

Digitization: Issues and Problems

Leena Ramrao Bhitre

Abstract

This article is in connection with issues and problems related to digitization. It deals with the aspects: long time preservation, technical issues, cost, digitization management, software selection.

Keywords: Long-time preservation; Technical issues; Cost; Digitization management; Software selection.

Introduction

Today in the age of information technology, it is very important that each and every information is utilized in better way. Digitization of information is the need of today. In that respect we have to look for how it is possible to make use of our knowledge in connection to digitization. There are various issues and problems related to digitization i.e. Storage, Capturing of digital data, Universal access, technological aspects, selection of software etc. Some of the issues and problems are discussed in this article.

What is Digitization?

Creating a digital object by converting a source object in analog form.

What is a Digital Object?

Digital objects are individual entities in a digital collection Any object which can be

digitized can be a digital object

Digitization Process

Identification of the items for the collection

Selection of the 'kind' of content

Choice of formats

Choice of hardware

Choice of software

Selection of medium

Storage and archiving

Management

And in this process following are the issues and problems in digitization

Long-time Preservation

Storage of digital data is the universal issue in digitization process. To preserve digital data for long time various storage media are used world over they are Hyperbooks, CD-ROM, Multimedia etc.

Hyperbooks

Hyperbooks are electronic books defined on the basis of the paperback metaphor. Metaphors have been widely used to facilitate learning of computer system.

A model with which the user is familiar is provided in order to enable users to exploit their existing knowledge of other domains when learning to use a system. The metaphor

Author's Affiliation:*Research Student; Yashwantrao Chavan Maharashtra Open University.

Reprint's request: Miss Leena Ramrao Bhitre. Plot-No. 10 Vasundhara Colony, Nandanvan Colony, Aurangabad.

E-mail:- rlr@rediffmail.com

(Received on 31.12.2012, accepted on 08.05.2013)

which appears to be the most suitable for presenting electronic books in the paperback metaphor stated by several authors.

Advantages of Hyperbooks

1. it can be shared by users in different geographical locations.
2. it is quick, accurate, easy to locate by using searching and browsing mechanisms.
3. AI and language techniques are exploited in fields such as cataloging, indexing and abstracting.

Hyperbook contains text and graphical images. The graphical object is always visible and appears in the text as in a paper book. The feature of the text is that it contains embedded links. i.e. active text is connected to other parts of the book.

CD-ROM

The first product in evolutionary optical storage of information is CD-ROM. It is a permanent optical-based storage device, that in conjunction with an associated drive becomes a powerful peripheral for the PC. The CD-ROM puts multimegabyte permanently stored data bases at user front end of a PC. Its only drawback is that the end user cannot put his own information on it.

Advantages of CD-ROM

1. High storage capacity
2. no head crash
3. low error rate, correctable data
4. long life
5. random access
6. low cost
7. less shelf space.

Multimedia

Multimedia technology integrates text, images, graphics, video, animation, sound or music. It is interactive on the lines of hypertext

and hypermedia. The multimedia has a number of technological barriers for example, extra disk space required to store multimedia formats and broad bandwidth to transmit the text.

Technical Issues

Some of the major technological challenges and issues drawing the attention of workers in this area includes.

1. High band width computer networks supporting efficient multimedia document transfer.
2. open communication protocols
3. information access tools
4. meta databases
5. electronic publishing tools
6. data compression
7. digital storage

Scanning and conversion technologies

1. media integration technologies
2. advanced retrieval, indexing, natural language processing, routing and filtering
3. document description and representation standards
4. inter-operability
5. privacy, authentication and security
6. location independent naming of digital sources.

The magnitude of the ongoing digital libraries projects, their funding and the complex technical issues involved may appear quite daunting and discouraging for library and information specialists in a country like India.

Costing

Costing is a very complex issue as it depends on many variables The variables include- Source object, Hardware, Software, Manpower. digitization cost maybe

categorized as Manpower cost, Process cost, Overheads.

Manpower Costs

Costing: An exercise Manpower costs are worked out on the basis of unit 'time' per task

On an average one person can scan 100-150 pages per day One day is calculated as 8 working hours (effective 6 hours) Thus we say it costs per page approx 3+ mins.

A manpower cost is worked out on the basis of the following assumptions:

A monthly pay of Rs. 3500

A working month consists of 25 days

In other words 2500 pages per person per month Per page cost= $25000 / 3500$

Process Cost

Process cost is based on the hardware and software costs. Two main hardware items are:

Computer, Scanner

Software cost The cost of purchasing the software

Assumptions

A computer costing Rs.45,000. A Scanner costing Rs.25,000. Software costing Rs.10,000.

Total investment cost of Rs.80,000. A life of 3 years, Based on the assumption of 3 years of life. The investment can result in $25 \times 12 \times 3$ working days. This results in $100 \times 25 \times 12 \times 3$ pages of digitization The process cost per page works out to be

Total cost per page = manpower cost + process cost.

We have excluded the overheads such as Electricity costs, Space cost Other indirect costs such as supervision, general housekeeping etc

Digitization Management

The quality of digitization depends on several factors.

The condition of the source document ;Such factors as coloring of the source document.

Font type and size; Person doing the scanning and Un known factors.

Conclusion

For working out to solve the different issues and problems in digitization all the above factors are necessary to be considered. this can be sorted at various level just need is to discussed it in overall world.

References

1. 'Digitization: Workflows, Costing and management' by Dr. Shalini R Urs. www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED295607
2. E-prints in library and information science. <http://eprints.rclis.org/archive/00007385>.
3. Digital library in electronic information Era by Krishan Gopal. 2002.