

EMERGING TECHNOLOGIES IN LIBRARIES & INFORMATION CENTERS

LISC-01403
(Foundation –IV)



Digitization (Definition):

- “Digitization refers to the process of translating a piece of information such as book, sound recording, picture or video into bits. Bits are the fundamental units of information in a computer systems. Turning information into these binary digits is called digitization”. Digitization is one of the hot topics in librarianship today. To build a ‘digital library’ requires that the content of a collection be available electronically.



Digitization Process

- Digitization requires a basic process, which involves different sets of hardware and software technologies at each step. Determining the appropriate technology is directly linked to the anticipated use and purpose of the material being digitized. For digitizing the text and other material, following four methods can be used.
 - (a) Manual data entry scanning;
 - (b) Optical Character recognition (OCR);
 - (c) Excalibur Technologies and pattern recognition technologies;
 - (d) Document Imaging.

a) The simplest method converting an image of a page (or the real page of text into digital text is to enter it manually. This is usually a time consuming method but very useful from the point of view of information retrieval.

b) In the second method, scanners are used to take digital pictures of objects. Scanners can be simple desk top machines or very large and complex systems that process thousands of documents.

c) Another simple digitization process is of OCR i.e. scanning printed pages to build a digital database of text. This process uses OCR (Optical Character Recognition) software, which takes a picture of the page and then turns it into digital text, which can be edited or fully indexed. OCR software must distinguish between black and white areas of text.

d) Excalibur Technologies and Pattern Recognition Technologies are the next generation of OCRs, represented by Pixie, a product being developed by Excalibur Technologies. This software uses a technology called Adaptive Pattern Recognition, which attempts to mimic aspects of the neural patterns of the brain. The software can be taught to recognize variations and relationships in pattern, such as patterns of text rather than readable text. The retrieval of search terms uses what Excalibur calls “fuzzy matching”.

e) Document Imaging, a simple method of capturing text, involves taking an electronic picture of each page of text with the same type of scanner as one would use for OCR. However, the difference is that the images are stored as graphic files rather than text files. A similar technology is used for fax transmission. Each page is stored as one picture. The text on the page cannot be edited or indexed.

Advantages & Disadvantages of Digitization:

Advantages	Disadvantages
<ul style="list-style-type: none">• Quick Delivery Time.• Costs can be favorable compared to in-house costs.• Range of scanning equipment bureau available• Bureau absorbs equipment depreciation and obsolescence cost.	<ul style="list-style-type: none">• Copyright of digital images needs to be assigned to client in the contract, and not returned by the bureau.• Lack of control over scanning environment;• Need to transport material;• Degree of trust involved in Quality Assurance;• Service level agreements needs to be right.

Preserving the Digitized Document

- Rapid developments are taking place in both the hardware and software involved in digitization. This means that the present technology will soon be supplemented by newer technology. The stability of current systems and the digitized products is thus questioned. Systematic efforts will be needed to ensure that what we digitize today is not slide into obsolescence tomorrow. Migration to newer systems and media and regular refreshment are two possible solutions. However. They are both costly and time consuming; they also carry a risk of data loss.

Types of Scanners:

- Flatbed scanners



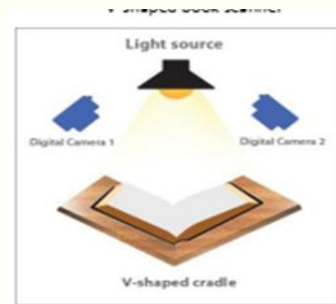
- Drum Scanners



- Sheetfeed Scanners



- Book Scanners



- Microfilm Scanners



- Digital Cameras



Reference

- Selvakumar, A. (2002). Acquisition and preservation of digital library resources. Retrieved from: <http://hdl.handle.net/10603/61632>