Digitization Tools and Techniques: An Overview
Senthilkumar K.R.

Abstract

The study provides an outlook on the various digitization tools and techniques for library materials - manuscripts, printed documents and archiving materials. The term "digitization" was coined in the late 1990s. Many libraries still do not make use of this useful technique at the root levels for the benefit of the research user community. To overcome this hurdle, here the researchers have presented the digital initiation in India at national level in addition to the relevant tools, software, techniques, file formats, storage and other techniques.

Keywords: *Digitization; digital preservation; digitization initiation*

Introduction

Digitization provides a solution to manuscripts problems such as conservation, preservation, accessibility, uniqueness, access, ICT (Information Communication Technology), as also demand and relevance to funders and space: the first and most vital reason being that the manuscripts start deteriorating after a certain period of time at a rapid rate. Current ICT techniques using open source platforms offer to create digital libraries to meet the needs of knowledge science communities of biological, environmental, medical, engineering sciences and humanities. Two Decades ago digitization was a novel concept when there was a need for more infrastructure, staff skill, fund…; so it was found to be a daunting task. However, the trend now slowly changed in a decade and presently a student of sixth standard is using electronic devices like smart phone, tablets and laptops. Nowadays almost each and every family from lower middle class student to research scholar level has started using smart phones and laptops for all their needs. This situation of digital technology has simplified sharing of information, video clippings, and pictures instantly to unknown persons from any part of the world. Almost we are living in a digital world where without technology leading a life is not so easy. Digitization is no longer a mere buzzword; it has become hard reality in the library world. There are three types of e-devices to digitize; "using scanner, digital camera and the latest smartphone". For transmitting any piece of information-either text, audio or video we have to select the proper digital device. If it is a paper, manuscript or document in a good condition, free from legal issues then a smartphone could be used to disseminate information instantly with the help of apps. For a research community we librarians have to present digital information using Digital Camera or Scanner in a simple way using a structured digitizing method.

Review of Literature

Kaur (2015) discussed the need for digitization of manuscripts with reference to Punjab university manuscripts and Punjab Digital libraries which started in the year 2004. Sageer and Francis (2015) identified different ways and methods for developing digital libraries for endangered library materials and the users attitude about the method of usage of manuscripts. Sarasvathy (2014) drew attention to the role of Indira Gandhi National Centre for Arts (IGNCA) in digitization of library materials, particularly manuscripts. For digitization IGNCA used the standards of UNESCO guidelines and also Metadata for all the materials. Earlier IGNCA had microfilmed

2.5 lakh manuscripts with 20,600 rolls out of which 17087 had been digitized and 13803 rolls duplicated. Rafiq and Ameen (2013)4 provided a survey of fifty-five university libraries and their activities on digitization in Pakistan. The results found that the digitization activities were still at preliminary level. Londhe et al (2011) focused mainly on technical requirements and processing method for digitization of manuscripts adopted in Jayakar library, University of Pune. This work explained the various technical stages to adopt for digitization of manuscripts and the metadata generation. The study found the importance of software for image capturing and processing. In South Asia the largest digitization initiative was launched in India with the biggest coverage of twenty-one scanning centres, inaugurated by Dr. A.P.J. Abdul Kalam. It is part of a Million Book Project. Along with this; digitization of thesis and dissertations, Indian Journals, digitization activities of NMM & IGNCA were evaluated by Anup Kumar Das in his dissertation in 2008. He observed that all these mega programs should be evaluated at regular intervals; training given and awareness must be provided to Indian researchers and scholars. He critically appraised NMM and IGNCA in conserving documentary heritage collection. The enormous output of research findings in Indian journals covering all the fields of knowledge should be made available at the international level with strong visibility. Kaur (2007) stated that preservation of digital materials for the benefit of the present and future generations is an urgent issue to address. The author explores various strategies and methodologies for preserving digital heritage materials; discusses issues concerning Digital Preservation (DP); enjoins upon the National libraries to take the responsibility for DP, with the cooperation of several stakeholders the Government, Publishers and Information Technology (IT) industry; gives an overview of initiatives taken by the National libraries world-wide, for preserving digital heritage materials; and discusses the Indian scenario. It also focuses on the pressing need for the national library, state libraries, museums and other libraries to work together more closely, to successfully organize digital archiving in the 21st century.

Objectives

Provide necessary technology and training for students and teachers to facilitate pleasure reading and for informational text. Provide reading materials to students and staff. Utilize specific programs/applications for supplemental instruction in reading, math, science, and social studies. Provide supporting accessories for technology for staff and students.

Results

Digitization Initiation in India

The National Digital Library of India (NDLI) is all set to collaborate with several top digital libraries from across the world. Giving this information in a written reply in Rajya Sabha, Minister of State (HRD), Dr. Satya Pal Singh said that NDLI has entered into a Memorandum of Understanding with British Library for integrating contents of its "Two Centuries of Indian Print Project". There have been many digital library initiatives in the world including in India. In India, the digital library initiatives have been sponsored by several Ministries, notably In Previous   Digital Library of India (DLI) Now National Digital Library of India (NDLI),   National

Mission for Manuscripts (NMM) and Indira Gandhi National Centre for Arts (IGNCA)

Digital Library of India (NDLI)

This initiative of the Government of India was started in early 2000 with an aim of digitizing (DSpace) 1 million rare books of Indian languages for the benefit of users of all over the world. he National Digital Library of India has opened up its 3.5 crore-strong digital contents, such as e-books, questions papers and solutions, lecture materials and thesis among others to help students study at home.16 DLI can be accessed through https://ndl.iitkgp.ac.in National Digital Library of India (NDL) hosted by IIT Kharagpur, sponsored by Ministry of Human Resource Development and established in 2015. Currently NDL has 7.2 million books in 70 different languages of 60 types of learning resources. The subject coverage includes philosophy and religion, literature, agriculture, fine arts, NCERT text books, theses from IITs, IIS, IIMs and Publications from the laboratories of CSIR and ICAR. There are over 917,026 entries found on self help groups in 0.5953 secs. This can be reached through https://ndl.iitkgp.ac.in/. NMM: National Mission for Manuscripts (2003) guidelines for digitization of library material of manuscripts, books, photos, maps. etc is available in both Hindi and English. It raises certain points like "why digitize?" and also deals with the process of digitization. More importantly guidelines are provided for handling/treatment of materials, selection of equipment, image capture, digital conversion, quality control, storage and management of digitized images and also on calculation of the cost of the proposed project. Indira Gandhi National Centre for Arts (IGNCA) that acts as the nodal agency of the National Mission for Manuscripts (NMM). The IGNCA has initiated many programs with its expertise in manuscript libraries through creating Kalasampada, multimedia projects, CAT-CAT database, catalogue of microfilmed manuscripts and digitization.

Need for Study

Objectives

- To collect, organize & collate print & digital information & disseminate at the point of care and for future use.

- To provide seamless access to information.

- To act as gateway to digital and electronic information.

- To develop in to a single access point library.

Methodology

Digital technologies have driven researchers to revisit not only the themes and questions of their research but also their research methodologies (Tinati et al., 2014), often leading to the creation of novel methods of research (Fielding, Lee and Blank, 2008; Hine, 2005; Hughes, 2012; Johns, Chen and Hall, 2004; Jones, 1999; Markham and Baym, 2009; Roberts et al., 2013; Salmons, 2010, 2012; Sappleton 2013). Early on, scholars acknowledged that tools, applications, content formatting and other Internet affordances not only repackage existing research methods, adapting them to the Internet, but also suggest completely new and often innovative

approaches to and methods of research (Jones, 1999; Sosnoski, 1999; Sudweeks and Simoff, 1999). Schneider and Foot (2004) suggested that the analysis of websites and their content requires new analytical tools, since increasingly complex web applications alter traditional relationships between media form and content (Schneider and Foot, 2004, p. 116). More recently, Rogers (2013) separated digital methods from virtual methods. He defined virtual methods as imported and migrated from offline research and as adapted to the online environment wherein they are employed (e.g., online surveys, e-interviews). On the other hand, for Rogers, digital methods are those 'native' to the medium (i.e., digital technologies), such as hyperlink analysis, web engine diagnostics, web archival research, web content analysis and social media research. Rogers considers digital methods appropriate for the analysis of digital data (e.g., hyperlinks, web content) and distinguishes them from digitised or virtual methods, aiming to draw researchers' attention to the 'medium' so as to 'reorient Internet research to consider the Internet as a source of data, method, and technique' (2013, p. 27). Others disagree with such a separation, considering it restrictive and arguing that it misses the full range of prospects and possibilities in digital methods (Roberts et al., 2013, p. 6).

Finding

The purpose of this study is to articulate and present in a simple way the structured digitizing method and to eliminate the bewilderment of the technical process of digitizing library materials such as manuscripts, printed documents. Also, the study analyses three e devices–smartphones, scanners and digital cameras are for digitizing and disseminating information for a research community.

Tools for Digitization

Digitization workstation consists of a stand-alone system where most or all the work is done on the same workstation or as a part of a network of workstations with imaging work being distributed and shared amongst various workstations. A typical digitizing station/project could consist of the following:

Micro computer with the latest configuration of Intel i5

Software

- Scanner/Digital camera/smartphone

- Storage system –internal, external

- Network

- High Power UPS

- Types of Images: BMP, TIF, JPEG, GIF and PNG

Most image formats use pixels per inch or dots per inch. The dpi chosen when scanning is entirely subjective. While scanning for preservation purposes a minimum of 300 dpi and even perhaps 600 dpi is required for a digital camera of 33 megapixels. The BMP (or bitmap) file format is supported by the Windows Operating System. The TIFF 9 or tagged image file format) is the format of choice for archival purposes. Indeed most

digitization projects would choose TIFF as the first file format to transfer analogue content into digital. The JPEG format is very popular mainly because it allows images to be delivered in small file sizes. The GIF, Graphics Interchange Format was the image format of choice in the early stage of the internet, before JPEGs. PNG portable network graphics format that uses geometric shapes is preferred rather than pixels

Useful Software

Software for manipulating and saving the digital image is absolutely essential. Paint program is a standard and common program for simple image manipulation. Apart from this there are some open source software and commercial software for using scanners– Adobe Photoshop, GIMP-Gnu Image Manipulation Program, IrfanView or ViewXn. When using a digital SLR camera, the software is usually provided as a package which will automatically install into the system. ViewNx is useful open source software for conversions of Raw files to JPEG or from TIFF to JPEG. Adobe Photoshop is often used for cropping and sizing JPEG images. Also it is useful for combining cropped JPEG images and to save as a PDF and to reduce file size to be saved from the scanned PDF. This option is better than reducing file size or reducing through print option/Adobe pdf because search ability of words etc. is good and the file size is also reduced. Abby Finereader is used for OCR-JPG images and lso its served OCR images to save as searchable PDF and to convert MS Word files

Digitization Techniques

The basic process of digitization is fairly simple through a wide range of sophisticated techniques and tools. Essentially, a digital image is composed of a grid of pixels (picture elements) arranged according to a set ratio of rows and columns. Each pixel represents a very small portion of the image, and is allocated a tonal value; namely, black, white or a particular colour or shade of gray. These tonal values are digitally represented in binary code (zeros or ones). So a digital image is actually a grid made up of zeros and ones. The binary digits for each pixel are called bits and are stored in a sequence. When the digital image is displayed on a computer screen or sent to a printer, the bits are interpreted and read by the computer to produce a physical representation of the original material.

Before beginning the digitization of manuscripts, the first step is to decide on digital devices such as a digital camera or a flatbed/moveable scanner. Second step is to survey the collection focusing on the type and size of the material to digitise and where the digitisation will take place. Third step is to contact other institutions that have digitised similar material and would have useful advice to share. Both digital photography and scanning are complex techniques and it is important that a person with the appropriate skills and ideally with an appropriate qualification is involved in the digitization work.

Scanners

Capturing a digital image is known as scanning. Image resolution i.e. the number of pixels in a row and colour depths, determines the quality of the scanning. There are several types of scanners, but it is preferable to save images in uncompressed format to enable later conversion to JPEG. In general an overhead scanner offers the most flexibility and decreases the risk of harming manuscripts as there is no direct contact, but these are very expensive and need technical expertise. The various types of scanners include: overhead or planetary, moving head, Sheet-feed, flatbed, specialized 'V' shaped book scanners, drum, robotic book and large format scanners. Among all these scanners, flatbed scanners and moving head scanners are the most familiar means of digitization. This type of scanner is comparatively inexpensive but small and medium size flatbed scanners will not be useful for palm leaf. Manuscripts as they are of various sizes of folio and many are not flat against the glass; the final and most important aspect is the use of harsh white light, possibly damaging the delicate palm leaf manuscripts.

Digital SLR Cameras

Digital cameras are designed specifically for making computer images. As a result digital cameras retain the strengths of scanners, can work without damaging manuscripts and can work with different shapes and sizes of manuscripts under various required lights. In addition we can make images of very high resolution and improve colour qualities of images.

Digital camera can save the images in either uncompressed TIFF format, or in uncompressed RAW format and at a later stage should be converted into JPEG. A tripod or a copy stand is needed to mount the camera downwards directly over the manuscript. This will prevent any distortion of the item being copied. An industry approved colour chart or grayscale reference card is essential in order to check the accuracy of the colour captured in photographs. One needs to plan for a necessary light source based on daylight colour and temperature. To prevent camera shake, one should plan for a remote shutter release or use tethered photography, where a link between the camera and the computer is established.

Smartphones

Undoubtedly, smart phones have become a part of our daily lives. By taking advantage of such massive use, we can use these miniature pocket computers to revolutionize research in digitization. New technology in smart phones offers high temporal and spatial resolution with built-in millisecond timing of stimuli display and touch screen responses. Smartphones are tools that are portable, easy to use, multimedia-enabled and identical in every country and for each user, with ready Internet transfer of collected data. These properties render it an instrument ideally adapted to studying cognitive functions. A smart phone allows us to dramatically increase the amount of data to collect, digitize and process without sacrificing the time of users.

Digital SLR Camera Vs Scanner

There are many advantages and disadvantages when there are two types of an electronic device, but we have

to select the correct one according to the purpose and objective. In this case there are two ways to digitize manuscripts using 'scanners' and 'digital cameras'. The Table 3.2 provides a list of important objectives and devices with the advantages and disadvantages.

Table-1 Digital Camera Vs Scanner

Ser. No	Objectives	Digital Camera	Scanner
1	Display	Good Quality camera with high pixel	Should have more dpi
2	Usefulness	Multipurpose	Single purpose
3	To digitize mss	possible with wide coverage lens	Need bigger sizes of scanner
4	To digitize photographs	Not possible	Possible
5	Different Aspects of objects	Possible to see all aspects-3D	Not possible-only flat items
6	Fragile Items	Fragile object can be handled in the place	Fragile object must be placed in scanner
7	Being Made use	Everybody can handle	Very rare-Not everyone can use
8	Speed	Fast - immediately	It will take several minutes
9	Size - characteristics	Easy - Very compact	Bigger in size, needs power cord
10	Result	Not very good quality	Better result
11	Close up	Required costly lens	No problem
12	Blurry corners	May be object in close-ups	No problem
13	Movement	Needs Tripod	No problem
14	Alignment	It should be pointed each time	No problem
15	Light	Needs Extra apparatus	No problem
16	Resolution	Needs higher megapixels	Measured in dpi/600 dpi equals to 34 pixel
17	Cost	Inexpensive	Expensive

Optical Character Recognition (OCR)

A scanned document is nothing more than a picture of a printed page. It cannot be edited or manipulated or managed based on the contents. Another technology which is involved in digitization is Optical Character Recognition (OCR). OCR is the usual process by which a page image is transformed into a text file. The purpose of the whole OCR process is to recognize the letters, words, and symbols printed on a page. Presently, there is a wide range of commercial OCR software in use. There is no proven OCR software to handle Indian language texts. Today, if Indian language materials have to be digitized, there are two options- maintain the files as digital images or manually key in the material.

Metadata

Digitized product that is to be put up on the Web needs information that makes it possible to be located. One of the principal challenges is to determine what information is essential for describing an electronic product. The Dublin Core (http://purl.oclc.org/metadata/dublin_core/) and other special initiatives for structuring and standardizing descriptive data propose to combine information about the technical characteristics of digital files (how they were created), their location, and a summary of their contents. The resulting information is known as "Metadata" and is located in the header of a tagged document. Their function is to provide users with a standardized means for intellectual access to digitized materials.

Digital Storage

The biggest concerns after digitizing manuscripts are where to store and how to manage an item. When digitizing, image file sizes are normally large formats of TIFF or BMP. It is essential to ensure that the digitizing station has the following: A connection to a network, where images can be stored in a server A removable HDD minimum of 1 TB A DVD burner

Conclusion

Libraries of different statuses have been working on this daunting task for more than a decade but only Research & Development libraries and special libraries have reached this level. They have created many digital initiatives and guidelines for solving technical issues. For this development for all the libraries, librarians should come forward to start using the latest digital resources and their applications to the readers.

References

http://ijrras.com/?media_dl=5981&edmc=5981

Das, Anup Kumar (2008). An evaluative study of some selected libraries in India undergoing the process of digitization. (Doctoral Dissertation) Retrieved from Eprints

Dr. Senthilkumar kr 2020 Comparison of E- Resources with their Usage Statistics in Southern Region, Library Philosophy and Practice (e-journal) https://digitalcommons.unl.edu/libphilprac/3270/

Rafiq, Muhammad & Ameen, Kanwal.(2013). Digitization in university libraries of Pakistan. International Digital library perspectives vol. 29(1), 37-46.

Senthil Kumar, K. R., Recent Trends of ICT Services and the Present Scenario of Some Selected Engineering College Libraries in Coimbatore District, Tamilnadu: A Study (February 2017). Asian Journal of Applied Science and Technology (AJAST), Volume 1

Senthil Kumar, K., Recent Trends of ICT Services and the Present Scenario of Some Selected Engineering College Libraries in Coimbatore District, Tamilnadu: A Study (February 2017). Asian Journal of Applied Science and Technology (AJAST), Volume 1, Issue 1, Pages 199-202, February 2017 . Available at SSRN: https://ssrn.com/abstract=2928955

Sarasvathy, P. (2014). Role of IGNCA in digital preservation of manuscripts: an over view. Retrieved 15th July 2016 from: http://eprints.uni-mysore.ac.in/16671/

A Scientometric Study On Niscair Journal Of Annals Of Library And Information Studies From 1999 To 2013 K Senthilkumar – 2015

Gyankosh- The Journal of Library and Information Management Year : 2013, Volume : 4, Issue : 1 First page : (89) Last page : (93) Print ISSN : 2229-4023. Online ISSN : 2249-3182. Free web page: A tool on usage of academic library development Kumar KR. Senthil

Londhe, Nagesh L., Desale, Sanjay K. & Patil, Suresh K. (2011). Development of a digital library of manuscripts. Program, 45 (2), 135-148

Narenthiran, R.(2012). The Digitization of palm leaf manuscripts. SALIS 2012

Ramesh C Gaur (2006). Digital Libraries in India with special reference to experiences at IGNCA, New Delhi, India.   Retrieved   from https://www.researchgate.net/publication/242749586_Digital_Libraries_in_India_with_special_reference_to _experiences_at_IGNCA_New_Delhi_India

kr, senthil, "User pattern of Libraries by students of Government colleges in Tamilnadu : A Study" (2019). Library Philosophy and Practice (e-journal). 2788. https://digitalcommons.unl.edu/libphilprac/2788

Kaur, A. (2007). Preservation of digital heritage materials. SRELS Journal of Information Management, 44(4), 385-394

Kaur, Harpreet (2015). Digital preservation of manuscripts: an Indian perspective with special reference to Punjab. 4h IEEE

https://www.hindustantimes.com/education/national-digital-library-opens-up-its-digital-contents-builds-covid-19-research-repository/story-36M8asvDd9GSMiKXTMDN4H.html

https://www.bhphotovideo.com/explora/photography/tips-and-solutions/scanning-without-a-scanner-digitizing-your-film-with-a-dslr