DESIDOC Bulletin of Information Technology, Vol. 18, No. 5, September 1998, pp. 19-24

Trends in IT: The Media Switchover Dilemma

M Jagan Mohan Rao, B Annapurna, KH Babu and Dr RN Biswas*

Abstract

Many Libraries and Information Centres all over the world are facing the dilemma of striking a balance between the age-old print media and new electronic media. An experiment run in the United States during 1991-95 under code name 'TULIP' and the 'CORE' project based at the Cornell University, US, provide insights into the issue and facilitate the resolution of the dilemma. The salient findings of these experiments and the likely shape the search for online information may take are briefly discussed.

1. INTRODUCTION

Libraries and information centres in the research and development establishments subscribe to a number of journals, most of which are of foreign origin, technical in nature and expensive. A trend that is discernible in the field of publishing of technical literature is, though the print media is retained, electronic media is also used parallely and in increasing volume every year. For example, all the IEEE journals are now available in CD-ROM for besides print copy editions. The number of journals switching over to online or CD-ROM format are increasing every year. All the important patents, standards like DOD and other international standards are now supplied in CD-ROM form. Many technical books these days contain floppies or CD-ROMs to augment the information in the book. The major appeal of CD-ROMs is that they bring information back in the local setting like print media, and unlike it, the space required to store the information is much less. The CD-ROMs are attractive alternative due to the ease and speed of access to the information and the elimination of connect time expenses involved in the online services.

2. ESTABLISHMENT OF DIGITAL LIBRARIES

In a number of establishments/academic institutions, 'DIGITAL LIBRARY has been established to deal with the information in digitised as well as in electronic form. The dilemma facing many institutions is the apportionment of financial resources between conventional and new sources of information. The present status and an intelligent guess on the future shape the field of information is likely to take are two vital inputs to resolve both the long- term and short-term investment issues. An experiment involving many of the parameters like generation and publication of information in computer accessible form, its communication to the users, hardware and software issues involved, user preferences etc. that influence the shape of things to come were conducted in the United States during 1991 to 1995. The code name of the experiment was TULIP (The University of Licensing Program). The focus of the experiment was to test one key aspect, namely, networked delivery of journals to user's

*Defence Research & Development Laboratory Hyderabad - 500 058.

desktop and the extent of their usage. The results and conclusions of the experiment are helpful in providing pointers to the resolution of not only to journals issue but also the future of information in electronic form. Another study with similar objectives though specific to the field of chemistry is the CORE (Chemical Online Retrieval Experiment)[2] project also in US, based at the Cornell University. It is a large experimental electronic library project involving realistic laboratory environment. The objectives and the essential conclusions reported so far in these projects are presented in brief in this article.

3. DIGITAL LIBRARY

3.1 Definition

Digital library[3] may be defined as electronic information collection containing large and diverse repositories of digital objects, which can be accessed by a large number of geographically distributed users. Digital objects include text, images, maps, sounds, videos, catalogues and indices, and scientific, business and government data sets as well as hyper textual multimedia composition of such elements. In the present day context, in India, digital libraries are likely to be a part of the main libraries traditionally holding the print media documents and would be viewed as complementary rather than competitor to the latter.

The key feature of digital libraries is that they are network-based distributed systems with individual servers responsible for maintaining local collections of digital documents ranging from sets of electronic texts to video-on-demand services. A digital library would provide a coherent, consistent view of many repositories of information allowing users to seamlessly connect and interact with information with no regard to geographic location or time. The term 'seamless' in the context of digital libraries means that the user need not be aware of the computer and communication related heterogeneous environment that could exist between him and the information sought.

The key components of digital libraries can be listed as follows:

(a) Geographically distributed digital information collection and users

(b) Information represented by a variety of digital objects

(c) Large and diverse collections

(d) 'Seamless' access

4. ADVANTAGES OF DIGITAL LIBRARIES

The advantages of digital libraries are many. In the area of search for information, fast online search is replacing printed abstract journals. Ubiquity, in the sense a single electronic copy can be accessed from many locations by many simultaneous users, is another advantage. The other advantages are in the area of preservation. Since the digital information can be copied by many without error and preserved without deterioration, making up-to-date current literature available to users and access to new kinds of multimedia materials have no equivalent in any traditional format.

5. THE TULIP STUDY

The TULIP study by a consortium of one of the leading technical literature publishers (Elsevier Science) and nine renowned universities (Carnegie Mellon, Cornell, Georgia Institute of Technology, MIT, California, Michigan, Tennesse, Washington and Virginia) is to jointly test the systems for networked delivery and use of journals at the user's desktop. The technical conclusions of the study are the following:

(a) The first technical conclusion was on the web. The web is a set of protocols that give Internet users interactive access to a huge variety of content on the Internet. In many people's view, the web is the content itself and not the underlying protocols. Though the web does not provide some real-time functionality features like image-zooming available in the windows applications, users favoured and shifted to the web.

(b) The bandwidth restriction and other limitations of the prevailing transmission

schemes do not permit large scale Internet FTP (File Transfer Protocol) transfer. FTP is a program used on Internet to connect one's computer with the remote host computer and allow search and selection of files for transfer.

(c) The electronic collection building of institutions is likely to be layered, composed of local servers for primary relevant material and remote servers for material of secondary importance.

6. EASE OF USE

The TULIP study has brought out that the end users have certain requirements regarding functionality, ease of use and critical mass of electronic information services. With respect to ease of use, the services should be as intuitive as possible and should be preferably from a familiar interface. The access to all information should be from one source. The search capabilities should be very effective with high processing speed. The information should be timely and there should be linking of information. The quality of image text should be good. The number of journals covered should be sufficient and their time coverage should be adequate. While there is enthusiasm about the desktop access to electronic information, the end of print media is far away still, if at all it happens. The reason is that there are practical benefits of paper products besides 'emotional' ties with paper for many users.

7. NEW DEVELOPMENTS

New developments are taking place every day in the areas of communication and Internet technologies. More powerful and versatile browsers like Microsoft Internet Explorer Release 4 came to the market in October 97. By the end of 1997 an exclusive net-Internet 2^4 will start operating in US for sophisticated academicians who are presently hampered by the slow speed of all and sundry users of present Internet. The Internet 2, the special Internet, is planned to operate at 2.4 gigabytes per second speed with the half of a new routing and switching technologies called Gigapops built on ATM switching system. The spread of Internet 2 facility across the continents cannot be far behind. This is exactly the place where the other recent development in communication called FLAG (Fibre-optic Link Around Globe) will fit it. The project FLAG when completed will be the world's longest cable link measuring about 28,000 km and will provide data transmission speed of 5.3 gbps (gigabytes per second) comparable to futuristic low flying satellite constellations like Iridium and Tele disk. All these developments are going to have direct and indirect impact on the digital library functions and popularity.

8. THE CORE PROJECT

The CORE project is primarily concerned with the conversion of large text and graphics collections to an electronic format and use the same to serve the document delivery needs of scholars in a distributed networking environment. The technological problems of full-text retrieval and delivery are investigated in this project. The five objectives of the CORE project are to

(a) Define a suitable architecture for delivery of full-text information in a distributed networking environment with heterogeneous workstations;

(b) Convert and mount a critical mass of chemistry journal data in a database format suitable for effective retrieval and display;

(c) Study the elements of full-text interface functionality necessary to serve the needs of scholars in a network document delivery environment;

(d) Advance the understanding of suitable document mark up for electronic full-text databases; and

(e) Investigate information retrieval questions germane to the coming era of full-text delivery.

9. DATA ACQUISITION

The data acquisition in CORE project was both by scanning paper and scanning film. It is reported that while paper scanning was relatively straightforward and routine, the film scanning was more complicated and the quality was much lower than that of the former. The users familiar with high quality of the normal printed journals would find the

reprints from the scanned matter needing much improvement.

The study of response time issues confirmed the need as in the case of TULIP experiment the desirability of faster response for browsers. In order to serve as replacements for the paper, online full-text systems should faithfully project the illustrations in the journal articles. The CORE project investigated this aspect using the data provided by American Chemical Society (ACS). The ACS was one of the leaders in computer typesetting of primary journals for more than two decades and their proprietory data markup has been effective for the database maintenance. For the definition of device-independent and system-independent methods of representing text in electronic form the ISO standard (ISO-8879, 1986) widely known as SGML (Standard Generalised Markup Language) is followed.

In the CORE project, the ACS structured files were translated to confirm to this ISO standard. The issues and problems associated with the conversion of database to SMGL, its design and use have been addressed to in the CORE project.

10. LESSONS FROM CORE PROJECT

Some of the lessons learnt from the CORE project about building and use of large digital library systems are summarised below:

(a) Work involved in the creation and management of data even from only one publisher and in only one subject area is enormous and certainly no less than interface work.

(b) In particular, data which involve more than flat ASCII file are tricky.

(c) There is inadequate standardisation of special character sets, procedures for subscripting/superscripting and so on which is essential for user comfort.

(d) In order to find all the special cases for the format conversion software scanning few articles are not enough, but gigabytes of data should be scanned.

(e) The value of pictures to the users has clearly emerged. They frequently look first and mostly at the pictures.

(f) The users are very interested in browsing. The search metaphor is not the only way to look at digital library and further work to cater for users who do not know what they want until they see it is needed.

11. TRANSITION TO DIGITAL LIBRARIES

The pace at which Digital Libraries would evolve depends on many factors. The TULIP report generated some important insights concerning this question. These are summarised below:

(a) At the moment, managing large digital collections locally is harder and more expensive than managing a comparable print collection.

(b) Not everyone is ready for digital collections, nor will they be soon. A couple of technological breakthroughs and more importantly crossing psychological barriers will be needed to make the digital library widely acceptable. Whether one talks of large local stores of data, or regional networked collections, or single remote hosts, the number of academic libraries, really ready to support digital collections, is still small.

(c) Users will only move to electronic publications when they find the content they need in sufficient quantity. The sufficient quantity of needed information is akin to the 'critical mass' required to sustain the chain reaction in the atomic world. Having journals in electronic form and bringing them to the desktop, are necessary but not sufficient conditions for the scholarly user. One must deliver a certain 'critical mass' of needed information to warrant learning a new system or accessing information in a different way.

(d) For the publisher, expanding electronic publishing (internationally) offers challenges very different from paper publishing. Publishers have been delivering a uniform product which required comparatively much less local training or support all over the world. With electronics, this is not the

same; the process is more complicated and requires a different magnitude of involvement from the publisher.

12. ONLINE SEARCH: PROXY SEARCH ENGINES

The popularity of online search through search engines depends on how easily the relevant information can be reached by the users through them. This in turn is influenced by the indexing of the resources and making them available for easy access. In this context, Proxy Search Engines (PSE) which help to reduce network traffic and the number of requests to be answered by primary servers are important[5]. A proxy web server intercepts user's requests for web pages, checks if there is a copy of the required page in the local cache, and responds to the user with this page if it is up-to-date. If there is no such copy or it is outdated, the proxy server retrieves the original copy, places it in the cache and sends it to the user. In the near future, it is likely that online search would be carried not by isolated PSEs but by a network of interacting information brokers exchanging information in a rational way. These brokers are called 'What's Hot' brokers and enable information demand to meet information supply.

13. BENEFITS AND DRAWBACKS OF ELECTRONIC JOURNALS

The most important benefit of electronic journals is the saving in the time involved from submission of the article to its reaching the reader[6]. The electronic journals also save the time of the user normally spent in sifting information by allowing Boolean search of the full-text to browse and read only the selected item. The drawbacks are mostly technological in nature and include low resolution of display units, reduced and restricted images and poor printouts. Also the requirement of availability of the user end gadgets and software are not totally standardised and frozen for effective exploitation of the journals.

14. CONCLUSION

The TULIP and CORE studies have brought out most of the relevant factors that govern the switchover decision from print to electronic media in respect of technical journals. The conclusion of these studies help the institution in decision making on the journals issue. At the extreme, some may choose to continue with the print media while others prefer to switch over entirely to electronic media. Both these have their pitfalls and prices to pay. As a middle course, some institutions may opt to partly switch over to the new media in respect of selected journals and those no longer available in the print media or if available have both print and electronic media copies of the same journal. In conclusion, it is safe to say that emerging information technologies are making a definite foray in the field of library science. Digital library is still evolving, as it is a multi-disciplinary concept involving highly sophisticated hardware and software technologies. The real place of digital libraries in our set up would largely be governed by the demand from users, availability of necessary infrastructure and economics of information transfer from source to end user.

REFERENCES

1. TULIP Final Report, ISBN 0-444-82540-1, Elsevier Science.
2. Entrich, Richard, *et al.* Making a digital library: The contents of the CORE project, *ACM Transactions on Information Systems*, 1997, **15**(2), 103-23.
3. Rajasekhar, TB. What is new in computers? *Digital Libraries: Resonance*, 1997, **4**, 60-73.
4. Bob Woods. Faster and costlier 'Internet 2' coming up. *Computer Today*, 1997, **13**(152), 100.
5. Kosymnin, Arkadi. From bookmark managers to distributed indexing: an evolutionary way to the next generation of search engines. *IEEE Communications Magazine*, 1997, **35**(6), 146-51.

6. Lakshmana Moorthy, SA & Karisiddappa, CR. Electronic publishing: Impact and implications on library and information centres. *In* Digital Libraries Dynamic Store house of Digitized Information, edited by NM Malwade, *et al*. New Age International Limited Publishers, New Delhi, 1996, pp 15-35.