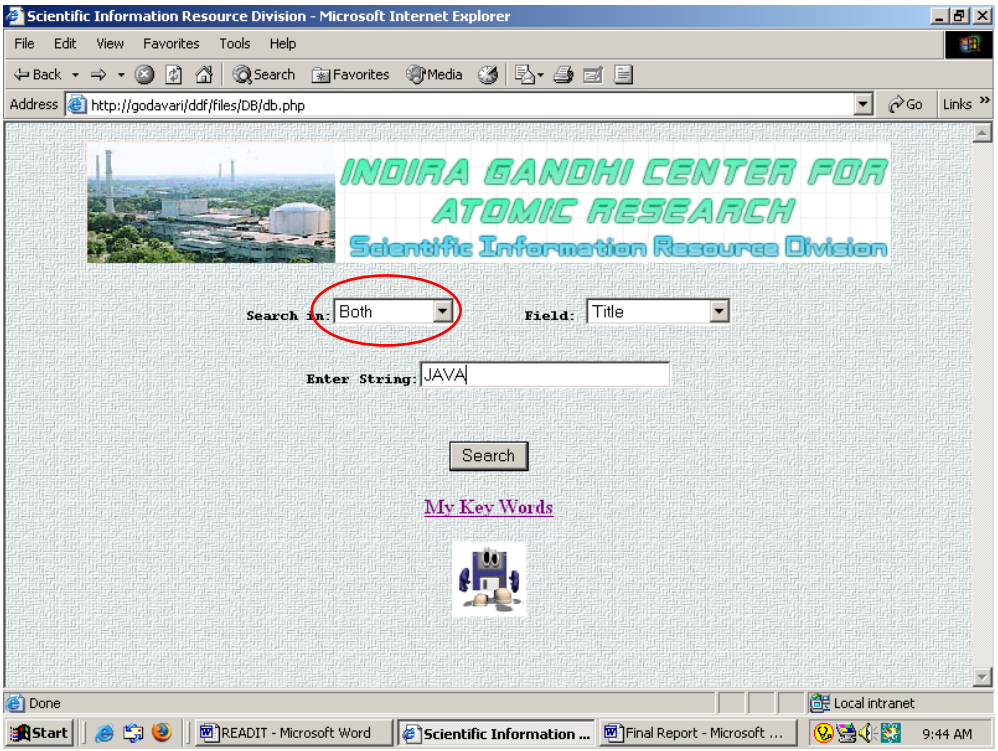
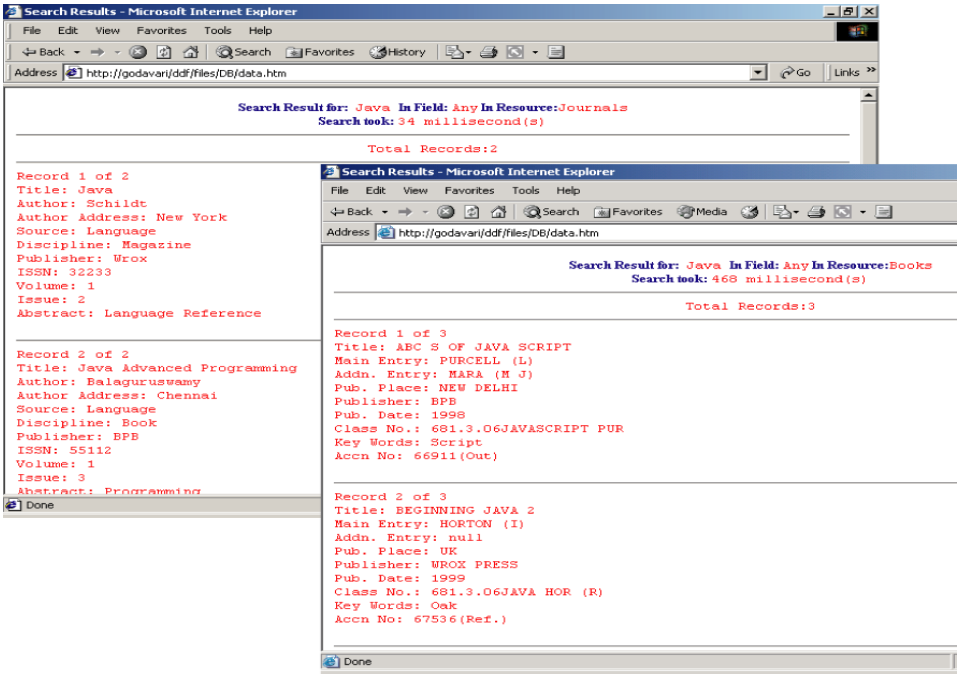


Figure .5 Search Results Screen



6.Suggestions

Based on the experiences at IGCAR, the following suggestions are made to enhance features and functions of the federated search:

- Merged Ranking
- This allows consistent ranking schemes to be applied across all sources, so that the most relevant hits always come out on top when results from different sources are merged.
- Result Filtering
- Unwanted results, such as those unacceptable content, outdated information, and so on, can be filtered out of result lists with custom filtering modules. Similarly, duplicates can be removed to reduce clutter in the result list. If desired, the ranking of results with duplicates can be enhanced to reflect that more than one information source considered them to be important.
- International Language Support
- Today, business is becoming increasingly international, requiring the ability to access information in various languages. An interface that can display several views of search results in multiple languages, along with the tools to connect to content sources from around the world is needed.
- Enterprise Security
- Access to specific documents and any repository should be restricted. Internal authentication models for repositories, proprietary applications, as well as third-party access management software need to be provided. For external information sources, support for source-specific credentials and cookies can be provided -ensuring only the files a user is entitled to see are shown in result lists and category views.

7. Conclusion

A federated search system proves to be useful for any organization that hosts enormous amount of data spread across different configurations. The user is provided with just a single interface using which he/she can access data present on any database. This widens the scope of search since many resources can be scanned. The system even scans vendor databases and fetches the results for the users. The key feature of the system is that, it can connect any system with any configuration. The system suffers the shortcoming of not able to provide the special functionalities that are actually supported by the third party databases.

References

1. Reaz Hoque. CORBA for Real Programmers. Morgan Kaufmann Press, San Diago, 1999.
2. Gerald Brose et al. Java Programming with CORBA. Ed. 3. Wiley, New York, 2003
3. Yarger R.J. et al. MySQL and mSQL. O'Reilly & Associates, Beijing.(Schroff Pub., Mumbai), 1999.
4. Soundararajan, E; Joseph, J V M Jayakumar, C and Somasekharan, M. Digital Library: The Integrated Search System. *In:* International Conference on Digital Libraries: Proceedings. New Delhi, ICDL, 24-27 February 2004.
5. Soundararajan, E; Joesph, JVM; Jayakumar, C and Somasekharan, M. Digital Information Dissemination: A Comprehensive Approach. 22nd Annual Convention and Conference of Society for Information Science: proceedings. Chennai, SIS, 21-23 January 2004.
6. Sun Microsystems URL: http://java.sun.com/docs/books/tutorial
7. W3Schoold URL:http://www.w3schools.com/php/php_intro.asp
8. Object Management Group URL: http://www.omg.org

Coordination Model for Communication Among Agents in Multi-Agent Based Intelligent Systems

Vinay Chavan[1] and Sanjay E. Yedey[2]

Abstract

Autonomous Agents and Multi-agent systems(MAS) represent a new developing area of research. Multi-agent concepts and methodologies are finding increasing application in controlling complex, unpredictable systems in real time. The area of intelligent manufacturing systems aims to use multi-agent systems. A Multi-Agent System (MAS) is a system composed of a several autonomous agents, which cooperate with each other to reach common objectives, while each agent pursues individual objectives. In order for a MAS to solve common problems coherently, the agents must communicate amongst themselves and coordinate their activities. A collective social behavior emerges through this communication. Therefore theories of agent societies are required which defines the world which hosts the society and the laws ruling the world. An intelligent global behavior can emerge from a system. This intelligence can not reside inside agents only but should somehow be spread among the agents and the interaction space among them. Therefore MAS is consist of not only the agents but also the abstractions supporting interactions.

Our aim in this paper is to point the most important theoretical concepts and practical issues associated with the design of MAS and coordination model for agent communication in Multi-agent systems. The coordination model presented as an interaction abstraction aimed at globally ruling the behavior of different system components. Two different types of coordination models, data-driven and control-driven , are considered.

Keywords: Intelligent agents, Multi-agents, Communication, data-driven coordination, control-driven coordination

1. Introduction

In MAS several autonomous agents cooperate with each other to reach collective goal, while simultaneously each agent is responsible to meet its individual objectives. The agents must communicate amongst themselves and coordinate their activities to achieve common goals coherently in the given environment. It facilitates communication and coordination through several rules and abstractions. There are different ways in which agents coordinate each other. There are several coordination models defined which can be broadly classified as: Data driven; and Control driven.

2. Theoretical Concepts

There are different definitions made by researchers for defining intelligent agents. One of them is that 'An agent is an autonomous entity which performs a given task using information gathered from its environment to act in a suitable manner so as to

1 *Sr. Lecturer, Dept. comp. sci. and Tech, Porwal College, Kamti , Nagpur*
2 *Sr. Lecturer, P. G. Deptf Comp. Sci.and Tech., Shree Hanuman Vyayam prasarak Mandal,Amravati-444605*. ***Email*** *: arc_san2000@yahoo.com*

complete the task successfully. It should be able to change itself based on changes occurring in its environment, so that a change in circumstances will still yield the intended result'.

An agent is characterized by the following properties:

- Autonomy: Operation without the direct intervention of humans or others, and have some kind of control over their actions and internal state.
- Social ability: Interaction or communication with other agents
- Reactivity: Perception of their environment and response in a timely fashion to changes that occur in it.
- Pro-activity: Response to their environment and exhibit goal-directed behavior by taking the initiative. An agent is capable of handling complex, high-level tasks. The decision as to how such a task is best split up into smaller sub-tasks.
- Temporal continuity: agents are continuously running processes.
- Mobility: the ability to transport itself from one machine to another, retaining its current state.
- Anatomy: Similar anatomy as that of objects. However It's state and behavior is expressed differently.

The agent's execution model contains:

- Static knowledge on itself and other agents (acquaintances).
- Expertise knowledge the agent represents, which can be described in various forms such as production rules, frames, logical expressions, etc.
- Reasoning: the inferences which draw the problem resolution.
- Communication; the communication protocols between the agents.
- Cooperation strategies used by the agents to cooperate with others.

As MAS are usually made up of large number of independently designed software components (agents), it is very difficult to exactly design and handle the system as a whole. It's not enough to simply put components together and let them interact. Rather a systematic approach is needed focusing on the role each agent plays in the system, the mechanisms upon which composition can be based, and their composition laws. A more dynamic and complex environment, also termed as agent space, is required which supports a complex communication among agents. A software architecture, which is interaction-oriented, rather than composition-oriented is needed

The agents in multi agent systems work in a team to achieve common goals by interacting with each other. Several researchers put their efforts in the areas related to agent communication like agent language, architecture, distribution strategies etc. But these efforts focus on the agents' internal structure where as a comprehensive view of a multi-agent system should not rely only upon the analysis of the internal behavior of each agent, due to their intrinsically interactive nature. In fact, these systems are likely to exhibit a complex global behavior, emerging from the mutual interaction among components, that is hard to be described and managed when communication is considered from a single agent's viewpoint. In multi-agent systems a team effort leads to a social behavior. Therefore, to fully understand multi-agent systems, theories for agent societies are needed. These theories should define what is the world that hosts the society, which laws rule the world, and which are the individuals that can populate it. In addition, if any intelligent global behavior can emerge from a system, there should be a place where it should be found and monitored. In a multi-agent world, this intelligence cannot reside inside agents only, but it should be somehow spread among agents and the interaction space among

them. That is, the world where agents live is not composed of agents only, but also of the abstractions supporting the interactions, as well as of the history of these interactions. Therefore, a multi-agent system should define not only the world where agents live, but also the media that permeate the space and enable agent interactions.

The agents themselves can be developed in different environments and using different languages. They usually interact by exchanging complex symbolic information and possibly have to agree on complex interaction protocols. Several proposals have been designed to develop agent-communication languages such as KQML, FIPA etc. Also several middleware systems, mostly those based on CORBA, Open Agent Architecture have been proposed to lead to a multi-tier platform enabling interoperability among software components. Although these works are important to achieve interoperability, they mainly focus on peer-to-peer communications and do not account for a more comprehensive view of the interaction as a primary component of agents' societies. Therefore, both agent communication languages and middleware systems have to somehow be extended in scope in order to include not only language and protocol specifications but also the definition of *coordination law*s, to allow for a global understanding and management of interactions.

Coordination means an interaction and management of interdependencies among different agents in MAS. A coordination model facilitates such interactions. It also deals with creation and destruction of agents, their communication activities, their distribution and mobility in space as well as the synchronization and distribution of their action. The agents being added to the system are registered so that it can be known to the system itself and the users for the system. Their activities are made available for use after registration. There are several reasons why agents in MAS need to be coordinated:

1. Agents have their own goals, knowledge and views, which may interfere with other agents' actions. They do not possess a global view of the entire agency to which they belong, as may not be feasible in any community of reasonable complexity. Also coordination is vital to prevent chaos during conflicts and to meet global constraints. For example Agents performing network management may have to respond to certain failures within seconds and others within hours. Coordinating agents' behavior is therefore essential to meet such a goal.
2. Agents in MAS possess different capabilities and expertise. Therefore, agents need to be coordinated in just the same way that different medical and para-medical personnel need to coordinate their capabilities to treat a patient.
3. Agent's actions are frequently interdependent and hence an agent may need to wait for another agent to complete its task before executing its own. Such interdependent activities need to be coordinated.
4. The most renowned coordination is based on the Contract-Net Protocol (CNP). In this approach, a decentralized market structure is assumed and agents can take on two roles: a manager and a contractor. The basic premise of this form of coordination is that if an agent cannot solve an assigned problem using local resources/expertise, it will decompose the problem into sub-problems and try to find other willing agents with the necessary resources/expertise to solve these sub-problems. Assigning the sub-problems is solved by a contracting mechanism. It consists of contract announcement by the manager agent, submission of bids by contracting agents in response to the announcement, and the evaluation of the submitted bids by the manager, which leads to awarding a sub-problem contract to the contractor(s) with the most appropriate bid(s).

5. Another important component of agent scheduling is the communication protocols among agents. In order to achieve this coordination, the agents might have to interact and exchange information; therefore they need to communicate by sending messages. KQML (Knowledge Query and Manipulation Language) is a good example of communication language.

3. Coordination Models

The design and management of a multi-agent system may take advantage from the choice and exploitation of a coordination model, that is, a high-level interaction abstraction aimed at globally ruling the behavior of the different system components.

As said earlier. a coordination model provides a formal framework in which the interaction of software agents can be expressed and their activities, distribution, synchronization etc. is maintained. It is comprised of three elements viz., coordinables, coordination media and coordination laws. The coordinables are the components of the multi agent systems, i.e. they are the agents themselves, whereas coordination media provides abstraction for interactions among coordinables and control them. The coordination laws define how the events are handled when agents interact. The laws can be defined in terms of a communication language, that is a syntax used to express and exchange data structures, and a coordination language, that is a set of interaction primitives and their semantics. Coordination models can be classified as *control-driven* or *data driven* ones.

3.1. Control-driven Coordination Model

In a control-driven coordination model agents typically interact with the external world by generating and handling events on well defined input output ports. These interactions are taken care by coordination media. The coordination laws establish how events and state changes can occur and how they should propagate. Therefore, the coordination media handle the topology of the interaction space among agents, without paying any attention to the data possibly exchanged between processes. From the viewpoint of the coordination media, the only data of interest are the communication events.

3.2. Data-driven Coordination Model

In data-driven coordination models coordinables interact with the external world by exchanging data structures through the coordination media. These data structures basically store data, which is shared among different agents. The coordination laws establish how data structures should be represented and how they should be stored and extracted from the data spaces. Unlike control-driven coordination models, the coordination media has no perception of state changes of the coordinables and does not provide for any virtual connection among coordinables. Different application contexts exhibit different needs with respect to coordination, and the choice of a coordination model is likely to have a great impact in the design of multi-component applications. In general, control-driven coordination models suits better those systems made up of a well-defined number of entities in which the flow of control and the dependencies between the components have to be regulated, and in which the data exchanged in not so important. These include, for example, computational intensive parallel applications, distributed management systems, definition of complex software architectures. Instead, data-driven model, seems to better

suit open applications, where a number of possibly a-priori unknown and autonomous entities have to cooperate. In this case, focusing on dependencies between the components, as a control-driven model would do, would somehow clash with the autonomy of the components and the dynamics of the open environment. Focusing on data preserve autonomy and dynamics of autonomous components, which are usually designed to acquire information rather than control, as in the case of software agents.

4. Conclusion

Multi-agent concepts and methodologies are finding increasing applications in controlling complex, unpredictable systems in real time. Agent-like approaches routinely accept uncertainty and distribution, leading to control schemes where decision-making and responsibility are distributed much more widely than in conventional engineering practice. The area of intelligent manufacturing systems aims to use multi-agent systems both as a modeling tool and a control software system. The challenge for the researchers is to identify and focus on the general issues of coordination and control, which are important for such applications, and to use the applications to stimulate and test advances in research on:

- How and where to represent information about the systems and the agent groups that operate on them
  - Mechanisms for delivery of adequate responses to time-critical demands
  - Coordination methods that are computationally economical enough for real-time use
  - Decentralized management of limited common resources
- Adaptation of results from machine learning to collective learning in multi-agent systems
  - New schemes of decentralized control to take account of the “real time” dimension than in conventional engineering practice

References

1. Marilyn, Deegan, & Simon, Tanner. (2002). *Digital futures: strategies for the information age*. London: Library Association.
2. David V. Pynadath and Milind Tambe. (2002). *Team Coordination among Distributed Agents:Analyzing Key Teamwork Theories and Models.* Chicago,American Association for Artificial Intelligence.
3. Michael Wooldridge. (1999). *Multi-Agent systems: A modern approach to distributed artificial intelligence.* Massachusetts,MIT Press.
4. Rosaria Conte. (2002). Agent-based modeling for understanding social intelligence. *PNAS*, May 14.
5. T. G. Dietterich. (1998) Machine-learning research:Four current directions. *The AI Magazin*e, 18(4): 97 Œ136).
6. D. Frey, T. Stockheim, P.O. Woelk and R. Zimmermann. (2003). Integrated multi-agent-based supply chain management. *In:* Proceedings of the 12th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises *(WETICE-2003*) *IEEE,* p.24.
7. Y. Guo, J. Muller, and B. Bauer. (2004). A multi-agent approach for logistics performance prediction using historical and context information. *In: Proceedings of The Third International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS'04*), New York, IEEE Press.

Library the Gateway

Preservation of Digital Information

Joseph JVM., Soundararajan E., Jayakumar C. and Somasekharan M.

Abstract

The issue of digital preservation is of utmost importance in order to maintain our mission of preserving a record of our culture. The exponential growth in the creation and dissemination of digital objects by authors, corporate, academia, governments, and libraries has emphasized the speed and ease of short-term dissemination with little emphasis for the long-term preservation of digital information. Understanding the issues and limitations that we face with respect to the preservation of digital information will enable us to develop a plan for preservation action in our digitization efforts. Digital information is inherently more fragile than traditional technologies such as paper or microfilm. The fragility of digital storage media, combined with a high degree of technology dependence, considerably shortens the "grace period" during which preservation decisions can be deferred.

Keyword: Document Management; Data encapsulation; Metadata; Extensible Markup Language (XML); long-term data storage; data migration; system emulation.

1. Introduction

Digital technology has numerous advantages and may help relieve the traditional conflict between preservation and access. Separating usage from the original, digital technology affords multiple, simultaneous uses from a single original in ways that are simply not possible for materials stored in any other form. "Digital preservation refers to the various methods of keeping digital materials alive into the future," according to a report from the Council on Library and Information Resources [1]. Digital preservation is defined as 'the means of taking steps to ensure the longevity of electronic documents. It applies to documents that are both "born digital" and stored on-line or to the products of analog-to-digital conversion (Bullock, 1999) [2]. Since technology is continuously evolving, this technological component is constantly transforming on both software and hardware levels. Because of these technological advances, the time frame in which we consider archiving becomes much shorter. Complex digital records are increasingly difficult to deal with because there are few universal standards for long-term preservation and access.

2. Need for Preservation/ conversion format

Rapid technological changes that require frequent upgrades of hardware and software. Migration and conversion of records in different format may also be needed to protect the information in records not yet eligible for disposal.
"Preservation formats will not last forever. They certainly will evolve as digital objects evolve. Text documents which are today in most of the cases reducible to two-dimensional images will in future no more be convertible in that way without unacceptable loss of information. The choice of preservation formats has therefore a lot to do with medium and long-term IT-planning. On the other hand it makes no sense to use a specific file format

Scientific Information Resource Division, Indira Gandhi Centre for Atomic Research Kalpakkam ; Email: joe@igcar.ernet.in

for current business if one does not know how it can be appropriately preserved, i.e. converted in a stable preservation format. Preservation formats therefore need to be maintained, format developments need to be carefully watched in order to keep always a migration path open to new formats when old formats will become obsolete."[5]
There are Maintenance Programs designed for organizations that want to stay on the leading edge of technology and ensure technological consistency throughout the organization and entitled to receive all upgrades of that software that occur during the Maintenance Program period.
According to Rich Lysakowski and Zahava Leibowitz "Adobe Systems (San Jose, CA) is an example of a company with a truly unique philosophy and commitment regarding long-term product compatibility. Adobe has a broad and public commitment to provide 100% backward compatibility of the PDF format *for at least the next 25 years*. No other software vendor has such a forward-looking philosophy and commitment to its customers. An important business reason for this commitment is that Adobe is beholden to the publishing industry, which has standardized on Adobe's PostScript language and many of Adobe's other multimedia publishing products. PDF is a new and improved but compatible version of PostScript, that also supports high quality publishing on the World Wide Web. The publishing industry simply cannot afford to retool every few years. Publishers must be able to access, view, and print their archived documents and books on demand for many decades to come".

Technology obsolescence, different data representation formats, limited storage media longevity, special software and hardware required, are some factors that may reduce the future readability of digital information.

3. Digital Preservation

Digital archives are distinct from digital libraries in the sense that digital libraries are repositories that collect and provide access to digital information, but may or may not provide for the long-term storage and access of that information. The glossary of the National Preservation Office defines preservation, as it relates to print publications, as "all managerial and financial considerations including accommodation and storage provisions, staffing levels, policies, techniques and methods involved in preserving library and archive materials and the information contained therein" [3]. The ability to continue to use digital information is threatened by the limited life of the media on which it is distributed. The National Media Laboratory estimate that the life expectancy of magnetic tape is between two and thirty years, incase of optical media between five and one hundred years[4]. Incase of electronic publication, the access of data is determined by the software and the hardware used, they become obsolete over a period of 5years approximately. While preserving the incoming content from various sources, to decide to keep the content format as original or to covert to current format is questionable. Due to legal responsibilities the content is stored in both formats

3.1 Document Management

Document Management is the process of managing documents throughout their lifecycle. This includes the processes of retrieving, sharing, tracking, revising and distributing

documents. It increases the productivity of the people and reduces the stress. All the documents of an organization have to be retained for various reasons; this includes varieties of formats (text, images, video and audio). While we try to keep the information on the site as accurate as possible, we disclaim any warranty concerning its accuracy, timeliness, and completeness, and any other warranty. Archiving of Digital Information focused on materials already in digital form and recognized the need to protect against both media deterioration and technological obsolescence.

3.2 Migration

Migration is a set of organized tasks designed to achieve the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation. The purpose of migration is to preserve the integrity of digital objects and to retain the ability for clients to retrieve, display, and otherwise use them in the face of constantly changing technology. Many number of migration tools that migrates data, applications and settings from one Source system to another Destination system. The process of migration is done by adopting simple methodology of installation, in less than the expected duration. In reality, not all the applications can be migrated to the destination system due to the operating system constraints that varies from 16.32 and 64 bit architectures.

3.3 Emulation

Emulation refers to the process of mimicking, in software, a piece of hardware or software so that other processes think the original equipment/function is still available in its original form. Emulation is essentially a way of preserving the functionality of and access to digital information, which might otherwise be lost due to technological obsolescence. One of the benefits of the emulation strategy compared with migration is that the original data need not be altered in any way. Once the data is archived with appropriate metadata and software, no other action is required apart from media refreshing until access is desired. One emulator can also be used as a solution for several data objects requiring the same operating environment. This should help maintain the integrity and "look and feel" of the material. Emulator can create and use multiple virtual machines (VM) running a variety of Operating Systems (OS). Emulator refers to real OS as the Host and the emulated OS as the Guest. Some emulators do not emulate an entirely different Processor/System. Only Operating Systems that run on Intel x86 compatible hardware will be useable with Emulator. This includes DOS, Windows and x86 versions of Linux, Unix and others, but few cannot emulate an Amiga or Mac for instance.

3.4 Conversion to standard formats

The use of standards is one strategy, which may be used to assist in preserving the integrity of and access to digital information. Use of standards can also help ensure best practice in the management of digital information. Resources, which are encoded using open standards, have a greater chance of remaining accessible after an extended period than resources encoded with proprietary standards. Preservation metadata is intended to store technical details on the format, structure and use of the digital content, the history of all actions performed on the resource including changes and decisions, the authenticity information such as technical features or custody history, and the responsibilities and rights information applicable to preservation actions.

Data conversion is prevalent due to widespread increase of data exchange among organizations due to the increase of participants of network communications and different usage of databases. Data conversion is generally necessary for data exchange because not all systems (formats) are compatible with other system data formats. Each system has its own unique format for organizing and storing digital data. Data conversion transforms the data into a format that is compatible with the target system through the use of import and export. To import a file the user must convert the data structure from a media source such as a tape cassette or disk to its own internal file structure. To export a data file the file must be transformed from its own internal file structure to that of a designated file structure and copied to a media source. Various users utilize data for different applications and depending upon their interest data representation may be viewed differently.
Using different software and hardware platform can alter the data format --the way data is organized and stored. Thus, making data incompatible when exchanging data between different systems. Standardized formats are somewhat different than the neutral file structure. Metadata must accompany a data set to determine the validity of a data set. Metadata or "data about data" provide documentation about the data content including data sources, production history and availability. Metadata is helpful for determining not only the validity of a data set, but also it usefulness and liability. One set of standards may not apply to some users and their applications. It is impossible to make all users agree on data classifications and feature definitions, interface tools for access or media sources for data transfer/communication.
The separate information description for the digital object structure, collection organization, and presentation interface is required. This will allow the:

- Separation of data archiving from data accessioning
- Separation of management of the collection organization from storage of digital objects
- Separation of presentation and querying of the collection from collection management

The information model used to describe the context of each component should have structured representation through use of Metadata.

3.4.1 Extensible Markup Language (XML) – Metadata Descriptive Standard

The main reason for XML's popularity is that it provides an underlying technology that gives "portability" of information across platforms, applications, and organizations. XML is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.

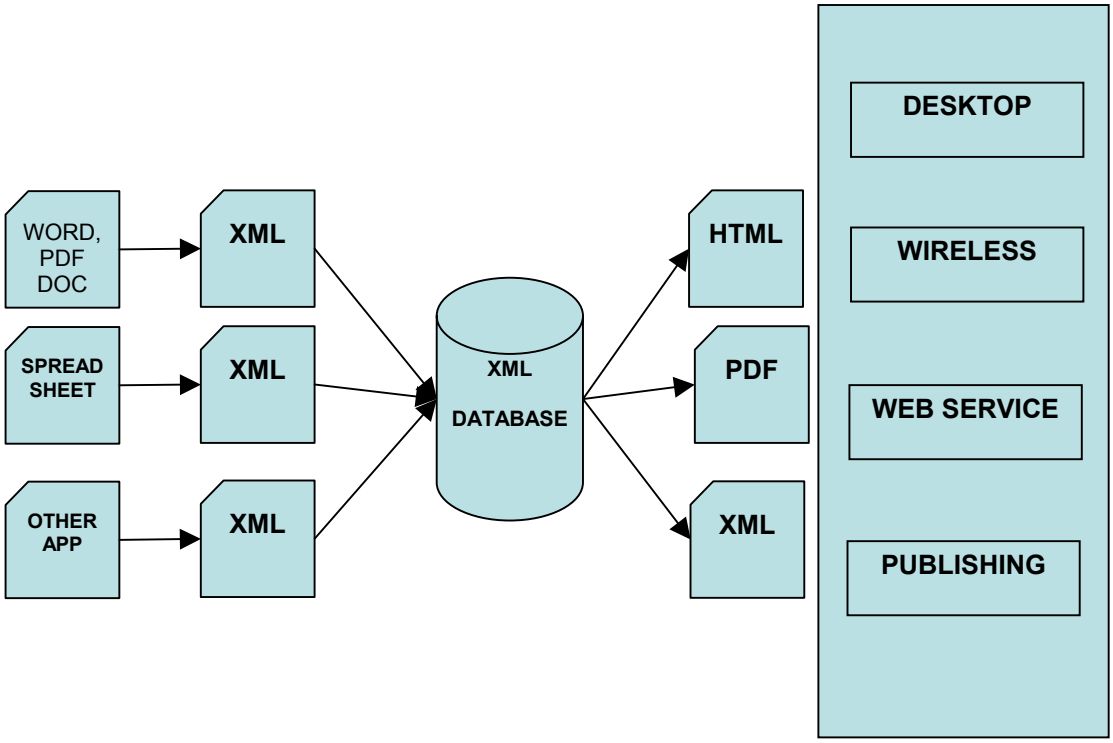


Figure 1: XML storage and publishing

XML benefits for Information Preservation

- XML is hardware and software independent. Documents created in XML-format can be retrieved long after the software and hardware they were created in has become obsolete.
- A Document Type Definition (DTD), which requires an author to conform to predetermined structures, can prevent many human variations, and this prevents time spent on editing the document after it is written.
- Since XML is format-independent, it can generate documents suitable for several output channels such as printed documents, field service updates, CD-ROM distribution, online delivery (World Wide Web), on-demand printing, help files, and machine-readable data.

XML is a versatile technology, which taps the potential of the World Wide Web and other technologies for disseminating information quickly, accurately, and independently of specific hardware platforms or software applications.

3.4.2 Issues related to XML

Data storing and retrieval in native XML format is difficult with respect to relational database management system (RDBMS). XML to relation or nested tables will ensure the consistent schema and will enable to annotate missing and incorrect links. Depending on the database vendor, the database might itself already internally support nested tables or some similar concept that can be used for this purpose. Native XML databases have a number of features that are useful for working with document-centric XML. The most important are the XML data model, which is flexible enough to model documents, XML-aware full-text searches, and structured query languages like XQuery. Successful XML database implementations are as follows: GEMt is based on Tamino; UltraXML is based on X-Hive/DB; and Syncato is based on Berkeley DB XML.

4. Preservation of Technology Investments

Investments in hardware, software, network, and digital content have limited effectiveness if not accompanied by strategic investments in staff. Investments in infrastructure improve users' requirements and provide capacity for future service improvements. To provide a cost-effective solution, it is important that the systems and components in place today can work or be upgraded for the future. The following standards may be important to establish an appropriate network infrastructure.

- High-speed (FDDI or switched) backbone
- Router-based campus and wide-area networks
- Twisted-Pair Ethernet (100baseT) building wiring
- 802.11g compatibility
- TCP/IP protocols
- SNMP network management

In order to promote stability across the network and security of the data, the organization should move towards a centralized management of both servers and network components. Centralized technical support team can provide troubleshooting, backups, administration and updates for all servers on the network. Network management has to be centralized so monitoring and administration can be handled through the technical support team.

5. Conclusion

Organizations/Publishers or individuals who did not previously consider themselves to be archivists are now being drawn into the role, either because of the infrastructure and intellectual property issues involved or because user expectations demand it. Digital materials include texts, databases, stills and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained. Long-term preservation of digital heritage begins with the design of reliable systems and procedures, which will produce authentic and stable digital objects. “Born digital” materials should clearly be given priority. Organizational Infrastructure should include the policies, procedures, practices and people to address digital preservation requirements. Technological Infrastructure should focus on requisite equipment, software, hardware, a secure environment, and skills to establish and maintain the digital preservation program.

References

1. Donald J. Waters. *Digital Preservation?* CLIR Issues (November/December 1998).

2. Bullock, A. (1999). *Preservation of digital information: Issues and current status.*

3. Foot, M *A. Preservation Policy for Digital Material: A Librarian’s Point of View.* JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib) 27th and 28th November 1995 at the University of Warwick. British Library R&D Report, 6238. 1996.

4. Keefer. A *Preservation of Electronic Publications* Presentation to the SLA Mediterranean Conference at Barcelona, 26-27 February 1999.

5. URL :< http://www.erpanet.org/events/2004/vienna/Vienna_Report.pdf>

6. V.K.J. Jeevan. *Digital library development: identifying sources of content for developing countries with special reference to India.* The International Information & Library Review Volume 36, Issue 3 , September 2004, Pages 185-197.

7. Rich Lysakowski and Zahava Leibowitz. *Looming Information Age Crisis Expected To Cause Trillion-Dollar Losses Over Next 20 Years. CBSNA 2000.*

8. E.Soundararajan, JVM.Joseph, C.Jayakumar and S.Venkadesan. *Digital Asset Management in a Networked Environment*. LIST 2003, Feb 21-22 Held at BHC Trichy.

Digital Libraries and Changing Role of Information Managers

Sanjay Gupta[1], Birmohan Singh[2] and Anukampa Negi[3]

Abstract

The information technology explosion and its applications in every aspect of life have changed the entire scenario of the present world. The IT revolution and information explosion has led to the emergence of electronic information era. These days, the users are not satisfied with the printed available material, they require that printed information be supplemented with more dynamic multimedia documents. Thus digital libraries are becoming an important element in the era of information technology and it seems that the digital library concept should be considered a dynamic and essential component of an organization/institute.

To keep pace in the rapidly changing environment, infrastructure and service facilities should be made available according to the needs so as to compete and survive in the era of competitiveness. Major changes in the information environment have transformed the role of librarians into those of information managers. An attempt has been made in the paper to give an outlines of Digital Library and changing role of information managers.

1. Introduction

In ancient days, the libraries were considered as storehouse of books and other reading material whereas the librarians were considered merely as the custodian of these documents. With the passage of time and rapid advancements in electronic information technology, the new means of communication have altered the way in which information is handled, stored and exchanged across the world. These advancements have transformed libraries from a mere static storehouse of documents to a dynamic powerhouse of information, serving all professional and non professionals in utilizing useful and need based collection of documents. The medium of information storage has changed from clay tablets, palm leaves, papyrus to papers and now to electronic and optical media. This has brought a revolutionary change in the way libraries adapt and function, shifting from print to electronic dissemination centre. All these desired the library professionals to switch over from traditional library system to digital library concept.

The Digital Library in a broad sense is a computerized system that allows users to obtain a coherent means of access to an organized, electronically stored repository of information and data. The term Digital Library may be used to describe any of the following

- Collection of electronic journals and books
- On-line educational portal
- Repository of multimedia files
- Archives of information created from local knowledge
- Electronic version of libraries
- The entire Internet

[1] *Librarian, Sant Longowal Institute of Engg. & Tech., Longowal*
[2] *Sr. Lect., Deptt. Computer Sc. & Engg, S.L.I.E.T., Longowal*
[3] *B.Lib.I.Sc. Student*

It is surprising, yet true, that there is no single universally accepted definition of the term **'Digital Library'**. Different persons have defined the term Digital Library differently as under:

According to E.A.Fox, The Digital Library may be defined as "New way of carrying out the functions of libraries encompassing new types of information resources, new approaches to acquisition, new methods of storage and preservation, new approaches to classification and cataloguing, intensive use of electronic systems and networks and dramatic shifts in intellectual, organizational and electronic practices".

Ray R Larson defined "Digital Library is a global virtual library- the library of thousands of networked electronic libraries"

Based on the view of different persons, a Digital Library may be referred to as combination of library system with computer network technologies or computerized network system where all the library information is stored in electronic format, which can be accessed and transmitted through networks enabling retrieval of required information by a large number of users. Users may access to desired information using a computer terminal at their pace of work.

2. Objectives

The primary objectives of Digital Library include:

- to collect, store, organize and access information in digital form.
- to meet the requirements of patrons by providing better services.
- to provide personalized and retrospective services in an efficient way.
- to have large digitized database accessible to multiple users at the same time.
- to save time of library staff by avoiding routine jobs.
- to provide a coherent view of all information in any format.
- to serve widely dispersed communities throughout the network.
- To minimize massive storage and space problems of large libraries.
- To reduce cost involved in various library activities.

Digital library is a combination of traditional and media collections, so they encompass both paper and electronic materials. The three main features of digital library are the storage of information in digital form, usage of communication networks, and copying by either downloading or on-line/ off line printing from a master file.

Digital Information Users

The users of digitized information may be broadly divided into four groups as given below:

(i) Those who have started using latest technology and digitized information.
(ii) Those who have been using these technologies and digitized information and are expanding it rapidly.
(iii) Those who have the fear of using new technologies for information.

(iv) Those who are intermixed between the above three groups but have no training to use the technologies for accessing global information.

Components of Digital Library:

The basic components that are required for digital library are:

- Servers for database storage.
- Adequate number of PCs connected in a LAN.
- Local databases in machine readable form, CD-ROMs
- RDBMS that supports variety of digital formats.
- Search engines to index and provide access to resources.
- Electronic document management functions that will aid in overall management of digital resources.
- Well trained manpower.

Advantages of Digital Library

- Universally accessible.
- Easy access to electronic resources.
- Optimizing use of IT environment.
- Knowledge content itself will be ubiquitous and inexpensive.
- Ability to deal with large datasets.
- Support wider range of materials.
- Access to latest information.
- No storage problems.
- Faster information retrieval.
- Increase in end users.

Barriers/ Hurdles/ Problems

- The cost involved in the creation and maintenance of digital library environment is high.
- Attitude of library professionals.
- IPR is not being given due importance.
- Preservation of electronic information.
- Pricing in the digital environment is going to be complex..
- Lack of indigenous, efficient and effective library software.
- Lack of information policy and information culture.

3. Role of Information Managers

Library plays an important role in supplementing class room teaching and it is the librarians who disseminate the precise information contained in the documents to its users in an efficient manner. With the advancements in electronic technology, it becomes very difficult for an individual to get the required information without wasting their time. It's the information managers, who have to make the required information available to the individual without wasting his/her time. The information professionals have to keep pace with the latest developments that are taking place due to advancement of technology. It is not possible that the existing information professional to go to library schools however, training can be arranged for them to have an bird's eye view about the developments that are taking place. The changes taking place in the corporate environment are challenging the librarians to reorient his/her skill to provide information solutions, not just information or information sources. Information professionals have to take into consideration the following aspects before collecting information for the end product, which they have to offer:

- It must be according to needs of its users.
- It must be equipped with powerful, easy to use, intelligent search engines.
- It must have attractive user interface.
- It must be reasonably priced.
- It must allow access from, and delivery to the users' workstation.

In the present scenario, the information professionals have to keep in mind the following points to keep pace with the modern trends, if they want to survive:

- Continuous updating
- Developing strong professional teams in their respective libraries.
- Arranging continuous education and training programs for users and staff.
- Accept the challenges being imposed due to advancement of technology.
- Arrange lectures / training programs from time to time.

Information managers are only competent enough to provide efficient and intelligent access to the world's information sources to its novice users. The knowledge and experience of libraries can be helpful in directing users to get an easy access to relevant information on the digital library network.

4. Conclusion

The digital library concept is growing at a fast pace. Emerging technology of digital libraries is an offshoot of information revolution which can drastically improve the efficiency and effectiveness of management of physical and financial resources of libraries. The information managers/ librarians have to equip themselves with the capabilities to link with global trends for the ultimate benefit of information seekers. Arrangements for managing and supporting information technology will take different shapes depending upon the size of its parent institution, the institution's existing computing and communication resources and the capabilities of the library's management and staff. The professionals can join the revolutionized technological race without sacrificing their conception of traditional libraries. The information providers have to keep themselves abreast with the latest trends if they desire to survive in the present world.

References

1. Fox, E.A.et.al. Digital Libraries: Introduction. Communication of ACM, 38(4)1995.
2. Harter, Stephen P. Scholarly Communication and the Digital Library: Problems and Issues. Proceedings of International Conference on Digital Libraries and Information Services for the 21st Century. Sept.10-13,1996, Seoul, Korea.
3. Gopinath, M.A. Curriculum for Digital Information Systems: A Learning Package for Library and Information Professionals. Library Science with a slant to Documentation and Information Studies.**35 (2),**1998. 73-78
4. Malwad, N M. Digital Libraries: Dynamic Storehouse of Digital Information, New Delhi, New Age International. P327-29
5. Nasib S. Gill and K.C. Dabas. Digital library : Its Emergence and Challenges. CALIBER-98
6. Hulser, R.P. Digital Library : Concept preservation in a Digital World. DESIDOC Bulletin Information Technology, **17(6),** 1997, 7-14
7. Kent, A. Encyclopedia of Library and Information Science. New York, Marcel Dekker Inc, **54**, 1994, 143-154

Digital Library: The Change Management

Aparajita Suman

Abstract

Change management is not a singular concept, rather it includes a set of best practices and experiences, which are used to handle both the internal as well as external changes. Change management includes effective management of new methods and systems in an ongoing organization It is both an area of professional practice as well as a body of knowledge. Change from an existing setup to a new environment has its own set of inherent problems and the problems become multifold when applied in a service institution as library. Libraries have been pioneers in adopting any new technology, the same holds true for information and technology also. The role of libraries has gradually changed from the traditional storehouse of information to access providers. There has been a paradigm shift in the ways libraries used to be managed. The problems associated with the transition in the case of libraries have both a content and a process dimension. This paper aims at providing a broad overview of the concept of change management and its relevance for managing the transit of traditional paper-library to digital-library.

Keywords: Change Management, Traditional Library, Digital library, Transition

1. Introduction

Karen Kaiser Clark states that "Life is change. Growth is optional. Choose wisely (1). Management when defined simply is nothing but making wise choices at the correct time. Change management refers to effective management of new methods and systems in an ongoing organization. It is not a singular concept rather it includes a set of best practices and experiences, which are used to handle both internal as well as external changes. Change management is both an area of professional practice as well as a body of knowledge. Change management includes the changes that lie within and are controlled by the organization and those that come to terms with the changes occurring in the surrounding environment i.e. the events originating outside the organization and the response to them.

The embedded complexity in the case of library and information centers makes the process of change management a task, a process, and an area of professional practice at the same time.

2. Change Management: The Concept

The basic definition of Change management has following 3 important aspects (2,3):

- The task of ongoing change.
- An area of professional practice.
- A body of knowledge.

Information Officer,Virtual Information Centre, ICICI Knowledge Park, Genome Valley, Turkapally,Shameerpet Mandal, RR District, Hyderabad 500078.
E-mail: *s_aparajitha@yahoo.co.in*, *aprajita_drtc@rediffmail.com*

'Managing change' refers to the making of changes in a planned and managed or systematic fashion where the aim is to implement new methods and systems in an ongoing organization in a more effective manner. Now, if one thinks of internal changes induced by outside developments, then the most familiar instance of this kind of change is the change or version control aspect of information system development projects. Here, the events originating outside the organization trigger these internal changes. Literally the two types of changes can be distinguished as a knee-jerk or reactive response and an anticipative or proactive response.

Professional expertise is required to manage the changes; whether reactive or anticipative. The process of change needs to be treated separately from the specifics of the situation.

The content or subject matter of change management consists chiefly of the models, methods and techniques, tools, skills and other forms of knowledge that go into making up any practice. The subject matter of change management is drawn from psychology, sociology, business administration, economics, industrial engineering, systems engineering and the study of human and organizational behavior.

3. Types of Changes

Change can be better managed if we are able to categorize them. Changes can be broadly categorized as:

- Provoked by pressure or necessity
- Induced by gentle persuasion rather than force
- Enforced change
- Motivation by example and evidences
- Designed according to individual needs and requirements

4. Problems and Prospects

As discussed earlier, the problems faced during bringing about changes in any organization have both content and a process dimension. Introducing a digital library setup has different dimensions depending on the host organization. For instance, the user group and its requirements of the digital library at a health university will be different from the users of a defense laboratory library. In this situation, it is very difficult to suggest a universal change management strategy, it needs to be defined and designed depending on the need of the organization. The differences become more prominent in the case of organizations with international user group with varied subject interests because the values differ, the cultures differ and at the very basic level even the problems differ.

But, as the overall processes of change and change management remain pretty much the same, so change management has evolved to become a discipline from a set of good professional practices.

5. Processes

The process of change management can be studied as an amalgamation of following sub-processes [4]:

- The Change Process as "Unfreezing, Changing and Refreezing"
- The Change Process as Problem Solving and Problem Finding

- The Change Problem
- Change as a "How" Problem
- Change as a "What" Problem
- Change as a "Why" Problem

It becomes easier when problems are formulated in terms of "what, Why and How" questions. The formulation depends on where in the organization the person posing the question or formulating the problem is situated, and where the organization is situated in its own life cycle.

6. Libraries and Information Centers

Change from an existing setup to a new environment has its own set of inherent problems and the problems become multifold when applies in a service institution as library. Libraries have been pioneers in adopting any new technology, the same holds true for information and technology also. The role of libraries has gradually changed from the traditional storehouse of information to access providers. There has been a paradigm shift in the ways libraries used to be managed.

The problems associated with the transition in the case of libraries have both content and a process dimension. The reason is that not only the library operation has got automated but also at the same time there have been drastic changes in the way information content used to be presented and organized. The tools and techniques that were suitable for traditional documents don't hold good for born digital documents, so, a whole new approach needs to be developed to handle the situation.

The process of change management in a library setup raises basic issues such as 'what and how" questions to cluster in core and buffer units. And the 'why' question is typically the responsibility of top management. The skills and strategies involved in managing change in libraries and information centers are multifaceted.

6.1 Skills

1. **Political Skills**

   These skills are required to motivate people and convince the management at the same time for the maximum benefit of the organization and users

2. **Analytical Skills**

   The ability to analyze the situation judiciously and to act wisely is the primary requirements of a good manager. Further, if the property to be managed is change, then the analytical skills are of greatest significance.

3. **People Skills**

   People skills are nothing but the ability to manage people in the groups, one that are part of the change and those that are bringing about the change.

4. **System Skills**

   These include a set of skill required for designing of a new system or bringing out balanced renovations in existing systems. One should be able to have a balanced

outlook towards all the components of the system ranging from input to the final output

5. Business Skills

Business skills have acquired greater significance because of the increasing emphasis on the evolution of self-sustaining profit generating information centers.

6.2. Strategies

Change strategy can't be defined in a singular fashion rather it has to be adopted as a mixture of strategy sets keeping an eye on the ultimate grand vision (5). Following factors determine the choice of the strategies:

- **Degree of Resistance** - Strong resistance argues for a coupling of power-coercive and environmental-adaptive strategies. Weak resistance or concurrence argues for a combination of Empirical-Rational and normative-reductive strategies.
- **Target Population -** Large populations argue for a mixture of strategy sets to be adopted
- **The Stakes** - High stakes again demand a mixture of strategies.
- **The Time Frame -** Short time frames require a power-coercive strategy.
- **Expertise** - Availability of desired expertise gives the management ample of confidence to a mixture of strategies
- **Dependency** - If the organization is dependent on its people, management's ability to command or demand is limited. Conversely, if people are dependent upon the organization, their ability to oppose or resist is limited.

Dennis Sparks, Executive Director of the National Staff Development Council, offers these 13 tips for managing the complex and difficult change process (Sparks, 1993) [6], these tips can be applied suitably in the case of managing the change from a paper-based library to a digital-library.

1. Educate the leaders of change.
2. Use a "systems" approach to ensure that all aspects of the organization are considered when planning and implementing change.
3. Use a team approach that involves many stakeholders in the change process.
4. Share power with everybody to encourage the implementation of the change efforts.
5. Make plans, develop plans, but know that they will have to be adapted to change, as needs change.
6. Realize that there is a tension between establishing readiness for change and the need to get people implementing new approaches quickly. While getting people intellectually ready for change is something to be considered, it should not take so much time and effort that people lose interest and motivation.
7. Provide considerable amounts of training and staff development for those involved.
8. Choose innovative practices, picking approaches that have been used or researched can help the implementation of those approaches.

9. Change happens only through people, the emotional effects of change on educators need to be considered and understood by all involved in the change process.
10. Be prepared for "implementation dip." Experts note that things often get worse temporarily before improvement begins to appear.
11. Help others in developing an "intellectual understanding" of the new practices. While the outcomes are important to assess, people also need to understand the underlying meanings and functions of the practices.
12. Seek out "paradigm shifters" and "idea champions" who are interested in making substantial changes in practice.
13. Take the long view; realize that change takes time and should not be forced to occur too quickly

7. Good Practices

Change can be managed better if the following set of measures are adopted and implemented.

- Planning a long-term, sound, strategic vision. Here, the focus should be on detail for establishing and measuring delivery of immediate actions, not medium-to-long-term plans.
- Establishing communicating methods, this enables immediate review and decision-making.
- Empowering decision making at a local operating level.
- Encouraging, enabling and developing capable people to be active in other areas of the organization via 'virtual teams' and 'matrix management'.
- Adjusting recruitment, training and development to accelerate the development of people who contribute positively to a culture of empowered dynamism.

8. Conclusion

The concept of change management can be understood and its relevance has been proved for managing the "*transit of traditional paper-library to digital-library".* The underlying principle is to involve and agree support from people within system (system = environment, processes, culture, relationships, behaviors, etc., whether personal or organizational). The key lies in communication, enabling and facilitating involvement from people, as early and openly and as fully as is possible.

References

1. *The Planning of Change (2nd Edition).* Warren G. Bennis, Kenneth D. Benne, and Robert Chin (Eds.). Holt, Rinehart and Winston, New York: 1969.
2. *Human Problem Solving.* Allen Newell and Herbert A. Simon. Prentice-Hall, Englewood Cliffs: 1972.
3. *Organizations in Action.* James D. Thompson. McGraw-Hill, New York: 1967.
4. http://home.att.net/~nickols/change.htm, accessed on 25.05.05
5. http://www.businessballs.com/changemanagement.htm, accessed on 01.05.05
6. http://www.themanager.org/Knowledgebase/Management/Change.htm, accessed on 05.06.05

Digital Document Delivery: A Case Study

Deshpande A. Dighe.R. and Pattnaik.J.K.

Abstract

A web based inter library photocopy request system has been implemented at the Centre for Advanced Technology Library, Indore. The features and functions of the system are described. Based on the evaluation of the system, this paper reports to share the experiences and modifications carried out in the system accordingly.

Keywords: Photocopy request, Resource sharing, Digital document delivery**.**

1. Introduction

The Centre for Advanced Technology (CAT) library at Indore offers 'Interlibrary Loan' (ILL) service facilitating users to obtain those document/s that are not available in CAT library. To enhance the ILL facility, web-based request form has been designed to obtain the requests and forward the same for processing to fulfill the demand. This system has been implemented. A brief description of the features of the system and its functions are described followed by the report drawn on evaluating the system.

2. Ill Request Form

A request form for the end users to furnish the requisite details of the document/s has been designed based on based on ISO 10161 Interlibrary Loan Application Protocol, which defines standard set of messages & format for ILL transactions. Three separate request forms have been designed, allowing variations in the data fields for different types of publications:

1. Journal Articles
2. Conference Papers
3. Book/Monograph

Java Servlet, Java Scripts & HTML are used in developing the software. To store the requests and responses, Oracle RDBMS has been used as backend database. User can login using his/her employee identity number. For security reasons, end users cannot modify the form. A warning concerning copyright violation is reproduced on the web-form. Using administrator login, library staff can modify and send requests to the supplying library.

3. Functions

3.1. User End

The requester performs the following steps:

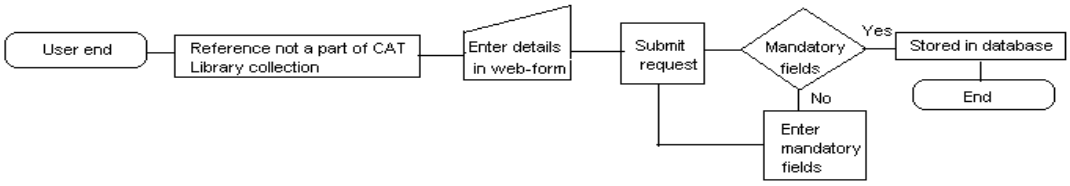
1. Access the library homepage on the Intranet.
2. Checks availability of the journal/book in the library holdings list/catalogue

Library, Centre for Advanced Technology, Indore - 452013, E-mail: artidesh@cat.ernet.in

3. Access the required photocopy form.
4. Enters his employee identity number on the form, which then automatically displays his full name & section.
5. Enters the required publication data. Certain of the fields are mandatory, & the system does not accept the request if mandatory fields are not filled.
6. Clicks on the send button to send the form to the library's ILL database.
7. Repeats the procedure for further photocopy requests.

The same has been drawn as a

Flow chart of functioning of the system at user end:

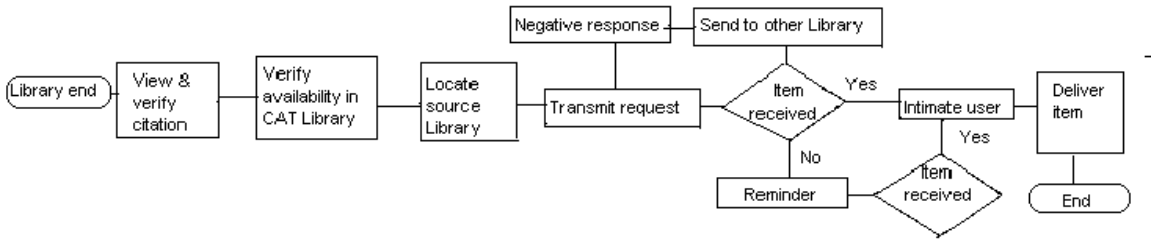


3.2. Library end

1. Library checks the ILL database for new requests.
2. It checks the bibliographical data, & makes corrections if necessary.
3. It checks the availability in the library holdings list or catalogue.
4. Checks availability in DAE libraries, or in other libraries.
5. From the addresses' database, selects the library to which the request is to be sent. Multiple requests from different users can be combined & sent to a single library.
6. If the selected library does not have an e-mail address, the requests are printed, together with the postal address, & sent by post.
7. If the responding library cannot supply the photocopy, the request is re-sent to a different library.
8. If the document is not received in time, a reminder is sent to the supplying library
9. When the photocopy is received, the system provides for the information to be sent automatically by email to the user. In case he does not have an e-mail address,
   his telephone number is displayed.

The same has been shown in Figure 2.

Figure 2 Flow chart of functioning of system at library end



3.3. Reports Generated

The design of the system provides for generation of different types of reports:

By user

1. List of photocopy requests made, giving their status. The user can view this list.
2. Intimation to the user of receipt of the material.

By Library

1. Email & printed requests sent to a particular library. Automatic intimation to user for his request processed.
2. Reminder generation for unfulfilled requests.
3. Printing of hardcopy ILL requests, if required.
4. List of requests received during a given period & yet to be processed.
5. List of requests processed, arranged by library, for which documents are awaited.
6. On receipt of document, automatic acknowledgement to resource library & collect notice to the user.

4. Evaluation

The implemented system was evaluated to identify the problems. A few of the major findings are reported here. That includes statistics regarding the requests received under this system over a period of time, the quantum of successful responses, the sources from which the ILL service was organized by (CAT) library at Indore and the mode of delivery. . The same is presented subsequently.

Table -1 Requests for the month of April 2004

Date of request	Date of request mailed	Date of article received	No. of days taken by CAT Lib. To mail	No. of days taken to fulfill the request
1	5	6	4	5
1	5	13	4	12
5	5	26	0	21
5	6	*	1	*
6	23	6-May	17	30
12	19	7-May	7	25
19	19	21	0	2
19	19	11-May	0	22
19	23	23	4	4
19	23	12-May	4	23
20	23	13-May	3	23
20	23	18-May	3	28
20	23	23	3	3
20	23	5-May	3	15
26	27	*	1	*
26	27	19-May	1	23
28	30	24-May	2	26
28	28	19-May	0	21
29	30	*	1	*
29	30	24-May	1	25
29	30	24-May	1	25
			Total 60	333

Total No. of requests - 21
Requests fulfilled - 18
Percentage of request fulfilled – 18/21 = 86%
Average days taken to forward the request to resource library - 60/21 = 2.85
- 3 days to forward the request to resource library.
Average days taken to fulfill the request – 333/21 = 15.86 = 16 days to fulfill the request.

Table-2 Response rate for April, May & June 2004

Month	% of requests fulfilled	Average days taken to forward	Average days to fulfill the requests
April	86	3	16
May	85	1	11
June	85	3	28
Total :	**256**	**7**	**55**

For all three months
Percentage of request fulfilled: 85
Average days taken to forward: 2
Average days to fulfill the request: 18

Graph – 1 Percentage of requests fulfilled for all 3 months

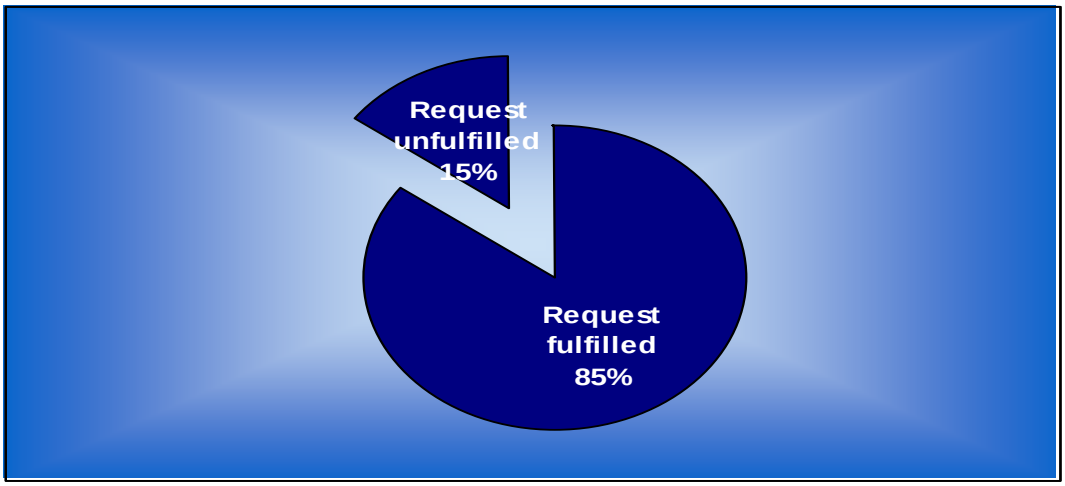
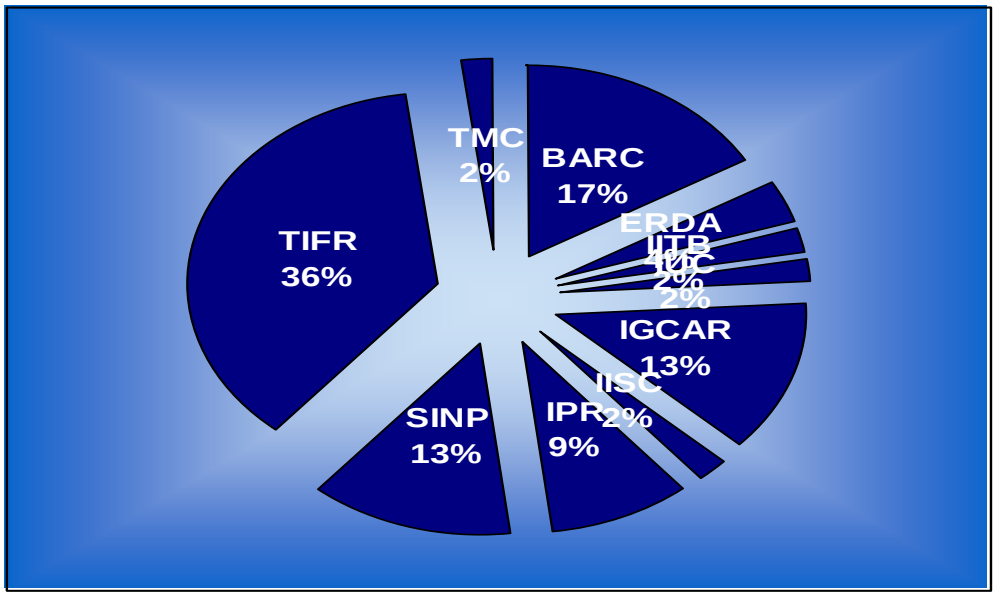


Table-3 Source Centres

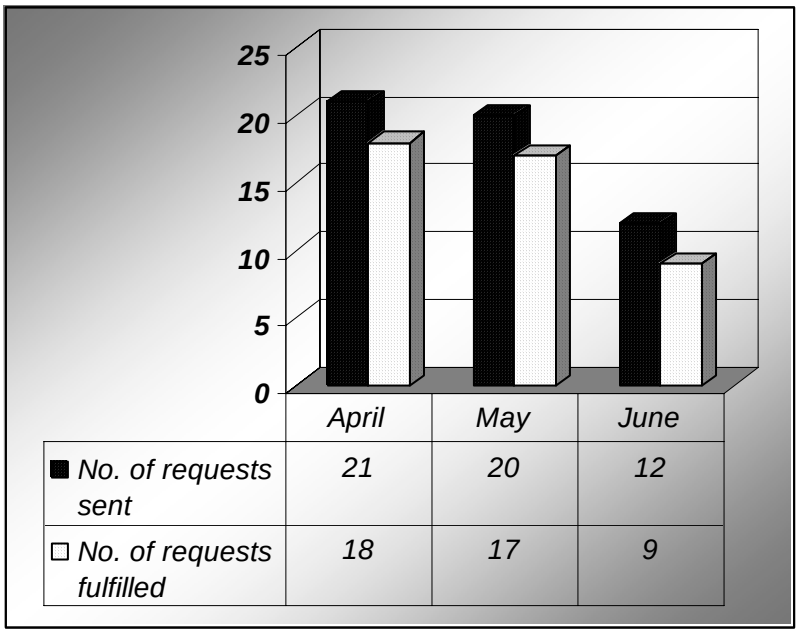
Resource Centres	No. of requests	No. of requests fulfilled	Ratio	%
BARC	9	6	0.67	67
ERDA	1	1	1	100
IITB	1	0	0	0
IUC	1	1	1	100
IGCAR	7	6	0.86	86
IISC	1	1	1	100
IPR	5	3	0.60	60
SINP	7	5	0.71	71
TIFR	20	20	1	100
TMC	1	1	1	100
Total	**53**	**44**		

Note: Ratio or percentage do not reflect the efficiency of the resource libraries, as other factors like availability, number of requests, request on payment basis, etc also affect their response.

Graph-2 Percentage of requests fulfilled by different resource libraries.



Graph – 3 Number of requests for the month of April, May & June.



5. Modifications

1. For wrong bibliographical details, option was not available earlier to communicate to the users. Now MAIL BACK or PRINT (in case a person is not having e-mail) option has been provided.
2. Other users who do not have the Employee's Identity Number were unable to submit the request, as the form is a login based. administrator has to fill the form on behalf of these users. Now it was made possible for other users like TSO (Training School Officers), JRF (Junior Research Fellows) and VS (Visiting Scientists) etc to submit the requests.

3. Search is customized. Search on journal title has been made possible through search option. It is now possible to search the requests, processed and pending.
4. New options. PREVIOUS and NEXT buttons are provided in detailed bibliographical page to switch over. DELETE option is provided to ILL administrator in REQUEST and RESPONSE form.
5. In response, same requests to different libraries are displayed together. Earlier it used to be scattered according to mail date.

Example:

Request	Author	Periodical Title	Vol No	Year	Unit	Reminder No	Mail Date
100	Morgan	BNL Technical Notes	48	1975	A3	23-	June2003
100	Morgan	BNL Technical Notes	48	1975	B0	25	June-2003
99	Tieu and Ulrich	Journal of the Optical Society of America	60	1970	C3	24	June-2003

The interlibrary response time could be considerably reduced if libraries could supply electronic copies. The urgent requests could be fulfilled by digital document delivery. For requests from the journals, which are accessible full-text on-line, the articles can be downloaded in pdf. format and sent as an e-mail attachment. For other journals, articles can be scanned, converted to pdf. and sent as an e-mail attachment. Putting on ftp server can also be considered. The receiving library could print a hard copy, & delete the electronic file to meet copyright requirement. Warning concerning copyright violation is reproduced in our Web-form (Appendix - 1). It mentions clearly that any resource library reserves right to refuse to accept a copyright order if, in its judgment fulfillment of the order would involve violation of copyright. However since supply of an electronic copy may not be acceptable to some publishers, that option has not been considered this option for the present.

6. Conclusion

Web-based ILL system has reduced the work at users as well as library end. We could avoid duplication of data entry work. Least time is taken by CAT library to forward request to other library by saving time in entering data about the document & resource libraries repeatedly, preparing letter, printing, sending etc. Due to ease in reminding, the number of requests not fulfilled is now very few. It has curtailed the manual work of telephoning & delivering the document to the users, as on receipt of the document the intimation is automatically sent. It has made the retrieving process easier with the search facility & is easy to know the status of the request. We plan to extend the system to other libraries for operational and management of their document supply. A centralized DAE ILL system can be planned on ANUNET, the DAE network.

Acknowledgement

We thank Mr.Deeptikant Pathy and Mr.Anurag Bansal of our computer centre who has developed the software for ILL system.

References

1. Mignon Adams. Rethinking Interlibrary loan for Scientist. Science & Technology Libraries, Vol. 24, no. 3/4 , 2004, pp.401-409.
2. J. K. Pattnaik, R. Dighe, and Zeenat Rustagi. Design & implementation of a web-form for interlibrary loan & photocopy request. Proceedings of the 21st Annual Convention and Conference of Society for Information Science, IIT Roorkee, 9-11 April 2003.
3. Fay Turner. Document ordering standards: the ILL Protocol & Z39.50 item order. Library Hi Tech, Vol.13, no.3, 1995, pp.25-38.
4. InterLibrary Loan Henry G. Bennett Memorial Library, Southeastern Oklahoma State University. URL :<http://www.sosu.edu/lib/ill.htm>
5. Interlibrary Loan Application Standards Maintenance Agency, National Library of Canada.URL: <http://www.nlc-bnc.ca/iso/ill>

APPENDIX -1

Journal Article Photocopy Request Form

CC No 1456

Name Arati Uday Deshpande

E-Mail artidesh @cat.ernet.in Phone No 8604

Author

Periodical Title

Article Title

Volume Issue No

Issue.Date Month Day Year

Pages Eg.(5- or 5-10)

Comments

Send Reset

[Warning Concerning Copyright Violations

The copyright law of the India (The Indian copyright Act, 1957) governs the making of photocopies or other reproductions of copyrighted material. The copyright law gives the right to libraries & archives to furnish a photocopy or other reproduction for a limited purpose or fair use. However any Library reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law]

Developing a Framework to Aid Decisions with Reference to the Use of Different Storage Devices for Information Management

Rathnakar Acharya and Vivekanand

Abstract

Managing and storing digital information is an essential requirement of all organizations today in the digital world. The question of storing information invariably leads to the next logical question of which particular storage media to use. In yesteryears, the decision of using a particular storage device was to a large extent determined by the cost factor. With growing technological advancements in the area of computing and storing, the price factor for deciding upon a particular storage device has significantly lost its prominence. In such a situation, other contextual factors such as the type of information to be stored, the level of accessibility for different types of users, the criticality of the information and the different sources of information, would become more critical. This paper will try to formulate a framework that would help decision makers in making an informed choice with reference to the use of particular storage media, given the particular set of contextual factors.

1. Introduction

Digital libraries have received much attention in research and development literature and their scope is ever increasing in the information driven world. Digital libraries are defined as electronic collections that are much richer in content and more capable in functionality than mere databases or mere information retrieval systems (Borgman, 1999). Such digital libraries are increasingly becoming accessible over the Internet. Harvard University, for example, is embarking on collaboration with Google that could harness Google's search technology to provide to both the Harvard community and the larger public a revolutionary new information location tool to find materials available in libraries. Initially, staff from Google is to collaborate with Harvard's libraries on a pilot project to digitize about 40,000 of the 15 million volumes held in the University's extensive library system (*hul.harvard.edu/publications/041213faq.html*). Google will to provide online access to the full text of those works that are in the public domain. In related agreements, Google will launch similar projects with Oxford, Stanford, the University of Michigan, and the New York Public Library. Such ventures are bound to be followed by universities, institutions and extend to commercial organizations across the world. The scope and importance of digital libraries is upwardly mobile and the trend will continue. The major advantages of digital libraries include storing resources in an easy-to-track digital format; allowing remote, rapid, and fair access to digital library collections, and providing search techniques that offer users increased flexibility and power (Wiederhold, 1995).

There have been significant advances in the technical development of digital libraries in areas such as information storage, information retrieval, and system integration, resulting in dramatic improvements in their performance. This paper deals with the issue of information storage with special reference to the decision made with respect to the usage of storage devices incorporated in designing and developing digital libraries. This paper is targeted at managers of enterprises wanting to set up information and knowledge sharing

Alliance Business Academy, , Bangalore – 560076 ; Email – rathnakar_a@yahoo.co.in

systems. It is believed that while deciding on adopting an Information System for storing and disseminating information, decision makers generally do not give much importance to the nature and kind of storage device that would be used and the decision is normally left to the vendor of the Information System solution provider. Given that there are various kinds of storage devices, each with its own set of features and price tag, a conscious decision to choose the optimum storage device could to some extent affect the operationality and the feasibility of such information systems.
The objective of this paper is to inform the decision makers about the available storage devices along with their functionality, so that they are able to take informed decisions with the respect to the use of storage devices for their information systems.

2. Need for Storage Management

The business environment today requires real time responses to change whether it is new demand by customers, change in the supply chain or unexpected competitive moves. To a large extent success in the "on demand" world depends on the ability to leverage Information Technology (IT). Because IT infrastructures are traditionally built along functional lines the applications operating systems, database and hardware platform often fail to communicate effectively. Greater dependence on information means, greater dependency on storage. Automation of routine storage management tasks can help enable IT staff to focus on providing better services to users and improved data protection to shield important information from unforeseen threats and providing greater resilience to respond to unanticipated events.
Integrated storage is essential to integrate information systems. The traditional model of storage directly attached to distributed server makes it difficult to share data among host servers because of networking bottlenecks as well as format and device incompatibilities. These weak links can make sharing data among host servers difficult. However by consolidating data into centralized pools of storage we can avoid the pitfalls of the traditional storage model. But this storage consolidation can also address another critical aspect of the on demand paradigm, the efficient use of resources. The economics of scale and simplified administration enabled by centralized storage can help reduce cost and improve responsiveness to changing demand.
Today it is an age of mission critical, storage intensive applications that with the advent of Internet and popularization of e-Commerce have forced organizations to provide 24/7 data availability to their end users. 24/7 world wide data availability ensures that users can access data at any point of time. This means that storage infrastructure must always be ready to handle transaction data. Many applications today heavily rely on storage infrastructure. Some of the storage intensive applications are:

- On-line University
- E-learning
- Data warehouse and data mining system
- Web based e-mail
- Online trading and transaction
- E-commerce transactions
- Multimedia applications

These storage intensive applications have raised unprecedented demands on the current storage infrastructure, which has been unable to keep up with the demand. If the information demand is not handled adequately, this may result in information spill over, server crash and the loss of important and confidential data.

3. Storage devices

Secondary storage is designed to store very large amount of data for extended periods of time. These secondary storage devices can have memory capacity of Gigabytes or Terabytes. The important characteristics of the secondary storage devices are:

- it is nonvolatile
- it takes much more time to retrieve data from secondary storage than it does form RAM
- it is much more cost effective than primary storage
- it can be placed on a variety of media, each with its own technology

3.1 Magnetic tape

Magnetic tape is a cheaper storage medium and can handle enormous amount of data. The downside is that it is the slowest for retrieval of data, because all the data are placed sequentially. Sequential access means that the system might have to run through the majority of the tape. Magnetic tape storage often is used for information that an organization must maintain, but uses rarely or does not need immediate access. Industries with huge number of files use magnetic tape systems. Modern version of the magnetic tape systems use cartridge and often a robotic system that selects and loads the appropriate cartridge automatically. There are also some tape systems like digital audio tapes, for smaller applications such as storing copies of all content of personal computer's secondary storage (backing up the storage).

3.2 Magnetic disks

Magnetic disks are very popular because they are built in our portable or Personal Computers Systems as a storage media. Also they allow much more rapid access to the data than does magnetic tape. In a magnetic disk data can be accessed directly. The Hard Drives in the system use stacks of rigid magnetic disks (hard disks). Hard disk provides several GB of data storage to our personal computers. Data access in hard disk is very fast, measured in milliseconds. For these reasons hard disks are extremely popular and common.

3.3 Hard drive

Organizations also use hard drive systems for backup and long term storage. A common technology is the Redundant Arrays of Inexpensive Disk (RAID). These machines use a large number of small hard disk drives. The data are distributed across all of the disk drives. The logic is that if one drive fails, the impact is greatly lessened. The access speed of RAID makes it more appropriate than magnetic tape storage if the data are to be accessed more often or require more rapid retrieval.

3.4 Magnetic Diskettes

Magnetic Diskettes (floppy disks) can also be used as one of the storage deceives. It functions similar to that of the hard disk drive. The data storage capacity of magnetic diskettes is lesser than that of the hard disks, which is in terms of few Megabytes. They are much slower than the hard drives. The big advantage of this floppy disk is that they are portable.

3.5 Optical storage devices

Optical storage devices are also a form of secondary storage device on which data are recorded by tiny holes burned on the surface of a reflective platter and are read by another laser in computer's disk drive. Compared to magnetic media, optical disk drives are slower than magnetic hard drives. On the other hand they are much susceptible to damage from contamination and are fragile. Optical disks can store much more information both on routine basis and also when combined into storage systems, like RAID optical storage system, can be used for large capacity data storage. These technologies known as optical jukebox, stores many disks and read leasers. CD ROM storage devices feature high capacity, low cost and high durability.

Comparison of secondary storage devices
The table given below lists down the advantages, disadvantages and applications of the various storage devices discussed in the previous section.

Storage Medium	Advantages	Disadvantages	Applications
Magnetic tape	• Lower cost per unit stored	• Slow data retrieval speed	• Corporate data archiving
Hard disk drive	• Higher capacity • Faster data retrieval speed	• Fragile; High cost per unit stored	• Personal computer through mainframes
RAID	• High capacity • Reduced risk of data loss • Low cost to per unit stored.	• Expensive • Semi-permanent installation.	• Corporate data storage that requires frequent, rapid access.
Floppy diskettes	• Low cost per diskette • portability	• Low capacity • Very high cost per unit stored • fragile	• Personal computers
CD ROM	• High capacity • Moderate cost per unit storage • High durability	• Slower retrieval speeds than hard disk drives; only certain types can be rewritten	• Personal computers through corporate data storage.

4. Decision influencing factors

Following are some of the important factors to be taken for better storage infrastructure:

4.1 Technical factors:

Limited scalability support: When we add more numbers of storage devices to the server, it may hamper the server performance. With the ever increasing amount of data this doesn't prove to be a reliable solution.

Excessive network traffic generated because of data by multiple clients: While performing backup and recovery operations across a distributed network and excessive amount of data traffic is generated, this can cause network shutdown in extreme cases.

Bandwidth mismatch between input/output devices and the storage devices: It is possible to increase the storage devices manyfold up to many terabytes. But the I/O bandwidth of the devices, on the other hand, has failed to keep up with the advancement in the storage devices. This may reduce the overall performance of a network.

4.2 Economic factors:

High cost of distribution networks: Owning or hiring a network to connect the number of devices spread around the globe increases the expense of the entire infrastructure.

Increased financial cost: Managing storage resources in the current infrastructure is expensive. Moreover hiring component and trained storage managers is a costly exercise.

5. Various options available:

Some of the storage devices and techniques are:

1. CD storage and Libraries
2. DVD storage and Libraries
3. Disk Storage systems
4. Disk Arrays
5. Tape Libraries and subsystems.
6. Embedded storage
7. Direct Attached Storage (DAS)
8. Storage Area Networks (SAN)
9. Network Attached Storage (NAS)

We are in the Information age, where fortunes spring from innovative ideas and the clever users of information. Business in the Information age must compete in challenging market requirements – one that is rapidly changing, complex, global, hyper- competitive and customer focused.

The pace and magnitude of change affecting organizations continue to accelerate causing increased uncertainty in company operation and strategies. Digital Library promises to make it easy for the business to perform in an effective manner.

The business environment in the information age places many demands on the Digital Libraries. In fact, the Digital Libraries may respond reactively to these requirements already in existence or proactively to an anticipated requirement in terms of an efficient data storage and data management systems.

6. Contextual Demands

The contextual demands on the Digital Libraries include the following:

6.1 Information Overload

The amount of information available from the different sources doubles every year. This information needs to be stored inside the organization in a embedded storage system. As the volume of the information increases, it becomes difficult to maintain these enormous data in one place. Hence it may be maintained in a Storage Area Network (SAN) or in distributed server.

6.2 Technological Innovations and obsolescence

The demand to produce goods and services efficiently and to come up with new products and services places lot of importance on technological breakthroughs that may give them an advantage over their competitors. Technology is playing an increased role in both manufacturing and service organizations. New and improved technology for data storage and management enables organizations to produce quality information.

6.3 Customer orientation

The Customer's sophistication and expectation increases as they become more informed about the availability of the quality of services. Customers also want customized information with high quality and at a faster rate.

6.4 Need for real time operations

Organizations in this information age no longer have the luxury of information float, which is the time between when new data is generated and its capture. It is expected today, that as soon as new data is generated it must be captured and updated in the storage devices for converting them into pieces of useful information.

6.5 Global spread of end users

The information requirement is not only for the users inside the organization. It must be available to anyone (legitimate users), any where, and any time. High performance telecommunication technologies can provide the accessibility without much time lag.

6.6 Changing life style

Due to the change in working time and geographical time difference, the users would like to access the data from the server at any point of time. Data must be available 24/7 and 365 days a year.

7. Organizational response

Organizations respond in many ways to the contextual demands discussed above. The major responses are stated below:

7.1 Collaborations

There are several types of alliances; sharing resources, establishing permanent relationships and creating joint research efforts, that helps in dealing with some of the above mentioned demands.

7.2 Strategic systems

Organizations seek to implement systems that will significantly impact the organization's operations, success, or survival. Such strategic systems provide organizations with strategic advantages in meeting organizational objectives.

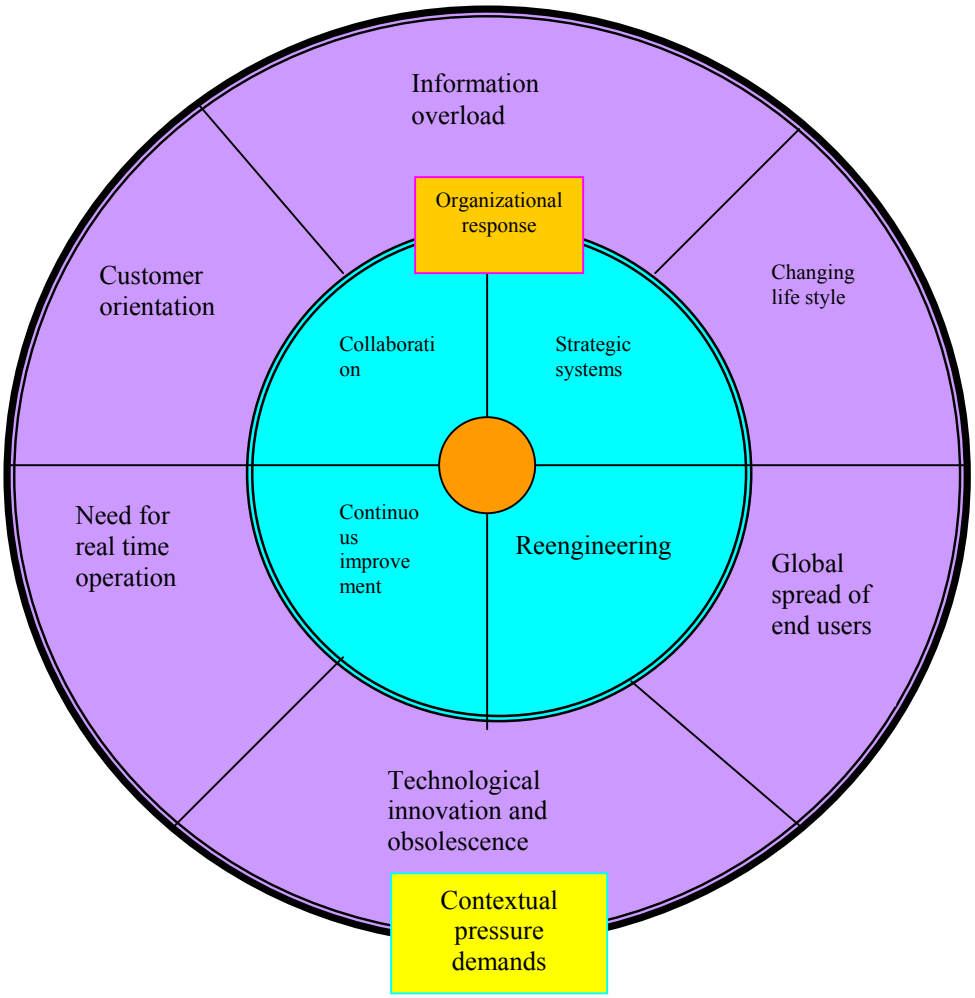
7.3 Continuous improvement

Many organizations make a continuous effort to improve their productivity and quality of information service. Continuous improvement is also essential for better decision making at all organizational levels. Appropriate decision making attempts to select the best, or at least a good enough alternative course of action.

7.4 Reengineering

Reengineering refers to the process of introducing a major innovation in an organization's structure or the processes, resulting in the possible overhaul of the organization's technological, human or organizational dimensions. If applied in the domain of information storage, it would mean, overhauling the information storage system to leverage strategic benefits from it.

The contextual factors and the organizational responses are presented graphically in the figure given below. A decision maker, when deciding on the use of a particular storage device, needs to not only keep in mind the advantages and disadvantages of various storage devices, but must also keep in mind the contextual pressures in which they have to operate. An optimum decision with respect to the use of a particular storage device would depend on the judicious mix and match of advantages of each of the storage device and the overriding contextual pressures.



8. Concluding remarks

This paper attempts to highlight the importance of storage devices in today's information intensive world and how decision makers can make use of the available information about storage devices in making informed decisions about the use of these storage devices. Different kinds of storage devices were discussed along with their advantages, disadvantages and their preferred applications. This was followed by a section on the contextual pressures that managers face with respect to managing, and in particular, storing information. The organizational responses to these contextual pressures were discussed and finally it was stated that, in order to make an informed decision with respect to the use of storage media, a decision maker must take into account not only the advantages and disadvantages of each storage media, but must also take in account the contextual pressures in which they have to operate. An optimum decision with respect to the use of storage media would depend on these factors.

References

1. Borgman, C.L. What are digital libraries? Competing visions. Information Processing and Management 35, 3 (May 1999), 227–243.

2. Wiederhold, G. Digital libraries, value, and productivity. Commun. ACM 38, 4 (1995), 85–96.

3. R.C Burns and D.D.E Long "Efficient Distributed Backing with Data compression, proceedings of the fifth Workshop on I/O in Parallel and distributed systems. ACM San Jose CAC Nov. (1997).

4. Information Technology SCSI Primary command-1 (SPC3) T10 working draft R.O Weber Editor T10 Technical Committee. International Committee for Information Technology (Sept. 17, 2002).

5. Network Management Protocol (NDMP) network data management protocol working group at http:// www.ndmp.org.

6. J.P Gelb "Systems Management Storage" IBM systems Journal 28, no. 1.77 – 103 (1989).

AGRILIBNET: A Web Portal

Rathinasabapathy G.[1] and Amudhavalli A.[2]

Abstract

Serial crisis is a challenge for most of the Indian agricultural university libraries. The agricultural university libraries are subscribing to a number of national and international journals to cater to the information needs of its target group of users. In this context, it is important to note that International journals do not widely cover issues that address research problems related to Indian condition. Regional agricultural and environmental issues and their research solutions have sparse coverage in the international journals. Hence, though efforts are initiated to form consortia amongst the agricultural libraries facilitating access to online journals, it is imperative to ensure also access to the Indian journals along with international journals including back issues to all academicians and researchers in the fields of agricultural sciences. The given resources' crisis warrants resource sharing. The advancement in IT is an enabler to accomplish this demand. To this end, an attempt has been made to design and develop a web portal – AGRILIBNET - for sharing the holding details of agricultural science libraries. This envisages to bridge the information gap and linkage amongst the agricultural science libraries in India.
Keywords: Library Resource Sharing, Library Networking, Web Portal, Agricultural Library

1. Introduction

Agriculture plays a vital role in the economic development of our country and it provides employment to 65% of the population. Further, it provides raw materials for various industries. In this context, it is important to note that the education, research and extension activities in the field of agricultural sciences need to be further strengthened to meet the food requirements of the growing population. India has built in a strong base of agricultural education. Earlier, agriculture was taught as one of the subjects in the conventional universities. Organized instruction in agriculture at university level was introduced in the beginning of 20th century when five agricultural colleges were established in 1907. At the time of Independence, India had only 17 Agricultural Colleges, 3 Veterinary Colleges and one Agricultural Engineering College. The Government of India was constituted several Commissions and Committees as a measure to plan the future requirements of Agriculture in this country and there was a realization on the part of policy makers that agricultural education ought to be recognized as a national issue (Bhatia, 2001).

The Radhakrishnan Commission Report on University Education (1948-49) recommended the establishment of 'Rural Universities' in India and this recommendation also been endorsed by the first Indo-American Team on Agricultural Research and Extension (1955. This led to the establishment of Agricultural Universities in the country

[1] Part-time Research Scholar and Assistant Librarian, Madras Veterinary College, Tamilnadu Veterinary and Animal Sciences University, Chennai 600 007. grspathy@yahoo.com
[2] Professor, Dept. Information Science, University of Madras, Chennai 600005 mudha75@yahoo.com

on the pattern of Land-Grant colleges in the United States of America. The first Agricultural University in India was established in 1960 at Pantnagar in Uttaranjal. At present, the country has 38 State Agricultural Universities, 5 Deemed to be Universities, 3 Central Universities dealing with various educational programmes and one Central Agricultural University (Appendix-1).

2. Agricultural Libraries In India

The Indo-American Agricultural Survey and Study Team in its final report on ICAR Institutes and Agricultural University Libraries (1969), recommended that every agricultural library is a special library which must be oriented to the service of agriculture. Its collection and the clientele are both discrete. The agricultural universities have recognized the need for proper library and information services and established libraries with a planned collection of books, periodicals, reports and other records to meet the study and research needs of a specific clientele.

At present all the 38 Agricultural Universities and their constituent colleges and the research institutes have libraries of their own. The Deemed Universities have also established libraries of their own to extend effective library and information services to their clientele. However, the services rendered in the Agricultural Libraries are confined to the agricultural scientists, researchers and students (Kaur, 1996).

Information is a vast dynamic and inexhaustible source that affects all disciplines and agriculture is no exception. Information in agriculture is used to support research and development, for the production of foods and services that touch all over lives. It can be used to improve the quality of life. But, the agricultural libraries are faced with a crisis – the cost of production and processing of documents is increasing much more rapidly than the budgets of the libraries which leads to non-availability of all the required resources in one library.

The role of agricultural information depends on the level of user group served and there are researchers, teachers and even policy makers and planners who have to keep abreast of new developments in agriculture all the time. Their main sources of information are research papers published in the scientific serial literature, conference proceedings (Perumalsamy, 2000). Though the 38 State Agricultural Universities and 5 Deemed Universities have established library and information service facilities, there is no sharing of resources among the libraries in a big way. Inter-library lending, the oldest concept of library cooperation, is also in the infant stage in India. But, effective and efficient library resource sharing is very important for the agricultural universities.

It was noted that out of 38 State Agricultural Libraries, 35 have their own web site. At least the list of current serials subscribed by the libraries can be made available over the web site of the university. But, that too is not available in the agricultural university web sites. The web sites of the agricultural universities contain only very brief information about the library and the current serials subscribed by the libraries, holdings details, web enabled online catalogue are not available in any of the agricultural libraries. The web site of the Indian Agricultural Research Institute (IARI) provides web enabled online catalogue and it is meant only for their library members which requires a login password to use the web OPAC. Efforts are being taken to form library consortia among agricultural libraries to subscribe to online journals. However, no effort has been taken so far to throw light on the back volumes of journals available in the agricultural libraries in India, which will provide very useful information to the agricultural scientists of this country.

3. Agricultural Library Resource Sharing

The serial crisis is a big problem for most of the Indian agricultural libraries. Many journal prices are increasing at a rate well over general inflation. The fluctuation dollar and other foreign currencies take its toll one hand and on the other hand the budgets of agricultural libraries have sustained deep cuts and there is really a crisis. In this regard, the crisis has to be seen as an issue of under funding as much as of over pricing of serials (Steele, 2004). But, it is mandate for the agricultural libraries to disseminate agricultural information to the students and researchers in agriculture to augment agricultural production of the country. It is necessary to provide latest information to the agricultural science educators, researchers and extension professionals. The main sources of information of these groups are research papers published in the scientific serial literature and conference proceedings. Therefore, agricultural libraries in India are subscribing to a number of national and international journals to cater to the information needs of the agricultural students and researchers. But, resource sharing among these libraries is limited which can help the libraries to ensure judicious spending on journals by avoiding duplication of journals, which are available in neighbor libraries.

In this context, it is important to note that International journals do not widely cover areas that address research problems related to Indian conditions. Localized agricultural and environmental issues and their research solutions have sparse coverage in the international journals (Bhattacharya, 2004). So it is necessary to ensure access to the Indian journals along with international journals to all students and researchers in the fields of agricultural sciences. Under the circumstances stated, library resource sharing seems to be the only way in the present environment of financial stringency to meet the demands of teachers, researchers and extension professionals in the fields of agricultural sciences. Further, there is no web portal for agricultural science students and researchers of India, which can provide single point access to the important web resources in agricultural sciences. The holding details of back volumes of journals available in various agricultural university and college libraries need to be brought to the knowledge of the agricultural science students and researchers and extension workers who are working in the width and breadth of the country.

4. Web Portal

Technically, a portal is a network service that brings together content from diverse distributed resources using technologies such as cross-searching, harvesting, and alerting, and collates this into an amalgamated form for presentation to the users and this presentation is usually via a web browser, though other means are also possible'. For users, a portal is a, possibly personalized, single point of access where searching can be carried out across one or more than one resource and the amalgamated results viewed. (Butters, 2003). The ideal library portal should have the most thorough coverage possible in several areas of the profession. The purpose of any library portal is to direct the user to specific information sources amongst the vast mass (O'Leary, 2000).

5. Objectives

This paper is an outcome of the research work undertaken towards the doctoral programme of the first author. The research intends to:

- develop an online database of current periodicals and back volumes available in Agricultural libraries in the Southern region of India towards resource sharing.
- design a web portal for the above developed data and host the web portal over the cyberspace for the use of educators, researchers and policy makers in the field of agricultural sciences.
- evaluate the usefulness and user-friendliness of the web portal

The objective of this presentation is to highlight the features, design and initiation in the direction of the proposed network among the chosen 6 State Agricultural University Libraries in Southern Region (SAU) (Appendix II)

6. Proposal

The proposed research project is named as AGRILIBNE (Agricultural Library Network). It is a web portal for sharing metadata of holdings of Agricultural Science libraries in South India to bridge the gap that exists in this area which can serve as an effective resource sharing.

6.1. Means & Methods

The details of the means and methods to accomplish the proposed AGRILIBNET are as follows:

- ❖ Collection of the bibliographical data of current periodicals and back volumes of the SAU and ICAR Institute libraries in India.
- ❖ Creation of the database of current periodicals and back volumes using *Ms-Access* as the back end.
- ❖ Designing and development of the Web Portal using *Ms-Access* as the back end and *Visual Basic* as front end.
- ❖ Enabling all the participating libraries to update their information available in the web portal from their desktop
- ❖ Provision of Password protection to all the participating librarians to ensure data security
- ❖ Registration of the domain name and hosting the Web Portal on the cyberspace for wide access.
- ❖ Evaluation of the web portal for its usefulness and user-friendliness among agricultural science professionals

6.2. Features Envisaged

The proposed AGRILIBNET will ensure easy access to union databases of serial holdings of State Agricultural University libraries through the Internet.

- The proposed AGRILIBNET will be hosted on the Internet so that its database of the holdings of the journal collection of the select six SAU libraries in India is made accessible worldwide. Since it is web-based, it will serve throughout the year.
- AGRILIBNET ensures to make available the valuable collection details of the journals and back volumes held in the Indian Agricultural University Libraries to every user one who can have access to the Internet. Any person who is interested to collect literature related to agricultural sciences can make use of the web portal.
- Updating of databases will be possible from the participating libraries from their desktop itself.
- Each and every participating library will be given a user name and password so to ensure security to the databases.
- Institutions, which do not have Internet access, can also make use of this facility through Internet browsing centers.
- AGRILIBNET will serve the information requirements of students, research scholars, faculty members and extension personnel in the discipline of agricultural sciences. AGRILIBNET will serve even for the agricultural researchers and agriculturists that are out of the agricultural university system.
- There is no ceiling for the participating institutions, as any number of libraries can become member of this web portal.
- Since it is a free service, the users need not pay anything to the web portal administration.

6.3. Requisites

AGRILIBNET web portal has to be registered and hosted on the Internet for wider access, which requires the following resources

6.3.1 Manpower

The AGRILIBNET will be hosted on the information superhighway after completion by the author. But, the sustainable development of the web portal requires support from the participating libraries of agricultural universities and institutes. Hence, skilled manpower is required to keep the web portal current and useful to the academic community. The manpower requirement is as follows:

- Professional Library Personnel with computing skills[*]

  Project Coordinator –1 and Library Assistants - 2

---

[*] *It does not mean that the above staff is exclusively required to maintain the web portal. The services of a professional library staff can be utilized on need basis in addition to their routine work.*

6.3.2. Budget

The following is the financial estimate towards the proposed AGRILIBNET:

1. Non-Recurring Cost

Domain Name Registration

..

Rs.2, 000/-

2. **Recurring Cost**

Web Portal Maintenance Charges

..

Rs.5, 000/year

3. Contingencies

..

Rs.5, 000/year

6.3.3 Equipments

The following are the minimum hardware and software required for the proposal:

Hardware

P-IV Computer with

- 40 GB HDD
- Base Memory 640 KB
- Extended Memory 3072 KB
- DDR RAM 128 MB
- CPU Type 80586
- FDD 1.44 MB
- CD-R Drive 52x
- Enhanced Graphics Adapter Present 256k
- CGA Colour Visual Display Unit
- 101 Key Board
- Mouse
- 0.5 KVA Stabilizer/UPS
- HP Deskjet Printer
- 56k Modem
- Scanner

Software

- OS Windows 1998/2000/XP
- MS-Access
- Visual Basic 6.0
- Internet Explorer 5.0

7. Projct Progress

The project has commenced and the details of the work initiated includes:

7.1. Data Collection

Data provision of the bibliographic details of the holdings of the journal collection has begun. Of the select six SAU libraries, data from the following universities have already been gathered.

1. Andra Pradesh NG Ranga Agricultural University, Hyderabad
2. University of Agricultural Sciences, Dharwad
3. Tamil Nadu Veterinary and Animal Sciences University, Chennai
4. Tamil Nadu Agricultural University, Coimbatore

Collection of data from the other two institutions - University of Agricultural Sciences, Bangalore; and Kerala Agricultural University, Thrissur is under progress.

7.2. Content Creation

The following databases have already been created for the data from the above said four SAU libraries.

1. Database of Current/Back volumes of Andra Pradesh NG Ranga Agricultural University, Hyderabad
2. Database of Current/Back volumes of University of Agricultural Sciences, Dharwad
3. Database of Current/Back volumes of Tamil Nadu Veterinary and Animal Sciences University, Chennai
4. Database of Current/Back volumes of Tamil Nadu Agricultural University, Coimbatore

On completion of data collection from the other two, which are under progress, the respective databases will be developed.

7.3. Web Design

Web Portal is being designed using *MS-Access* as the back end and *Visual Basic* as the Front End. The following are the few already designed:

I. Home Page of AGRILIBNET (Fig. 1)
II. User Authentication Screen (Fig. 2)
III. New Journal Entry Screen (Fig. 3)
IV. Journal Search Screen (Fig. 4)
V. Journal Edit/Update Screen (Fig. 5)
VI. Modify Account Screen (Fig. 6)

8. Conclusion

India is a very vast country with numerous agricultural institutions engaged in teaching, research and field works. The proposed project covers only the southern region and the six well-established SAU libraries, AGRILIBNET. However, this is envisaged as is an initiation in the direction of building a national level network and based on the evaluation of the project for its success it may continue to plan and programme for a wider network of all the agricultural libraries in the country.

References

1. S Bhatia. Agricultural Education: A Career Guide. New Delhi: Indian Council of Agricultural Research, 2001.
2. Sujit Bhattacharya, and B M Gupta. Towards developing a National Bibliographical Database in Science and Technology in India. University News, vol.42, no. 4, 2004, pp. 1–5
3. Geof Butters. What features in a Portal. Ariadne, vol. 35, 2004. URL: <www.ariadne.ac.uk/issue35/butters/intro.html>
4. H Kaur. Reading materials used by the Postgraduate students in Agricultural Universities: A Trend Report. IASLIC Bulletin, vol. 41, no.3 **,** 1996, pp/ 121-126.

I. 5.M O'Leory. Grading the Library Portals. Online, Nov. 2000. URL: <www.onlinemag.net/OL2000/oleary11.html>

5. K.Perumalsamy, Role of Library in Agricultural Information System. University News, vol. 38, no.2, 2000, pp. 6–9.
6. C Steele. Serials Cancellations. URL:< http://anulib.anu.edu.au/ news/stories/ serials.html>

APPENDIX -1
List of State Agricultural Universities and Deemed Universities in India

A. State Agricultural Universities

S.No	*Name of the University*	Address
1	**Acharya N G Ranga Agricultural University**	Rajendranagar, Hyderabad **Andhra Pradesh** PIN 500030
2	**Anand Agricultural University**	Anand, **Gujarat**
3	**Assam Agriculture University**	Jorhat, **Assam** PIN 785013
4	**Bidhan Chandra Krishi Vishva Vidyalaya**	Haringhatta PO Mohanpur Nadia, **West Bengal** PIN 741246
5	**Birsa Agricultural University**	Kanke, Ranchi, **Jharkhand** PIN 834006
6	**Central Agricultural University**	JROISEMBA, **Manipur** PIN 795001
7	**Ch Charan Singh Haryana Agricultural University**	Hisar **Haryana** PIN 125004
8	**Ch. Sarwan Kumar Krishi Vishwa Vidyalaya**	Palampur **Himachal Pradesh** PIN 176062
9	**Chandra Shekhar Azad University of Agriculrure & Technology**	Kanpur **Uttar Pradesh** PIN 208002
10	**Dr. Balasaheb Sawant Konkan Krishi Vidyapeeth (KKV)**	Dapoli **Maharashtra** PIN 415712
11	**Dr. Panjabrao Deshmukh Krishi Vishwa Vidyalaya**	Krish Nagar, Akola **Maharashtra** PIN 444104
12	**Dr. Yashwant Singh Parmar University of Horticulture & Forestry**	Solan **Himachal Pradesh** PIN 173230
13	**Govind Ballabh Pant University of Agriculture and Technology**	Pantnagar **Uttar Pradesh** PIN 263145
14	**Indira Gandhi Krishi Vishwa Vidyalaya**	Krishak Nagar, Raipur **Madhya Pradesh** PIN 492012
15	**Jawaharlal Nehru Krishi Vishwa Vidyalaya**	Jabalpur **Madhya Pradesh** PIN 482004
16	**Junagadh Agricultural University**	Junagadh, Gujarat – 362 001
17	**Kerala Agricultural University**	Vellanikkara, Trichur **Kerala** PIN 680654
18	**Maharana Pratap University of Agriculture & Technology**	University Campus, Udaipur **Rajasthan** PIN 313001
19	**Maharashtra Animal Science & Fisheries**	Seminary Hills, Nagpur

	Sciences University	**Maharashtra** PIN 440006
20	**Mahatma Phule Krishi Vidyapeeth**	Rahuri , **Maharashtra**
21	**Marathwada Agricultural University**	Parbhani, **Maharashtra** PIN 431402
22	**Narendra Dev University of Agriculture and Technology**	Faizabad **Uttar Pradesh** PIN 224229
23	**Navsari Agricultural University**	Navasari, **Gujarat**
24	**Orissa University Of Agiculture & Technology**	Bhubaneswar **Orissa** PIN 751001
25	**Punjab Agricultural University**	Ludhiana **Punjab** PIN 141004
26	**Rajasthan Agricultural University**	Bikaner **Rajasthan** PIN 334002
27	**Rajendra Agricultural University**	Pusa, Samastipur **Bihar** PIN 848125
28	**Sardar Krushinagar-Dantiwada Agricultural University**	Dantiwada, Sardar Krishi Nagar **Gujarat** PIN 385506
29	**Sardar Vallabh bhai Patel University of Agriculture & Technology**	Modipuram, Meerut **Uttar Pradesh** PIN 250110
30	**Sher-e-Kashmir University of Agricultural Sciences & Technology (Jammu)**	Railway Road, Jammu, **Jammu & Kashmir** -180004
31	**Sher-e-Kashmir University of Agricultural Sciences & Technology** (Kashmir)	Shalimar Campus, Srinagar **Kashmir** PIN 191121
32	**Tamilnadu Agricultural University**	Coimbatore **Tamilnadu** PIN 641003
33	**Tamilnadu Veterinary & Animal Sciences University**	Chennai **Tamilnadu** PIN 600051
34	**University of Agricultural Sciences**	Bangalore, **Karnataka** PIN 560065
35	**University of Agricultural Sciences**	Krishi Nagar, Dharwad **Karnataka** PIN 580005
36	**UPPDDU Pashu chikitsa Vigyan Vishwavidyalaya evam Go Anusandhan Sansthan**	Mathura **Uttar Pradesh** PIN 281001
37	**Uttar Banga Krishi Vishwavidyalaya**	P.O. Pundibari, Distt. Cooch Behar, **West Bengal** PIN 736165
38	**West Bengal University of Animal & Fishery Sciences**	68, Khudi Ram Bose Sarani, Belgachia, Kolkata, **West Bengal** PIN 700037

APPENDIX-II

State Agricultural Universities covered under the Study

S.No	*Name of the University*	Address
1	**Acharya N G Ranga Agricultural University**	Rajendranagar Hyderabad PIN 500 030 **Andhra Pradesh**
2	**Kerala Agricultural University**	Vellanikkara Trichur PIN 680 654 **Kerala**
3	**Tamilnadu Agricultural University**	Coimbatore PIN 641 003 **Tamilnadu**
4	**Tamilnadu Veterinary & Animal Sciences University**	Chennai PIN 600 051 **Tamilnadu**
5	**University of Agricultural Sciences**	Bangalore PIN 560 065 **Karnataka**
6	**University of Agricultural Sciences**	Krishi Nagar Dharwad PIN 580 005 **Karnataka**

Metadata Standards Available for Cataloguing Indian Manuscripts: Comparative Study

Harinarayana N.S. and Gangdharesha S.

Abstract

Cataloguing of manuscripts is one of the areas, which has received a scant attention by the Indian cataloguers. Information technology has enabled not only to provide better opportunity to preserve the content of the manuscripts for posterity but also helped to provide efficient access from remote areas to these manuscripts. Towards digitalization of the content of manuscript, Metadata standards are necessary. Although attempts have been made in this direction, no single standard has been adopted for cataloguing of manuscripts. This leads to a situation where inconsistency and non-uniformity in cataloguing practices of manuscripts exists. Hence, this paper aims to identify appropriate standards for the description of manuscripts, comprehensively and efficiently. The standards available for description of manuscripts such as 'Digital Scriptorium' and 'Text Encoding Initiative' (TEI-MSS) are critically examined and compared.

1. Introduction

Cushioned in the hoary antiquity of several millennia, the cultural heritage of India descends down to us through manuscripts. These manuscripts constitute sources of very valuable information for the social, cultural and literary history of India. Thus manuscripts are the written ambassadors of our history; they are the cultural heritage reserves of our country. They contribute in threading up the cultural, literary, educational, economical and historical continuity of life, lived in different stages of cultural evolution. India, an important country of the orient, possesses millions of manuscripts like the Arab countries. Manuscripts were produced in all parts of the country: they are in many languages and scripts and on a wide range of religious, philosophical, historical, literary and scientific subjects. Preservation and organisation of these valuable sources of information is a big challenge to the information professionals. Issues and problems in handling these manuscripts, particularly cataloguing them, are being discussed in this paper.

2. C**ataloguing of Indian Manuscripts**

A manuscript is a unique book; there are no duplicates. Even if the text is same, as in the case of a model and it's copy, the two books are different in script, physical appearance and even text, because a handwritten copy always contains some changes in comparison with its source.

According to M L Saini, *"If a collection of printed catalogues of Indian manuscripts were brought together at the present time it would extend over more than 50 feet (or about 15 metre) of shelving and would constitute in itself a specialist library*
We have reason to suppose that more than one million of Indian manuscripts are deposited in libraries public or private, and that more than 6,00,000 different manuscripts have been listed in some manner in printed catalogues since the origins of Indian studies a

[1] *Library and Information Science, University of Mysore, Mansagangothri, Mysore. harinarayana@vidyanidhi.org.in*
[2] *J R D Tata Memorial Library, Indian Institute of Science, Bangalore 560012. kishan_gangu@rediffmail.com kishan@library.iisc.ernet.in*

century and a half ago." And so many other features of the Indian manuscripts pose challenges tot he organizers and cataloguers. However, to highlight a few, the problems in cataloguing of Indian manuscripts are summarized as follows:

- Uncertain authorship. Authors are almost bee-hived in obscurity and the variants in Indian name only add confusion to obscurity.
- Proliferation of the different manuscripts of the same text both in India and abroad and also the mass of related peripheral literature render the task of identification and authentication difficult.
- Over whelming enormity of anonymous works.
- Lingual and scriptural variations also contribute to confusion. Synonymy in sounds and signs leads to interchangeability. Errors of scribes have lead to the complete divergence between the meanings of the text of the original and its manuscripts.
- Loose-leaf practice of manuscript maintenance has resulted into mutual shuffling of pages.
- Indian manuscript catalogues swarmed up inordinately both in quality and quantity, quite unrelated and independent of one another even in the same very language.

- Hence, cataloguing a manuscript is a difficult task, which requires a skill and competence. However, with the advent of IT, using Metadata standards, T comprehensive cataloguing and access to manuscripts can be made easier and simpler.

3. Metadata

Metadata is structured data, which describes the characteristics of a resource. The term *'Meta'* derives from the Greek word denoting a nature of higher order or more fundamental kind. It shares many similar characteristics to the cataloging that takes place in Libraries, Museum, Archives, and Manuscripts etc. This enables people searching for electronic information to find the information they are seeking more efficiently. Metadata is invaluable for both discovering and using resources because is succinctly describes, manages catalogues, and retrieves the resources, efficiently and effectively.

- Some of the most popular Metadata schemes include:
- Dublin core
- USMARC (United States Machine Readable Cataloguing)
- AACR-2 (Anglo-American Cataloguing Rules 2nd Edition)
- GILS (Government Information Locater Service)
- RDF (Resource Description Framework)

Each metadata scheme will usually have the following characteristics:

- A limited number of elements.
- The name of each element.
- The meaning of each element.

Syntax, though, is not strictly part of the metadata scheme, the data will be unusable, unless the encoding scheme understands the semantics of metadata scheme. The encoding allows the metadata to be processed by a computer program. A few of these are:

- HTML (Hyper Text Make-up Language)
- SGML (Standard Generalized Make-up Language)
- XML (Extensible Make-up Language)
- RDF (Resources Description Framework)
- MARC (Machine Readable Cataloguing)
- MIME (Multipurpose Internet Mail Extension)
- Z.39.50
- X.500
- LDAP (Light-weight Directory Application Protocol)

Typically, the semantics is descriptions of the contents, location, physical attributes, type (eg: text, image, map or model.) and form (e.g. Print copy, electronic file, and manuscript). The metadata elements supporting access to published documents include the originator of work. When and where it was published and the subset areas it covers where the information is issued in analog form, such as print material, additional metadata is provided to assist in the location of the information e.g.: call numbers, accession numbers, used in libraries. The resource community may also define some logical grouping of the elements or leave it to the encoding scheme.

4. Metadata Standards For Manuscript Cataloguing

Many metadata standards are available for cataloguing of manuscripts, out of which *Digital Scriptorium and Text Encoding Initiative* are briefly discussed here:

4.1 Digital Scriptorium

Digital Scriptorium, a project started in 1996 with a grant of the Andrew W. Millon Foundation to the Bancroft Library of the University of California at Berkeley and the Rare Book and Manuscript Library of Columbia University, in New York. Consuelo W. Dutschke, Curator of Medieval and Renaissance Manuscripts at Columbia, says “Digital Scriptorium is a web-based visual union catalogue of medieval and renaissance manuscripts: web based, in that it was intended form the start as on expendable and correctable tool, in the way that the static media of print or even CD-ROM cannot be; Union, because it meshes into one database the holdings of a number of institutions; and visual, because it includes at least one image of every manuscript described, thus allowing the user to test and eventually correct the catalogued information by means of his own specialized knowledge.”

At present Digital Scriptorium contains about 2,200 entries, and 2000 images. Digital Scriptorium has been created with an Access database compatible with the TEI-DTD. Digital Scriptorium database dictionary accompanies the Microsoft. “Access 97” database structure designed in 1997 at the University of California, Barkelay by John Hassan and Merrillee Profitt in consultation with C.W Dutschke (Curatov of Medieval and manuscripts at Coloumbia) and Martha Rust (then at Berkely), has been incorporated over a period of time. The database in intended to collect in an efficient and organized manner basic information about manuscripts held by the Digital Scriptorium partners, and to provide searching prints for images form these holdings.

The database is divide into two sections (1) Document and (2) Image, The document containing ownership i.e. (historical) information about the manuscript, then the descriptive (i.e. physical) and bibliographic (i.e. textual) information about the manuscript, as well as the caption information for individual images. Thus document section has four parts:

a. Manuscript
b. Part
c. Test
d. Image

In practice to first level "manuscript" identifies the single codex; the second level "part" contains most of the physical information about the codex; the third level, "Text" holds bibliographic information; the fourth level, "image" refers to what the end-user sees on the web. Each successive level is dependent upon the proceeding level (s), ion a potentially, but not necessarily many-to-one relationship to its predecessor. The Second section called 'Image' which is not defined here, contains the processing data about the image (i.e. image capture metadata, for example, type of capture, film type, shatter speed, frame etc.,). In this database totally 67 fields are provided to describe the given manuscript among which 14 fields are required fields.

4.2.Text Encoding Initiative For Manuscripts Cataloguing (Tei-Ms):

Text Encoding Initiative is an interaction and interdisciplinary project, established in 1987, to develop guidelines for the encoding of textual material in electronic form for research purposes. TEI became consortium in 2000. The University of Oxford, the Brown University, the University of Bergen and the University of Virginia host the TEI-Consortium; a sub-group is devoted to manuscript cataloguing. TEI is an international standard that helps libraries, museum, publishers and individual scholars to represent all kinds of literary and linguistic texts for online research and teaching, using an encoding scheme that is maximally expressive and minimally obsolescent. TEI is an interchangeable format TEI continues to develop and maintain encoding standards. It has a specific mark-up syntax as well as well defined tag set, but few tags are mandatory.

TEI, Header is a set of descriptions prefixed to the TEI encoded document that specifies four components:

a) File description (a full bibliographic description)
b) Encoding description (level of detail of the analysis- the aim or purpose for which an electronic file was encoded; editorial principles and practices use during the encoding of the text.
c) Text Profile (Classificatory and contextual information such as text's subject matter; the languages and sub-languages used; the situation in which it was produced, the participation and their setting)
d) Revision history (history of Changes during the electronic files development) contains bibliographic information supporting use of the resources.

5. Comparison of Standards for Cataloguing of Manuscripts.

An attempt has been made to identify the data elements prescribed by *Digital Scriptorium* and *Text Encoding Initiative* (TEI-MS) for preparing metadata manuscripts. It may be noted from the table given below that both the system provides a very detailed list of elements required for cataloguing of manuscripts. The tables given below provide the features of both the standards. Table 1 provides a corresponding data elements with

definition of both the elements. As many as 30 data elements are common to both, however there 'name' and 'definitions' vary slightly and table 2 provides those elements not found in TEI but available in *Digital Scriptorium* and vice versa in table 3.

Table-1

TEI – MS DATA ELEMENTS			**DIGITAL SCRIPTORIUM DATA ELEMENTS**	
Sl No.	**Name of the element**	**Definition of the element**	**Name of the element**	**Definition**
1.	<idno>	Suppliers the string of abbreviations and numbers, usually referred to as the 'call number', 'shelfmark', 'accession number', etc., that is used to identify the item.	Shelf mark	Collection name, item number, size designate that the particular institution uses to identify on specific manuscript.
2.	<altname>	Contains any form of alternative identifier used for a manuscript. Eg:- Nick name, former catalogue No. ocellus nominum etc.,	Nick Name	Names commonly used to refer to certain more famous manuscripts
3.	<country>	Contains the name of a geo-political unit	Country	Country of origin of the manuscript
4.	<region>	Contains the name of a smaller geo-political unit, intermediate between country and settlement.	Region	Modern political subdivision of a country (Region, Country, Department Land, Province etc.,)
5.	<settlement>	Contains the name of the smallest component of a place name.	City	Name of the city where the manuscript is held.
6.	<institution>	Contains the name of the organisation, within which a manuscript repository is held.	Institution	Name of the institution where the manuscript is held.
7.	<repository>	Contains the name of the repository(usually a distinct a physical building) with in which manuscripts are stored, forming part of an institution	Library	Name of the library where the manuscripts is held
8.	<author>	Identifies the primary author of the work or works contained in a manuscript	Author	Name (s) of Author (s)
9.	<title>	Supplier a title for the work or works contained in a manuscript, or a brief description of a manuscript.	Title	Title of the work.
10.	<respStmt.>	Supplies a name and description for someone other than an author credited with intellectual responsibility for some aspect of	Other associated names of authors.	Names of other associated with a given text.

		the work/s within a manuscript.		
11.	<origDate>	Contains any form of date, used to identify the date of origin for a manuscript or manuscript part	Date Begin date, end date Year = month = day	Date of origin of the manuscript Date of the manuscript when it is dated by the scribe.
12.	<textLang>	Describes the languages and writing systems used by a manuscript.	Language (s)	Names of the language used in the manuscript text.
13.	<note>	Contains any additional descriptive information about a manuscript / item.	Note	A non defined slot to be adopted to the needs of a specific manuscript.
14.	<bibl>	Contains a conventional bibliographic description for example of a modern edition of item	Bibliography	Bibliographic
15.	<decoNote>	Contains a note describing either a decorative component of a manuscript or fairly homogeneous class of such components.	Other Decoration	Any other significant decoration
16.	<explicit>	Contains the text of any explicit attached to a particular manuscript item, that is, the closing words of a text or a section of a text sometimes used as a kind of title, possibly followed by one or more rubrics or colophons.	Explicit	Ending words of a text often used to identify it.
17.	<incipit>	Contains the text of any incipit attached to a particular manuscript item, that is the opening words of a text.	Incipit	Beginning words of a text, often used to identify it.
18.	<rubric>	Contains the text of any rubric or heading attached to a particular manuscript item, that is a string of words where by a manuscript signals a text division (begin, end) which is in some way set off from the text itself, usually in red ink, or by use of different size or type of script, lining, or other such visual device.	Rubric	Author - and /or title-like information in the manuscript, often set off from the main body of the text by red ink, underlining, display script or spacing.
19.	<support>	Contains a description of the written materials making up the manuscript item	Support	They physical material that support the script.
20.	<extent>	Describes the approximate size of the manuscript specified in any convenient units.	Total Folios.	In reference to the individual part, the number of leaves in

21.	<collation>	Contains a description of how the leaves or bifolia are physically arranged.	Span of Folios	number of leaves in Arabic numerals.
22.	<watermarks>	Contains a detailed description of the watermarks identified in the paper of which a manuscript is composed.	Watermark	The identifying mark used by the paper maker.
23.	<dimensions>	Element is available for use any where in a description, the cataloguer may choose to discuss (for example) dimensions of miniatures at the same time as describing the miniatures, rather than specify that information within the <extent > element.	Outer Dimensions	Dimensions of the book block.
24.	<layout>	Contains a description of ruling technique and layout.	Layout	Indication of ruling technique and layout.
25.	<musicNotation>	Contains description of type of musical notation.	Music	Type of musical notation, when present.
26.	<bindingDesc>	Describes the present and former bindings of a manuscript, either as a series of paragraphs or as a series of distinct binding elements, one for each binding of the manuscripts.	Binding	Book binding
27.	<binding>	Contains a description of one binding, i.e. type of covering, boards, etc., applied to a manuscript.		
28.	<foliation>	Describes the numbering system or systems used to count the leaves or pages in a codex	Total Folios	In reference to the extent of leaves of the entire codex.
29.	<provenance>	Contains any descriptive or other information concerning a single identifiable episode during the history of a manuscript or manuscript part, after its creation but before its acquisition.	Provenance	Successive owners, private and institutional.
30.	<msPart>	Contains information about an originally distinct manuscript fragment now forming part of a composite manuscript.	Composite Manuscript	Statement as the whether or not the codex is composed of more than one part.

The Data Elements Which Are Not Found In Text Encoding Initiative Set But Found In Digital Scriptium

Table-2

DATA ELEMENTS	DEFINITION
PHYSICAL ISSUE (S)	Presence of seals if the item is an archival document; roll or other non-codex format; incunable or printed book
REPRODUCTION	Container for notes on existence of microfilm, negatives, etc
ACKNOWLEDGMENTS	Public recognition of contribution to knowledge about the manuscript.
SOURCE	Source of the information input into the database.
INPUTTER	Name of person responsible for selecting, organizing and inputting the information about the manuscript into the database, in reference to all four levels (Manuscript; Part; Text; Image) even though the field "Inputter" is only present at the "Manuscript" level; date of initial/main inputting.
REVISER	Name of person revising a given record and date of the revision.
REVISIT	Flag referring to any one or several fields in this level as a reminder that the level contains something that is incomplete/uncertain/possibly wrong.
PART NUMBER	A roman numeral designating the parts of a given codex that have independent origin from one another (frequently but not necessarily with differing place or date).
SPAN OF FOLIOS	In reference to the individual part, the number of leaves in arabic numerals.
CARDINAL POINT	Point of the compass or the word "central" for country of origin; second of four place-defining elements.
DOCUMENT	Identifies reports of legal activity.
ALPHABET	Indication of primary writing system of the manuscript.

SCRIPT	Type of script used, in practice usually a subset of the Latin alphabet.
NUMBER OF SCRIBES	Number of readily distinguishable hands with span of leaves copied by each person.
SCRIBE	Name of the scribe.
ARTIST	Name of artist/decorator
SEQUENCE	Arabic numeral to ensure proper sequencing on the web.
SUPPLIED TITLE	Term that indicates a category of books, rather than a specific text.
DOCKET	Abridged outline of proceedings in a legal document, giving names, dates, action, etc.
STATUS OF TEXT	Statement as to type of problem when a text is not complete or disordered.
FOLIO NUMBER(S)	The location within the manuscript of the photograph.
CAPTION	Text accompanying image on the Web, serving both to point out an interesting feature (justification for the choice of shot), and to allow for free-word searching.
ICONCLASS	A hierarchical code comprised of numbers and letters to identify iconography; see Iconclass manual/CD-Rom for the codes themselves; for overview, see website: http://iconclass.let.ruu.nl/home.html
NOTES TO PHOTOGRAPHER	Notes to photographer.

Table –3 The Elements Which Are Not Found In Digital Scriptorium Set But Used In Tei

DATA ELEMENTS	DEFINITION
<origPlace>	Contains any form of place name, used to identify the place of origin for a manuscript or manuscript part.
<mscontents>	Describes the intellectual content of a manuscript or manuscript part either as a series of paragraphs or as a series of structured manuscript items.
<msItem>	Describes an individual work or item within the intellectual content of a manuscript or manuscript part.
<colophon>	Contains the text of any colophon attached to a particular manuscript item: that is, an inscription, usually found at a break point such as the end of a text or codex, usually containing information about the production of the manuscript, such as the name of the scribe, the date and place of the copying, the person who commissioned the copying, etc.
<listBibl>	Contains a list of related bibliographic descriptions (standard TEI element)
<summary>	Contains a brief summary of the contents of an item provided by the cataloguer
<form>	Contains a description of the physical format of the manuscript.
<msWriting>	Contains a description of all the different kinds of writing used in a manuscript.
<handDesc>	Describes a particular style or hand distinguished within a manuscript.
<condition>	Contains a description of the physical condition of the manuscript.

<additions>	Contains a description of any significant additions found within a manuscript, such as marginalia or other annotations.
<history>	Groups elements describing the full history of a manuscript or manuscript part.
<origin>	Contains any descriptive or other information concerning the origin of a manuscript or manuscript part.
<acquisition>	Contains any descriptive or other information concerning the process by which a ms or ms part entered the holding institution.
<additional>	Groups additional information relating to the modern bibliography for a manuscript, its current curatorial status, and and other associated materials.
<adminInfo>	Contains information about the present custody and availability of the manuscript, and also about the record description itself.
<surrogates>	Contains information about any digital or photographic representations of the manuscript being described which may exist in the holding institution or elsewhere.
<accMat>	Contains details of any significant additional material which may be closely associated with the manuscript being described, such as non-contemporaneous documents or fragments bound in with the manuscript at some earlier historical period.

6. Conclusion

Catalogue is a tool or pointer for archiving and providing access to old treasures of knowledge. Systematic arrangement, uniform and mutual consistent cataloguing of manuscripts is needed. In the present information technological era the information present in the manuscripts should be digitized, preserved and provide access to these old treasures of knowledge, by developing common suitable metadata standard and make them useful by providing efficient assessing points to them and opens the communication channels is needed. Thus this study suggests that *Text Encoding Initiative and Digital Scriptorium* standards data elements can be adopted for cataloguing of Indian manuscripts. Further study is recommended.

References:

1. M L Saini. “Manuscripts literature in Indian languages”, ILA Bulletin, vol.v, no.1:6-21, Jan-March, 1969.
2. http://sunsite.berkeley.edu/seriptorium/index.html.

3. http://www.inform.nu/articles/vol 4/v 4nzp o61-066.pdf
4. http://www.homes.dsi.unimi.it/alberti /pubb1/ichim.pdf
5. Chintaharavan Chakravarti,. Value and importance of manuscripts in olden times. Journal of Royal Asiatic Society Bengal (letters),vol..16, 1950: pp. 253-61.
6. http://www.ils.unc.edu/moony/whatismeta.html
7. http://www.tased.edu.ac/tasonline/metada/
8. Dorothy K.Coveney. Cataloguing of Indian manuscripts. Journal of Documentation,vol..6, 1950: p.125.
9. Klaus Ludwig Janet. An annotated bibliography of the catalogues of Indian manuscripts Pt. I Wiesbaden, Steiner, 1965, p.9.

Implementing RFID in Library: Methodologies, Advantages and Disadvantages

Narayanan A., Sanjay Singh and Somasekharan M.

Abstract

A library is a growing organism. As it grows in size the problems associated with the maintenance and security of the documents also grows. The researchers have always helped the librarian in solving their problems. To solve the problems of arranging documents in order they have given classification schemes. To solve the problems of searching documents they have given cataloging guidelines. To solve the problems of space and time they have taught librarians to digitize the documents and share over network. To automate the counter activities they gave us bar-codes. Bar-codes have served the librarians and libraries for a long time, and now it is slowly getting replaced by RFID.
This paper discovers the technology, implementation methodologies, advantages and disadvantages of RFID in Library.

1. Introduction

RFID (Radio Frequency IDentification) invented in 1969, patented in 1973, first used in harsh industrial environment in 1980s', and standards presented in 2001, is the latest addition of technology to be used in the libraries for a combination of automation and security activities in the well maintenance of documents either inside the library or goes out-of library. RFID uses wireless radio communications to uniquely identify objects or people, and is one of the fastest growing automatic data collection (ADC) technologies, which is comprising one or more reader/interrogators and RF transponders in which data transfer is achieved by means of suitably modulated inductive or radiating electro-magnetic carriers. In addition it can be used as a data carrier, with information being written and updated to the tag on the fly. RFID systems carry data in suitable transponders, generally known as tags, and retrieve data, by machine-readable means, at a suitable time and place to satisfy particular application needs.

RFID is a combination of radio-frequency and microchip. RFI chips are of particular interest, because they have become smaller and smarter to the point where they can be added every kind of document and can be read and updated from a distance [1]. The data capacities of transponder normally range from a few bytes to several kilobytes. There are also 1-bit transponder (without chip) to fulfill monitoring and signaling functions called Electronic Article Surveillance (EAS). In writable transponders, the reader can write data to the transponder in three procedures. Inductively coupled RFID system uses EEPROMs, FRAMs and microwave systems commonly use SRAMs. The important feature of power supply to the transponder is drawn either from the field of reader (Passive tag) or from the battery incorporated in the tag (Active/Semi-active tag).

Scientific Information Resource Division, IGCAR, Kalpakkam E-mail:nan@igcar.ernet.in, sanjjay@igcar.ernet.in

2. RFID Technology in Libraries

The concept of RFID can be simplified to that of an electronic barcode and can be used to identify, track, sort or detect library holdings at the circulation desk and in the daily stock maintenance. This system, consist of smart RFID labels, hardware and software, provides libraries with more effective way of managing their collections while providing greater customer service to their patrons.

The technology works through flexible, paper-thin smart labels, approximately 2”X2” in size, which allows it to be placed inconspicuously on the inside cover of each book in a library’s collection. The tag consists of an etched antenna and a tiny chip which stores vital bibliographic data including a unique Accession number to identify each item. This contrasts with a barcode label, which does not store any information, but merely points to a database. These smart labels are applied directly on library books and can be read with an RFID interrogator/scanner. Line of sight is not essential for reading the tags with the scanner, therefore, the books require much less human handling to be read and processed. A middleware or Savant software integrates the reader hardware with the existing Library Automation Software for seamless functioning of circulation.

The information contained on microchips in the tags affixed to library materials is read using radio frequency technology regardless of item orientation or alignment. It provides a contact less data link, without need for line of sight, for example, the documents in the shelves or cardboard boxes can be checked without removing or opening. RFID has no concerns about harsh environments that restrict other auto ID technologies such as bar codes. Tags have a discrete memory capacity that varies from 96 bits to 2kbytes. In addition to tags, an RFID system requires a means for reading or "interrogating" the tags to obtain the stored data and then some means of communicating this tag data to library information system.

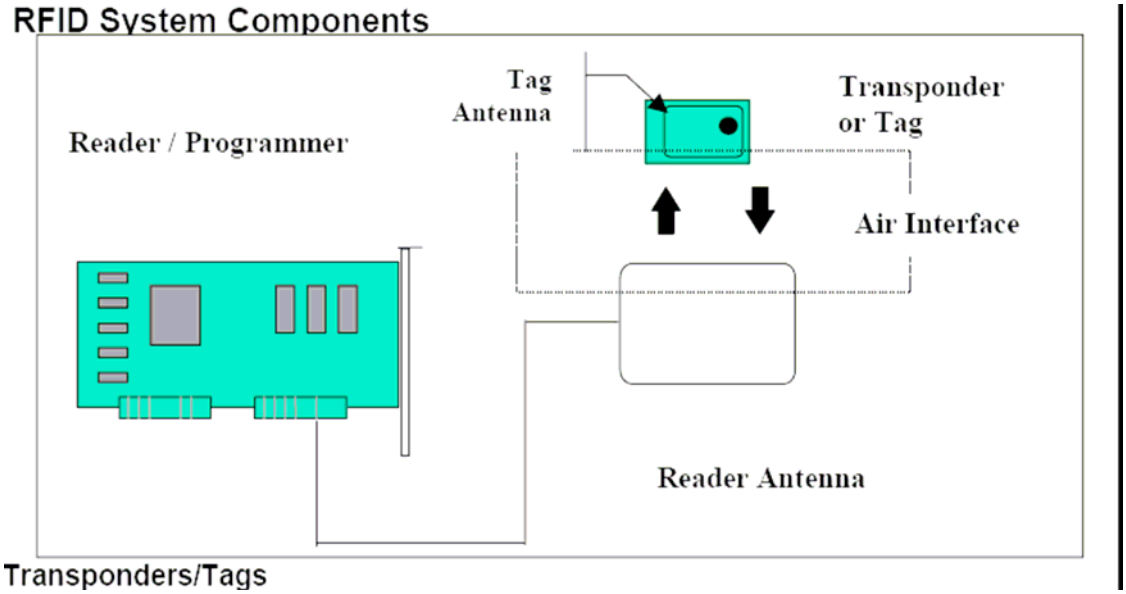
RFID-based systems have been implemented for efficient document tracking purpose through out the libraries that combine, easier and faster charging and discharging of documents, security of materials, inventorying, stock verification and shelf handling. RFID tag’s transponder listen for a radio query from the reader and respond by transmitting their unique ID code. Most RFID tags have no batteries, they use the power from the initial radio signal to transmit their response.

2.1 RFID Components

Normally a RFID package for library consists of eight components: RFID tags, a self check-out station, a staff check-out station, a self-return book drop with an automatic check-in feature, a tagging station, a set of security gates, a shelf scanner for inventory and an administrative station. The self-checkout station allows patrons to borrow books without assistance from the library staff. The staff checkout station is used when patrons prefer staff assistance. The book drop allows returned books to be processed instantly by updating the database the moment the items pass through the chute. The shelving station speeds the process of sorting the returned books for re-shelving. The shelf scanner allows library staff to take inventory and find wrongly shelved books without having to pull the books off the stacks.

3. How RFID Works

Figure-1



3.1 RFID systems:

In typical system tags are attached to objects. Each tag has a certain amount of internal memory(EEPROM) in which it stores information about the object, such as its unique ID, or in some cases more details of bibliographic data and product composition. When these tags pass through a Radio Field generated by a reader, the transponder in the tag transmit the stored information back to the reader, thereby identifying the object.

3.2 How Tags Communicate

The communication process between the reader and the tag is by wireless. The major differences between the different types of waves are the distances covered by one cycle of the wave and the number of waves that pass a certain point during a set time period. The wavelength is the distance covered by one cycle of a wave. The frequency is the number of waves passing a given point in one second. For any electromagnetic wave, the wavelength multiplied by the frequency equals the speed of light. The frequency of an RF signal is usually expressed in units called hertz (Hz). One Hz equals one wave per second. Basically what happens is that when the reader is switched on it starts emitting a signal at the selected frequency band (**in library HF is used with 13.56 MHz**). Any corresponding tag in the vicinity of the reader will detect the signal and use the energy from it, to wake up and supply operating power to its internal circuits. Once the tag has decoded the signal as valid, it replies to the reader and indicates its presence by modulating (affecting) the reader field.

Table 1. Frequency Bands and Applications

Frequency Band	Characteristics	Typical Applications
Low 100-500 kHz	Short to medium read range Inexpensive low reading speed	Access control Animal identification Inventory control Car immobiliser
Intermediate 10-15 MHz	Short to medium read range potentially inexpensive medium reading speed	Access control Smart cards
High 850-950 MHz 2.4-5.8 GHz	Long read range High reading speed Line of sight required Expensive	Railroad car monitoring Toll collection systems

3.3 Anti-collision

If many tags are present (in a row of books) then they will all reply at the same time, which at the reader end is seen as a signal collision and an indication of multiple tags. The reader manages this problem by using an anti-collision algorithm designed to allow tags to be sorted and individually selected. The number of tags that can be identified depends on the frequency and protocol used, and typically range from **50 tags/s for HF and up to 200 tags/s for UHF**. Once a tag is selected the reader is able to perform a number of operations such as read the tags identifier number, or in the case of a read/write tag write information to it. After finishing dialoging with the tag the reader can then either remove it from the list, or put it on the stand by until a later time. This process continues under the control of anti-collision algorithm until all tags have been selected.

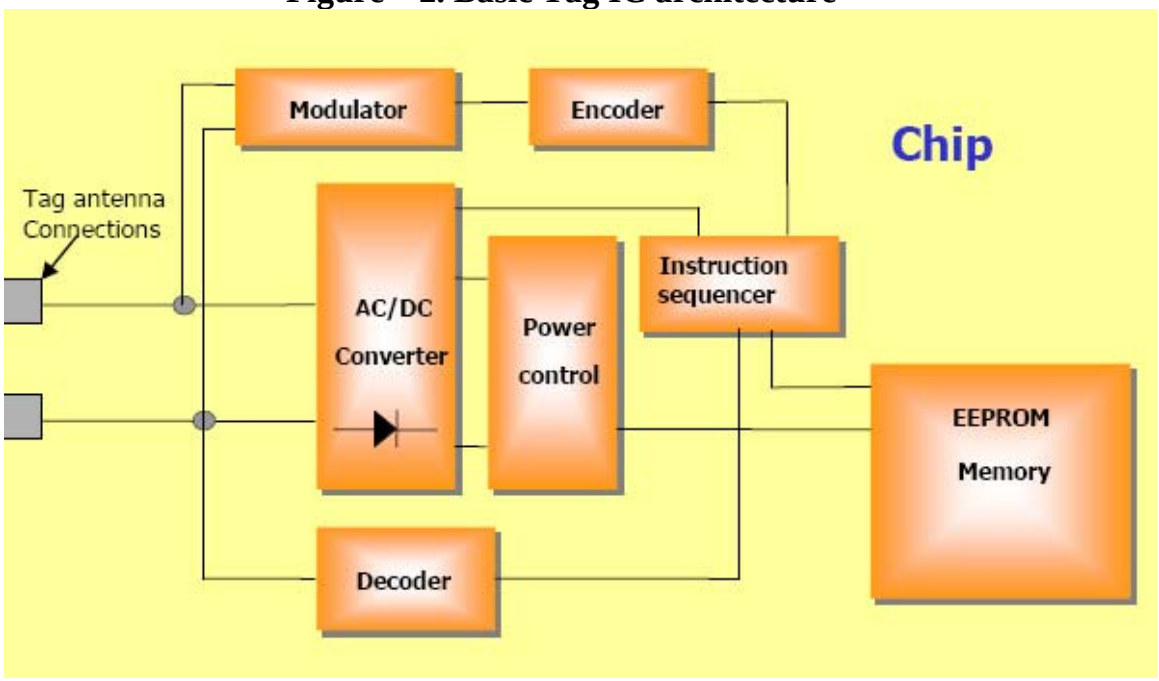
In fact very real challenges for the ICs'exist such as achieving very low power consumption, managing noisy RF signals and keeping within strict emission regulations. Other important function of the circuit is to allow the chip to transfer power from the reader signal field, and convert it via a rectifier into supply voltage. The chip clock is also normally extracted from the reader signal. Most RFID tags contain a certain amount of NVM (non-volatile memory) like EEPROM in order to store data.

The amount of data stored depends on the chip specification, and can range from just simple identifier numbers of around 96 bits to more information about the product with up to 32Kbits. In 1999 the AUTO-ID centre (now EPC Global) based at the MIT-USA, together with the number leading companies developed the idea of an unique electronic identifier code called the EPC(Electronic Product Code). The EPC is similar in concept the UPC used in barcodes today. Having just a simple code of up to 256 bits would lead to smaller chip size, and hence lower tag cost, which is recognized as the key factor for wide spread adoption of RFID. Like a barcode, the EPC is a 96 bit unique number which is divided into numbers that identify the manufacturer, product, version and serial number.

3.4 Tag IC's A single-chip design led to the RFID tag, a small device composed of a chip, an antenna, and an optional power source, that carries a unique identifier. The 1990s witnessed the use of such tags for card-keys, fuel-station payment systems, and automated toll payment. Such tags were typically specialized for a certain class of

applications and cost a few dollars each. The tags typically stored application-specifc data and were capable of modest processing on-tag [2].

Figure – 2. Basic Tag IC architecture



3.5 Tag Classes: one of the main ways of categorizing RFID tags is by the capability to read and write data. This leads to the following four classes and EPC global has also defined five classes which are similar to the one below:

3.5.1 Class 0: Read only – factory programmed. These are simplest type of tags, where the data, which is usually a simple id number is written only once into the tags during manufacture. The memory is then disabled from any further updates. Class 0 is also used to define a category of tags called EAS or anti-theft devices which have no id, and only announce their presence when passing through an antenna field.

3.5.2 Class 1: Write Once Read Many(WORM) - Factory or user programmed. In this case tag is manufactured with no data written in to the memory. Data can then either be written by manufacturer or by the user – one time. Following this no further writes are allowed and the tag can only be read. Tags of this type usually act as simple identifiers.

3.5.3 Class 2: Read-Write – This most flexible type of tag, where user have access to read and write data into the tags memory. They typically used as data loggers, and there fore contain more memory space than what is needed for just a simple id number.

3.5.4 Class 3: Read-Write (with on board sensors) – These tags contain on board sensors for recording parameters like temperature, pressure and motion, which can be recorded by writing into the tags memory. As sensor readings must be taken in the absence of a reader, the tags are either **semi-passive or active.**

3.5.5 Class 4: Read-Write (with Integrated Transmitters) – These are like miniature radio devices which can communicate with other tags and devices without the presence of the reader. This means that they are completely active with their own battery power source.

3.6 Active and Passive tags:

First basic choice when considering a tag is either passive or semi-passive or active. Passive tags can be read at a distance of up to 4 – 5 m using UHF frequency band, whilst the other types of tags (semi-passive and active) can achieve much greater distance of up to 100m for semi-passive, and several KM for active. This large difference in communication performance can be explained by the following;

- passive tags use the reader field as a source of energy for the chip and for the communication from and to the reader. The available power from the reader field, not only reduce very rapidly with distance but is also controlled by the strict regulations, resulting in a limited communication distance of 4 -5 m when using UHF frequency band (860 MHz – 930 MHz) .
- Semi-passive (battery assisted back scatter) tags have build in batteries and therefore do not require energy from the reader field to power the chip. This allows them to function with much lower signal power levels, resulting in greater distance of up to 100meters. Distance is limited mainly due to the fact that tag does not have an integrated transmitter, and is still obliged to use the reader field to communicate back to the reader.
- Active tags are battery powered devices that have an active transmitter onboard. Unlike passive tags, active tags generate RF energy and apply to the antenna. This autonomy from the reader means that they can communicate at the distance of over several KMs.

Table No.2 : Different Tag Classes

class	Known as	Memory	Power source	Applications
0	EAS/EPC	None/EPC-1bit on/off	Passive	Anti-theft/ID
1	EPC	Read only	Any	Identification
2	EPC	Read-Write	Any	Data logging
3	Sensor tags	Read-write	Semi passive/active	sensors
4	Smart Dust	Read-write	Active	Ad hoc networking

4. Selecting tags:

Choosing the right for a particular RFID applications is an important consideration, and should take into account many of the factors listed below:

- Size and form factor – where does the tag have to fit?
- How close the tags be to each other
- Durability – will the tag need to have a strong outer protection against regular wear and tear
- Is the tag re-usable
- Resistance to harsh environment(corrosive, steam…)
- polarization – what will be tag orientation with the respect to the reader field
- exposure to different temperature ranges
- communication distance
- influence of the materials such as metals and liquids

- environment(electrical noise other radio device and equipments)
- operating frequencies(LF,HF, UHF, MW)
- supported communication standards and protocols
- regional regulations(Europe, Asia, USA..)
- will tag data need to store more than just an id number like an EPC
- anti-collision how many tags in the field at the same time how quickly must they be detected
- how fast will tags move through the reader field
- does the tag need to have security data protection by encryption
  reader support – which readers products are able to read the tag read the tag

4.1 Difference between Barcode and RFID

- Information can be read from RFID tags much faster than from barcodes
- Several items in a stack/counter can be read at the same time using RFID
- Items do not have to be handled one-by-one nor removed from the shelves
- Inventory-taking is no longer a tedious operation
- RFID can stand more than 10,000 read/write
- RFID can have theft bit which can be in two states "ON/OFF"
- Shelf verification/rectification can be done on daily basis
- More information can be written in the RFID tag on incremental basis
- Need not open/remove books to capture information
- Items are identified on upper and lower shelves more comfortably

Basic technology comparison Barcode versus RFID:

RFID	BARCODE
Can be Read and Write	Read only
No line of sight required	Needs direct visible contact to reader
Multiple items can be read simultaneously (anti-collision)	Single item scan only
Item attendant data (mobile data-carrier)	Database look-up is always necessary
Guaranteed data retention of at least 10 years	Limited lifetime due to printing.
Stock verification made easier as No need of taking the books out from shelf. You can read multiple books from the shelf at a time.	Stock Verification Takes time because of the fact that each book has to take out from shelf and then scanned with the scanner.

Key features of an RFID-Library versus a Barcode solution:

- Fully transparent stock control on all available books and medias in the library.
- The Electronic Article Surveillance (EAS) feature offers a secure stock control system and no books get "lost". This feature is especially important for rare books or unique books, which should not get lost because of their value.
- Due to the fact that RFID-labels can be identified (read) very easily (without line of sight, over a certain distance, several books simultaneously), the so-called self-check-out system can be used and this will save manpower and increases the attractiveness of the library.
- Because the RFID-Library offers a fully controlled and categorized library stock, the number of similar books can be reduced because the turn-around time of each book can be shortened drastically.
- RFID-label stores data of the book and its system status that gives the possibility to check the book without the database.

5. Implementation of RFID:

The methodology for implementation can be divided into many phases taking into consideration of budget provision, the types of document holdings, number of volumes, types of items meant for circulation, and the number and types member the institution has. Care should be taken to integrate the library automation package while detailed tender specification are drawn. Since the technology is new to Indian library environment proper demonstration of the system can be arranged and should visit the library where the system is successfully running. While evaluating the tender the past experience of firm supplying the equipment, tags, reader and software should be thoroughly investigated. The fixing of tags to documents can be initially outsourced then in house arrangement can be done after proper training. The reader should be able to read the other manufacturers RFID Tags. The provision for reading the existing barcode in the document can be made and the required data can downloaded by interacting with the present database and can be written to the tag. The tags can be over layered with the self adhesive sticker containing the logo of the library or the institution for longer life. Until sufficient confidence is gained with the system, old system in practice can be continued.

5.1 Retrospective conversion of already existing stack requires a "programmer" or "conversion station." The conversion of existing barcoded items, including affixing the tags to library materials, takes 15-30 seconds per item depending on the amount of information added to the tag and the skill of the person doing the tagging. Pre-programmed tags, which are used for new acquisitions in libraries that want only identification numbers on the tags, take even less time because they do not involve scanning existing barcodes. The speed of conversion can be increased by dividing responsibility for removing and replacing library materials, converting the barcodes, and inserting the tags among at least three people. It is essential that the tasks be rotated so that no one repeats the same motions over an extended period of time. Almost all libraries tag **new acquisitions** as part of the cataloging process, however, libraries that have experienced losses of unprocessed library materials from technical services, might consider doing the tagging at the time of receipt in acquisitions. While inadvertent duplicates cannot then be returned, it should significantly reduce losses and facilitate tracking of items in technical services.

5.2 Readers: A typical RFID system includes three different kinds of readers, also known as sensors or scanner/wand. These devices designed to detect and read tags to obtain the information stored thereon.

(i)The types of readers include staff workstations for circulation desk charging and discharging, patron self-charging stations, and longer-range walk-through exit sensors to detect and read an RFID tag passage for purposes of determining whether it is a charged or discharged.

(ii)RFID exit sensors at exits are of two types, one reads the information on the tag(s) going by and communicates that information to a server. The server, after checking against the circulation database, activates an alarm if the material is not properly checked-out. Another type relies on a "theft" byte in the tag that is turned on or off to show that the item has been charged or not. It is then not necessary to communicate with the circulation database. The security system will work even though the online library server is not working.

(iii)The portable scanner or inventory wand, can be moved along the items on the shelves without touching them. The data goes to a storage unit, which can be downloaded at a docking station or a server later on, or it can go to a unit which will transmit it to the server using wireless technology.

6. Advantages of RFID systems :

6.1 Rapid charging/discharging: The use of RFID reduces the amount of time required to perform circulation operations. The most significant time savings are attributable to the facts that information can be read from RFID tags much faster than from barcodes and that several items in a stack can be read at the same time. While initially unreliable, the anti-collision algorithm that allows an entire stack to be charged or discharged now appears to be working well.

6.2 Simplified patron self-charging/discharging: For patrons using self-charging, there is a marked improvement because they do not have to carefully place materials within a designated template and they can charge several items at the same time. Patron self-discharging shifts that work from staff to patrons. Staff is relieved further when readers are installed in bookdrops.

6.3 High reliability: The readers are highly reliable. Some RFID systems have an interface between the exit sensors and the circulation system to identify the items moving out of the library. Were a patron to run out of the library and not be intercepted, the library would at least know what had been stolen. If the patron card also has an RFID tag, the library will also be able to determine who removed the items without properly charging them. This is done by designating a bit as the "theft" bit and turning it off at time of charge and on at time of discharge.

6.4 High-speed inventorying: unique advantage of RFID systems is their ability to scan books on the shelves without tipping them out or removing them. A hand-held inventory reader can be moved rapidly across a shelf of books to read all of the unique identification information. Using wireless technology, it is possible not only to update the inventory, but also to identify items which are out of proper order.

6.5 Automated materials handling: Another application of RFID technology is automated materials handling. This includes conveyor and sorting systems that can move library materials and sort them by category into separate bins or onto separate carts. This significantly reduces the amount of staff time required to ready materials for reshelving. Given the high cost of the equipment, this application has not been widely used.

6.6 Long tag life: Finally, RFID tags last longer than barcodes because nothing comes into contact with them. Most RFID vendors claim a minimum of 100,000 transactions before a tag may need to be replaced.

6.7 Fast Track Circulation Operation

The use of RFID reduces the amount of time required to perform circulation operations. The most significant time savings are attributable to the facts that information can be read from RFID tags much faster than from barcodes and that

several items in a stack can be read at the same time. While initially unreliable, the anti-collision algorithm that allows an entire stack to be charged or discharged now appears to be working well.

7. Disadvantages of RFID Systems:

7.1 High cost: The major disadvantage of RFID technology is its cost.

7.2 Vulnerability to compromise: It is possible to compromise an RFID system by wrapping the household foil to block the radio signal. It is also possible to compromise an RFID system by placing two items against one another so that one tag overlays another. That may cancel out the signals. This requires knowledge of the technology and careful alignment.

7.3 Removal of exposed tags: The RFID Tags can not be concealed in either spine or gutter of the books and are exposed for removal. If a library wishes, it can insert the RFID tags in the spines of all except thin books, however, not all RFID tags are flexible enough. A library can also imprint the RFID tags with its logo and make them appear to be bookplates, or it can put a printed cover label over each tag.

8. Evaluating RFID from different vendors

It is potentially overwhelming to evaluate competitive offerings of a new technology; hence the following guide lists some of the characteristics to be considered.

8.1 Security feature

The same RFID tag used to manage inventory can also be used to protect it from theft. Current offerings provide the choice between a purely RFID solution, or RFID with an EM (electro-magnetic) add-on for theft.

8.2 Tag memory capacity

More memory is not necessarily better than less - it often correlates with price, and data transmission speed. As a first step, consider what information you need to program into each tag, and then discuss with vendors.

8.3 Tag functionality

8.3.1 Read/Write vs. Read Only

- Some vendors offer tags which can only be “written to” once. That is, once the tag is programmed, the information stored in the tag’s memory cannot be changed. Alternatively, information stored in the memory of read/write tags can be updated as required.

8.3.2 Anti-collision

- All RFID vendors in the library market offer a product with anti-collision (the ability to read several tags simultaneously). However, the speed at which this can be performed, and the total number of tags that can be read, will vary. This relates specifically to inventory management with a hand-held reader, and check-in processes.

8.3.3 EAS (Electronic Article Surveillance) mechanism

- As mentioned above, RFID can be used to prevent theft in the library. This approach varies from vendor to vendor – the security mechanism may be integrated into the chip itself, or security gates may be linked to a separate server which interrogates the database to conclude whether an alarm needs to be triggered.

8.4 Cost

- Expect to pay from US$0.85 to over US$1 per tag.
- The price of hardware (per unit) varies extensively from different suppliers. However, the infrastructure requirement also varies.

8.5 Standards

- The emerging standard for library RFID solutions is to employ a frequency of 13.56MHz. However, no formal standards are currently in place [6].

9. Conclusion

Though the unique advantages and flexibility of RFID is the good news, the technology is still not yet widely understood or installed in the library environment, and the cost/ROI models far from established. RFID, its application, standardization, and innovation are constantly changing. Its adoption is still relatively new and hence there are many features of the technology that are not well understood by the general populace. Developments in RFID technology continue to yield larger memory capacities, wider reading ranges, and faster processing. The interest in RFID as a solution to optimize further the automation and tracking of documents are gathering momentum at an increasing pace, with more libraries joining the trails.
"RFID is increasing in popularity among libraries, as the early adopters of this technology have shown that, it makes good economic sense, both for large and small libraries."

References:

1. Dave Brich at Consultant Hyperion. URL:www.chyp.com as visited on 1-2-2005. As quoted in "News Comment – Contact less crazy" in Journal: Card Technology; pub by Elsevier. p2, February 2005.
2. Sudarshan S. Chawathe ,et al."Managing RFID Data" Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004.
3. Klaus Finkenzeller, "RFID Handbook, 2nd Edn" - translated by Rachel Warrington. John Wiley, 2003.
4. Steve Lewis, "A Basic introduction to RFID Technology and its use in the supply Chain". White paper by LARAN RFID Jan 2004.
5. Ashim A Patil, "i-TEK RFID Based Library Management Suite" a White paper. by Infotek Software & Systems PLtd, Pune, India. 02, April, 2004.
6. "RFID for Libraries" as viewed on 01-03-2005 at URL: http://www.biblio-tech.com/html/rfid.html.
7. "RFID Survival" as viewed on 25-03-2005 at URL: http://www.rfidsurvival.com/contact.html

Author Index