

---

---

## DIGITISATION OF NEWSPAPER: AN EASY ACCESS TO INFORMATION

***Keisham Sangeeta***

### **Abstract**

*Today in the digital world, digitization of library materials has advanced rapidly. Digitization solves the traditional library problems like conservation, preservation, storage and space. It provides wider access to information to different users. The present paper discusses the importance of Newspaper Digitization (ND) in providing effective library services. It also highlights some of the National Libraries, which undertake ND project successfully.*

**Keywords:** Newspaper Digitization; Digitization

### **1. Introduction**

The impact of current information revolution and information technology have changed library totally in a new shape, as one of the important media for communication of knowledge. Using modern ITs, the libraries of today have changed the pattern/ methods of collection-storage-dissemination-retrieval of knowledge/information. The traditional concept of libraries, which was completely based on print media, is changing towards digital/electronic form. Thus the role and function of a library become relevant in providing appropriate element of information and within a given period of time to the rightful user.

Today, in the digital environment digitization has become one of the key activities in the libraries, which is an excellent way of providing access to library materials. Through digitization, libraries are able to access, preserve, maintain the digital collections easily and provide effective services to the user requirements. Newspaper-services is considered as one of the important services provided by the libraries, which provide up to date and timely information to the users. Newspapers are the major channel of communication and main sources of information on current topics on all aspects of a country's social, economic, political and cultural life as well as pursuit latest trends on different subjects or topics. For decades, even hundreds of years after their publication, researches of all kinds turn to newspapers for information relating to their needs. However, it is not easy to provide access to newspaper texts because of the very poor nature of the paper the newspaper was printed on. In order to get ease of access to its news content, there is a great need for digitization of newspaper. Digitization has increased the means to make newspapers more readily available by adding advanced search features to access texts. Because of their heavy use, newspapers are seen as an important part in the Libraries digitization programmes.

---

---

## 2. Digitization

A large volume of information in this universe is in non-electronic or analog form. This information needs to be digitized before it can be handled by digital technologies. For example, printed information may be digitized and stored inside a computer using a scanner or a digital camera that is attached to the computer. The computer controls these devices, acquires and stores the digital images produced by them. In short, digitization refers to the process of converting an item such as printed text, manuscript, image or sound recording, picture or video into digital form.

## 3. Why Newspaper Digitization (ND)?

It is often said that digital information is transforming the way we learn, the way we communicate, even the way we think. One of the most important qualities of information in digital form is that by its very nature it is not fixed in the way that texts printed on a paper are. Digital texts are neither final nor finite, and are fixed neither in essence nor in form except when a hard copy is printed out.

The main purposes for digitization of newspapers are identified as:

- to promote wider access to information to a large number of users;
- to preserve and conserve the national and cultural heritage through digitization of back volume of newspapers, which constitute the raw materials of history;
- to provide use-friendly interface to access the news content;
- to search across the different newspaper titles and to draw together materials relating to a wide range of research and learning topics;
- to access to all or part of the content of different newspapers on the World Wide Web.

## 4. Factors and problems for ND

In the recent years, a practical distinction has emerged between current newspapers, which are "born digital", and those created in the letterpress era, which requires conversion to digital formats using scanning equipment. It has been possible only in the last ten years for libraries to consider alternatives to consulting the microfilm of the back runs of newspapers. Nowadays, it is possible to have microfilm scanned digitally, with the digital images capable of further manipulation by software, to permit a good degree of readability, with the texts also being searchable. Moreover, it has also become possible to have information indexed from all kinds of sources. This has been due to:

- rapid development of software to link computers, together with sophisticated software to provide information to users;
- increase in sophistication for the indexing of all kinds of digital information;
- development of hardware and software to create pictures digitally, store them, and transmit them to other computers;
- arrival of the World Wide Web.

Texts in manuscripts or in letterpress, such as newspapers have been more difficult for software to work with, as the variables of character recognition are so much greater. Newspapers are difficult for scanning and optical character recognition (OCR) because:

- 
- text characters are over or under-inked;
  - text characters are worn and not print properly;
  - paper may be creased, resulting in loss of text;
  - the spine curvature of bound volumes of newspapers makes text difficult for scanning machines to recognize, and also for OCR;
  - significant variation in font size within individual issues of a newspaper;
  - layout of text into columns is not always regular, and the layout could vary significantly over time.

## **5. Technical requirement for ND**

The following technical points need to be considered in securing full public availability for the newspaper texts:

- proper preparation of the original newspapers, in sequence, for either scanning or microfilming;
- high quality microfilming of the newspapers;
- quality scanning of the microfilm or the originals;
- quality assurance of the scanned images;
- high quality Optical Character Recognition of the scanned images;
- some form of division of the newspaper pages into zones or segments;
- use of software to index individual articles by subject words;
- use of software to enable searching of the entire set of texts;
- use of software to provide a host for the data, and to permit users to access it via the Internet;
- ensuring that the data is preserved for future use.

## **6. Trends in ND**

Today, a number of National Libraries undertake large project for ND to provide wider access to its newspaper collections. Some of the National Libraries, which successfully undertake ND programme are highlighted below:

### **6.1 British Library Newspaper Pilot**

In the first half of 2001, the British Library and other partners collaborated to produce a prototype system for the digitization, indexing, and presentation of historic newspapers from the newspaper collections of the British Library. The overall goal of the project was to allow for online accessibility to the library's historic newspapers, and in particular, those in the nineteenth century. Eighteen reels of duplicate negative microfilm were selected, containing newspaper issues which focused upon British national events. Full runs of the same few months in each year were chosen for each newspaper title. The intention was for users to look at a variety of perspectives on the same events or issues. Microfilms of arrange of ages or states were duplicated, the aim being to test the technical issues attendant upon the digitization of microfilm as well as the acceptability of the content.

---

The system and the data were created and have been live since 2001. Usage is around 25,000 user sessions per month, and this is without major promotion and publicity. It is estimated that some 20,000 pages of newspaper pages were processed in two months, containing some 500,000 articles and advertisements.

## **6.2 British Newspapers 1800-1900**

In 2004, the British Library secured funding from the Joint Information Systems Committee, JISC, as part of its digitization programme to film and scan up to 2 million pages of nineteenth century newspapers. The project, reflecting the British Library's vision of helping people to advance knowledge to enrich lives, offers users in higher education and further education. The aim of the project is to select and to digitize complete runs of UK newspaper titles that were published between 1800 and 1900. It is proposed to select a mixture of UK national, regional and local newspaper titles, whether microfilmed or not, which reflect the social and political development of the times in which they were published. The library's strategy for the successful completion of the project comprises in-house activities for material selection, conservation and microfilming undertaken by library staff and the procurement of a solution through competitive tender for the digitization.

## **6.3 Library of congress (LC)**

The Library of Congress newspaper and current periodical room has an online guide to its newspaper collections. The library also provides information on the United States newspaper program, probably one of the largest preservation projects undertaken to microfilm older U.S newspaper titles, state by state, since the early 1980s. It has also information regarding the National Digital Newspaper Program (NDNP). The National Endowment for the Humanities (NEH) is planning to facilitate digitization by funding the NDNP. Ultimately, over a period of approximately 20 years, NDNP will create a national digital resource of historically significant newspapers from all the states and U.S territories published between 1836 and 1922. This searchable database will be permanently maintained at the LC and be freely accessible via the Internet.

## **6.4 National Library of New Zealand**

The 'Papers Past' Project shows selected nineteenth century newspapers and periodicals, with digital images of over 4000,000 pages from 25 publications being available.

## **6.5 National Library of Canada**

The library has created an electronic collection. It intends to create digital images for newspaper texts as part of this initiative.

## **6.6 National Library of France**

The 'Gallica 2000' project has created a substantial body of texts in digital form available for searching. Under this project, the library is likely to create full digital images of French newspapers and put on the site.

---

---

### **6.7 National Library of Austria**

ANNO is the library's virtual newspaper reading room. It offers access to a growing number of digitized newspapers.

### **6.8 National Library of the Netherlands**

Digitization from Microfilm from Dutch newspapers 1910-1919 project is conducted in the framework of the Dutch national programme for preservation of library material, *Metamorfoze*, which is an initiative of the Netherlands Ministry of Education, Culture and Science. The selected part of the collection contains historic newspapers from the period 1910-1919.

### **6.9 TIDEN- A Nordic Digital Newspaper Library**

TIDEN is a collaborative project of the Helsinki University Library, the National Library of Sweden, Statsbiblioteket-the Danish State and University Library in Aarhus, and the National Library of Norway for the digitization of newspapers on microfilm.

## **7. Conclusion**

Due to digitization, radical changes are taking place in information dissemination. It is seen that a significant expansion of activity in making copies of older runs of newspapers are available via the Internet. Moreover, it is also likely that there will be far greater ease of access to texts available as a result of the application of sophisticated software. This will significantly increase the availability of researches to search quickly through huge quantities of printed newspapers, which will be vast improvement over the former manual process. As such, librarians in digital era need to build up infrastructures and soft facilities to meet organization of collections and services, ICT applications, knowledge management, web based applications as well as enhance core competencies to meet the challenges and threats in digital era.

## **References**

1. Chaurasia, N.K. (2006). Digitisation of Library Materials. *University News*, 44(30), 15-18.
2. Prasad, R. (2006). Preservation of Library Resources in Digital Environment. *University News*, 44(11), 6-10.
3. King, Edmund. Digitization of Newspapers at the British Library.
4. [http:// www.google.com](http://www.google.com)