International Journal of Researches in Social Science and Information Studies (IJRSSIS)
September 2014
Issue-III, Volume-II
Online Journal
ISSN No. 2347-8268

INFORMATION UTILITY IN THE NEW ERA

S. H. Gawali And S.S. Sawarkar
Librarian
J.B. College of Science Wardha

ABSTRACT :

An information utility is a distributed repository of a variety of materials such as books, periodicals, news, airline/train schedules, music, video, experimental data, commodity prices etc., of interest to the general public. It is ideally accessible to anyone, anytime, anywhere. Such an information utility is now emerging using the internet. It brings with it many new problems of intellectual property rights, security, accessibility, cost and ethics. In this talk we will highlight these issues.

INTRODUCTION :

Traditionally information has been accessed from a variety of libraries. Every University maintains a large library with a diverse collection of books and other materials such as audio tapes, video tapes, microfilms, microfiche, etc. Besides a central library, departments maintain their own special libraries of interest to a small group of researchers. The library system is well developed – companies maintain libraries of interest to them, individuals have their own libraries, most cities have public libraries. It has been recognized that access to information is essential in modern civilized society and investment on libraries have grown over the years.

Recently Google (famous for its search engine) has initiated a project to scan and place books of several libraries on their web site which will be accessible to all. There are copyright issues which have been resolved. Out of print books in the library will be put in Google site with arrangements to pay a fee to copyright holders.

Scholarly and other information available in libraries is not the only information people are concerned about. There are a variety of other

A Four-monthly Journal By
Vishwashanti Multipurpose Society,
(Global Peace Multipurpose Society)

information people need in their day-to-day life in a complex society. These include government rules and regulations, daily news, up to date information on prices of commodities, shares etc., schedules of public transport, to cite a few.

The advent of computers half a century ago set in motion a new paradigm of information storage and retrieval. Early researchers worked on methods of classifying information for ease of retrieval in a computer based system. Research was impeded due to non availability of large machine readable corpus of information as disks were of small capacity and manual transcription of information was slow and expensive.

This situation has changed now. There has been a convergence of a number of developments in computer technology in the last five years which has significantly affected the way computers can be used to access information. These developments are:

Emergence of CDROMs (Compact Disk Read Only Memories) and now DVDROMs (Digital Versatile Disk Read Only Memories) with very high information storage capability. One DVDROM can store upto 7.5 Giga bytes (7.5×10^9 bytes) (To store a typical 500 page book 0.25 Mbytes are needed). The cost of these storage devices is very low, around ten paisa per Megabyte.

Continuous increase in capacity of magnetic disks which can be used for on-line access. Today (2008) desk top PCs have 160GB disks. Storage capacity of disks is doubling every twelve months, at constant price

Development in computer network technology which has facilitated interconnecting computers not only within the country but also across countries leading to a world wide computer network. Network bandwidths are also doubling almost every 9 months at constant price.

Wireless technology also rapidly developed. This allows anywhere – any time access to information even when a person is mobile.

Method of digitizing, compressing and storing text, audio, graphics and video data have continuously improved. Standards have emerged for audio compression, e.g.MP3 format, graphics (JPEG) and video data compression(MPEG4). Standards allow easy interchange of these data.

What is an information utility?
What are the unique advantages of a computer based information utility?
How will such a utility affect our day-to-day work?
What is the relevance of these developments to India?
What is an information utility?

We attempt to define our concept of an information utility using the analogy of an electrical power utility. In early days of power generation, each city or community had a local generating station which supplied power to the consumers in its immediate vicinity. There was hardly any standardisation. Direct current (DC) was supplied in some cities and alternating current (AC) to others in their neighbourhood. Electrical gadgets could not be used when one moved to a city with a different power supply. Excess generation by a city could not be used by its neighbours. Engineers realised the need for standardization of supply voltages and frequency, need to interconnect generating stations and agreeing on distribution networks and strategies. This led to modern power systems with its attendant advantages of optimization of power generation, fault tolerance, development of a large consumer market for electrical gadgets, cost reduction due to economy of scale and availability of power to geographically remote areas. Thus a power utility is characterised by Distributed generating stations. Interconnection of generating stations and creation of a distribution network

Standardisation of supply to ease access and enable wide use. Regulation of power generation, tariffs, and adherence to standards.

INFORMATION RESOURCE

Textual data - This consists of books and journals and other useful information such as patents, international standards, specifications, etc., stored in a digital form in a computer's disk store. There are two ways of storing this information. One way is to photograph a page and scan the image with a scanner. The scanner digitizes the image storing a 0 for white and 1 for a dark spot. For good resolution one page will be represented by (800 × 1000) bits (or 100 Kbytes). This form of storage is called a bit mapped form. Bit patterns do not carry information for indexing. This is, however, the only practical way of storing old manuscripts, texts and journals. The image of a page may be retrieved and displayed on the video screen of a computer.

The other way of storing a text is to represent each character by its ASCII code. Texts generated using a word processor is already in this form. Most books and journals produced in the past few years will already be in this form. If a page has 6000 characters it will need 6000 bytes of storage. Further, it will be easy to index the document using arbitrary words in the text. If a table has numeric information, the numeric data would be stored in coded form which allows it to be processed. Photographs or other complex figures in the text, however, will have to be scanned and stored as bit maps.

As it requires less storage to store text in ASCII coded form, software is becoming available to scan printed texts using a scanner and convert them to coded form. Conversion by such software is , however, not 100% accurate and manual correction is required before the text is stored. Good conversion software for standard fonts are currently able to give 95

to 98% accuracy. For old texts using non standard or mixed fonts and for hand-written manuscripts such conversion software is not available.

Numeric data- consist of tables of various types such as physical property data of various materials, data from experiments, astronomical tables, stock prices etc. Such numeric data stored digitally may be used (if required) by curve fitting programs, spread sheet programs.

Graphics data- may be photographs, maps, drawings, land records etc. The simplest way of storing such data is to scan the image and store it as a bit pattern. There are better ways of coding and storing maps, drawings etc., which abstract the information contained in them. For example, maps may be stored using longitude/latitude as coordinates of cities, a linked list depicting road network etc. Data stored in this form eases retrieval.

Photographs (both colour and monochrome) are stored in bit mapped form using compression algorithms to reduce storage space. Formats known as bmp, tif, gif and jpeg are now commonly used.

Audio data- is digitized, compressed using a commonly accepted standard compression algorithm (called MP3 format) and stored. Musical scores may also be coded and stored along with the audio data (if required). Video data requires enormous storage space due to the need for repeating frames atleast 30 times per second. Thus the data is compressed in such a way that when decompressed the original data is recovered. Common standards for compression have been evolved. The current standard is called MPEG-4 (Motion Picture Experts Group - Version 4) and compresses one 90 minute video movie to occupy 7Gbytes.For details of methods of acquiring, compressing, storing, processing and dissemination of multi-media data one may refer to the book by V. Rajaraman

INDEXING

Indexing and interlinking multimedia data is extremely important for ease of retrieval. Key words in textual documents are selected and linked to related words with logical links by appropriate software. This is called a *hypertext*. For material in other media (audio, video) also, related elements are selected and linked in what is known as *hypermedia*. Such links would allow an user to navigate through multimedia material. For example, from a multimedia encyclopaedia stored in a CDROM one may request information on the Taj Mahal. The computer would search the data and retrieve a page giving textual information about Taj Mahal which would be displayed on the video screen. If there is a reference to music in the text it may link to an audio clip giving a recording of classical music of that time. Links may also be present to video clips on Taj Mahal and related subjects.

LINKING

The information collection of the utility will normally not be stored in one computer. It will be distributed in many computers known as servers. All these servers will be linked by high speed communication links. The fact that the information is distributed need not be known to a user as it is not relevant from his/her point-of-view. A user gets a "seamless" access to the information based on his/her request regardless of its geographical location.

USER

A user may access information from anywhere using a terminal or a computer, called a client, connected to the network to which the information servers are connected. New types of services which are now popular are music downloads provided by Apple Computers in a hand held device known as Apple iPOD. A large library is available and one

may download individual tracks of an album on payment of a fee. Another emerging facility is YouTube (recently acquired by Google) which provides video clips stored by numerous amateurs and professionals for free download.

Amazon has recently introduced a service using an e-book reader called Kindle. Kindle is battery operated, portable and uses e-ink technology which is easy to read. It uses a mobile network to enable users to download books from Amazon's book list and store them locally. The cost of books is a third of print version. One has to buy Kindle which costs around $ 250.

To summarise, the key components of an information utility are:

A large collection of digitized and compressed multimedia data.
All data logically linked together and indexed with key words (or elements) to enable easy search and retrieval.
The data collection is geographically distributed on a computer network.
Users are geographically distributed and connected to the network.
Seamless access is available to all "consumers" to the data stored on servers connected to the network.
Availability of search programs for accessing desired information.
Technologies which enabled creation of information utility

Unique advantages of an Information Utility

The fact that all types of data are in a digital form and data is distributed and accessible from anywhere endows some unique advantages to an information utility. We discuss some of them in this section.

Unlike traditional libraries documents are not physically handled by a user. A user views the necessary document and prints the portions of interest in his/her location. Further, many users can simultaneously view a document. Thus documents are not "lost" due to theft or

misplacing. Documents do not tear due to wear. Library need not have multiple copies of a popular journal. Rare manuscripts may be allowed to be viewed by many as only images are accessed by users.

Unlike printed text where tables of numbers can be studied but not easily processed, if they are in digital form the numbers in a table can be processed. In print material numbers are "passive" whereas in digital form the same numbers are "alive" as they can be processed and transformed. For instance, a table may be used in a spread sheet program and we can find out how altering some entries change other entries. A user may also try to fit appropriate curves to a set of numbers in a table and give his/her own interpretation of the data.

The ability to digitally store and retrieve multimedia data allows one to effectively provide access to audio and video information. Providing such access is particularly difficult in current libraries due to difficulty in handling such material and wear and tear if many persons access the material. Information Utility would also provide an effective method to preserve rare manuscripts, vintage movies, old music, etc., without denying access to users.

An Information Utility would provide users access to current information which normally takes a long time to reach a traditional library. For example, many authors store their latest research reports in their web page and permit free access. Many conference papers may be found in World Wide Web. An information utility provider can collect articles of interest to a research group and provide it in a local server thereby saving users' time.

The internet allows easy access to discussion groups in specified subjects. Users of an information utility can find researchers working in

similar areas and attempt collaborative work. An information scientist can enable this collaboration.

We are also seeing the emergence of journals published only in "electronic form" with no print version. This reduces cost and allows prompt, wide dissemination of the journal. Some journals also provide means for readers to criticise an article and link such review with the article. Such informal review is unorthodox but is very useful for a prospective reader.

Many other types of information not found in a traditional library may be provided by an information utility such as computer aided lessons, lecture demonstrations by musicians, dancers and artists, lecture video of important speeches (such as Nobel lectures) etc.

To summarise the unique features of emerging information utility are :

Safe storage and multiple access of multimedia data Ability to process numerical data published in the literature. Ability to store variety of data such as audio, video, graphics, data output from experiments, computer aided lessons, lecture demonstrations by artists, famous lectures etc.

CONCLUSION :

We have seen that the emergence of many technologies - both software and hardware - has led to a revolution in the way information is created, stored and disseminated. Rapid international access to the variety of information distributed across the world will to some extent alleviate problems faced by scientists in the developing world with regard to information availability. There are, however, many problems which need to be resolved before a world wide information utility develops. Some of these are:

Copyright problem. It is easy to copy digital information. Methods have to be found to prevent illegal copying without inhibiting legal users. Material being digitized are in both page image form and ASCII coded form. ASCII form is preferable for wide use of textual matter but technology is not yet available for digitizing archival material to ASCII.

A disturbing aspect, particularly for developing countries is the rapid obsolescence of both hardware and software. This would put higher financial burden in changing equipment.

Special precautions need to be taken to prevent corruption of data by vandals bent on mischief. Authentication and quality control of information available in the utility is now almost non-existent. This task is a difficult, time consuming and expensive.

Lastly, we are not used to the idea of paying for information. The question of who pays when information is accessed and how much payment is considered reasonable has to evolve.

REFERENCES :

1. Berkeley Digital Library Sun site (Information on digital libraries) (Web address: http//sunsite.berkeley.edu).
2.Borgman (Christine L.). What are Digital Libraries? Computer Vision. Information Processing. Vol.35; 1996; 227-43.
3. Communications of ACM, U.S.A. Vol. 38, No.4, April 1995 (special issue on Digital Libraries).
4. IEEE Computer, Vol.29, No.5, May 1996 (Special issue on U.S.Digital Library initiative).
5. Fox, Edward, Digital Library Source Book, 1993 (available at web address:
http://fox.cs.vt.edu//DLSB.html)

6. Fox, Robert, Tomorrow's Library Today, Communications of ACM, Vol.40, No.1, Jan.1997, pp.20-21

7. Harter (Stephen P) (1996). The impact of electronic journals on scholarly communication: a citation Analysis. Public-Access Computer System Review 7

8. http://www.dlib.org/dlib/july98/rusbridge/07 rusbridge.html

9. Rajasekhar T.B. Digital Libraries, Resonance (Indian Academy of Science Journal on Science Education), Vol.2, No. 4, April 1997

10. Shemeent (Jorge Reina). Encyclopaedia of Communication and Information. Vol.2; 2002.pp.449-553

11. Towards a World Wide Digital Library", Communications of ACM, U.S.A., Vol.41, No.4, April 1998

12. V.Rajaraman. Introduction to Information Technology" Prentice Hall of India, New Delhi, 2003.MF.24.11.08