NEWSPAPER DIGITAL ARCHIVES NEED IN NETWORKED WORLD

D CHANDRAN

Absctract

Newspaper has been considered as an important source of information. The generation of current varieties of information in large quantity makes it distinct from the information media. A large number of newspapers published throughout the world with mountain of information create severe problem for the information professionals regarding the storage and retrieval of such information. Considering the specialties and importance of newspaper as information media, the issue is of special concern for the information professionals. Several newspaper information retrieval systems and projects in different countries, mostly in developed world, are found to cater the retirements of users on a regular basis, besides most of the developed countries have established their own newspaper information retrieval systems. Newspapers are particularly both to preserve and to access they are large in format, prolific in output. A number of techniques were developed to preserve and retrieve the information available in newspapers. Among them, Digital preservation is the recent one. This led to create its own archive for newspapers. The author presented an overview about newspaper archive initiatives worldwide.

Keywords : Digital Archives/ Newspaper

1. Introduction

Since long back newspaper has been considered as an important source of information. The generation of current varieties of information large quantity makes it distinct from other information media. A large number of newspapers published throughout the world with mountain of information create severe problem for the information professionals regarding the storage and retrieval of such information. Considering the specialties and importance of newspaper as information media, the issue of concern for the information professionals. Several newspaper information retrieval systems and projects in different countries are to cater the requirements of users on a regular basis, most of the developed countries have established their own newspaper information retrieval systems.

Newspapers are particularly both to preserve and to access they are large in format and prolific in output. There is however, no other medium in our history that records

---

every aspect of human life over the three hundred years. For users of many kinds, newspapers represent a source of information that is of monumental importance, and they are unparalleled as a primary source media. Traditional manual and printed systems for archiving and retrieval of newspaper clippings get some inherent problems; especially, in terms of speed, accuracy, diligence and some othe intricacies of modern retrieval techniques. A number of techniques were developed to preserve and retrieve the information available in newspapers. Among them, Digital preservation is the recent one. This led to create its own archive for newspapers

A newspaper is a publication containing news, information and advertising, usually printed on low-cost paper called newsprint. It may be general or special interest, most often published daily or weekly. Recent developments on the internet are posing major challenges to its business model, however. Paid circulation is declining in most countries, and advertising revenue, which makes up the bulk of a newspaper's income, is shifting from print to online, resulting in a general decline in newspaper profits. This has led to some predictions that newspapers' role in society will shrink or even disappear, although historically, new media technologies such as radio and television never supplanted print media.

Newspapers are fundamental sources of current information on contemporary issues and are essential for building current awareness and keeping society at large, well informed. Newspapers are also fundamental for teaching and research in journalism, the social sciences, management, arts and education. Newspapers contain humongous information in the form of articles, features, profiles, case studies, market & industry reports, corporate indices, developments in science & technology and a whole gamut of sociological, economic, political developments in and around the world. The usefulness and scope of a Newspaper Archive for research is limitless. As newspapers chronicle events as they unfold – no historical study is complete without researching a Newspaper Archive. Newspaper Archive is primary literature. For decades microfilm has been the only way to store and preserve newspaper collections around the world. DigitalTechnology is tool and the privilege lies in human hands to select and use his tools wisely for his own optimum benefit. Digital Archiving has come exist to preserve the physical newspaper to digital paper or e-paper. The advancement in technology is led to the Digital Archiving process through digital preservation. .

2   Digital preservation

Digital preservation refers to the management of digital information over time. Unlike the preservation of paper or microfilm, the preservation of digital information demands ongoing attention. This constant input of effort, time, and money to handle rapid technological and organisational advance is considered the main stumbling block for preserving digital information beyond a couple of years. Indeed, while we are still able to read our written heritage from several thousand years ago, the digital information created merely a decade ago is in serious danger of being lost.

Digital preservation can therefore be seen as the set of processes and activities that ensure the continued access to information and all kinds of records, scientific and cultural heritage existing in digital formats.

Digital preservation is defined as: long-term, error-free storage of digital information, with means for retrieval and interpretation, for all the time span that the information is required for. "Retrieval" means obtaining needed digital files from the long-term, error-free digital storage, without possibility of corrupting the continued error-free storage of the digital files. "Interpretation" means that the retrieved digital files, files that, for example, are of texts, charts, images or sounds, are decoded and transformed into usable representations. This is often interpreted as "rendering", i.e. making it available for a human to access. However, in many cases it will mean able to be processed by computational means.

2.1 Archive

An archive refers to a collection of records, and also refers to the location in which these records are kept. Archives are made up of records which have been created during the course of an individual or organization's life. In general an archive consists of records which have been selected for permanent or long-term preservation. Records, which may be in any media, are normally unpublished, unlike books and other publications. Archives are distinct from libraries insofar as archives hold records which are unique. Archives can be described as holding information "by-products" of activities, while libraries hold specifically authored information "products". The word 'archives' is the correct terminology, whereas 'archive' as a noun or a verb is related to computer science.

Archivists would consider the Internet Archive to be a digital library as opposed to an archive as the Internet Archive's materials lack the characteristics of archives.

2.2 Internet Archive

The Internet Archive (IA) is a non-profit organization dedicated to maintaining an archive of Web and multimedia resources. The Archive makes the collections available at no cost to researchers, historians, and scholars.

2.3 Digtial archive

Digital Archive offers real-world solutions for the challenges of archiving and preservation in the virtual world. This flexible system allows you to archive assets in two ways.

a.   Use Web archiving for item-by-item harvesting and submission of Web pages and Web-based documents,

b.   Batch archiving to submit your collections on various storage media for ingest and automated metadata creation.

2.4 Benefits of Digital Archive

■ **Security**. Keep your materials safe, secure and accessible over time

■ **Reliability**. The Digital Archive is a dependable and trusted, standards-based third-party repository

■ **Focus**. The Digital Archive lets you focus on your content, not storage space and technology trends.

■ **"Born Digital"** that lasts. Archive live documents from the Web or collections that exist only on your own servers, CDs, or tapes.

■ **Metadata expertise**. Informtion Centres applies its vast experience with bibliographic metadata to help you capture extensive preservation metadata.

■ **Peace of mind**. No more worries about the whereabouts of collections scattered across multiple CDs or tapes.

2.5 Universal features

■ **Configurable access.** Allow the whole world to see your resources or limit their availability to a single IP address or small range of users.

■ **Administrative access.** View any or all items in your collection for quality assurance purposes easily at any time with the Administration Module.

2.6 Web archiving features

■ **Connexion integration.** Save on training time because the archiving process is integrated with Connexion.

■ **Web Harvester.** Locate and harvest all, or just specific parts, of Web sites you want to archive and preserve.

■ **Automatic metadata retrieval.** The Web Harvester automatically grabs existing metadata from the pages being archived so you don't have to input it yourself.

■ **Flexible storage options**. Use OCLC as your repository or your own local archive.

3. Benefits for libraries

The digital archive for newspaper service is intended to help libraries improve workflow, lower cost and provide value-added services within a union catalog approach. The pilot project will help us determine if these benefits are achievable in a production service. It will enable catalogers and other metadata professionals to make digital content widely accessible and usable through WorldCat-based services.

4. Online and Electronic Newspaper

Most modern newspapers are run online as well as being run as hard-copy, but with the development of the internet the market the hard-copy newspaper are being pushed out by the soft-copy.

Soft-copy are in every part newspapers, but in electronic format only (Like Internet Television and Internet radio).

Most soft-copy are copies of printed editions, but more and more publications are now going online only, like Southport Reporter. The development of papers like this one, on the internet, are leading on to the development of electronic newspaper, that will very soon be replacing hard copy printed papers with portable electronic paper.

4.1   Electronic Paper

In February 2006, the Flemish daily De Tijd of Antwerp announced plans to distribute an electronic-ink version of the paper to selected subscribers. This will be the first such application of electronic ink to newspaper publishing.

Electronic paper, also sometimes called e-paper or electronic ink, is a display technology designed to mimic the appearance of regular ink on paper. Unlike a conventional flat panel display, which uses a backlight to illuminate its pixels, electronic paper reflects light like ordinary paper and is capable of holding text and images indefinitely without drawing electricity or using processor power, while allowing the paper to be changed. One important feature needed is that the pixels be image stable, so that the state of each pixel can be maintained without a constant supply of power.

Electronic paper was developed in order to overcome some of the limitations of computer monitors. For example, the backlighting of monitors is hard on the human eye, whereas electronic paper reflects light just like normal paper. It is easier to read at an angle than flat screen monitors. It is lightweight, durable, and highly flexible compared to other display technologies, though it is not as flexible as paper.

Predicted future applications include e-paper books capable of storing digital versions of many books, with only one book displayed on the pages at any one time. Electronic posters and similar advertisements in shops and stores have already been demonstrated.

5.   Digital Newspaper projects

A good number of digital newspaper projects and programmes were started across the globe, with the intention to preserve the newspapers of various regions, a few important projects are :

5.1   National Digital Newspaper Program

The National Digital Newspaper Program is a joint project between the National Endowment for the Humanities and the Library of Congress to create and maintain a publicly available, online digital archive of historically significant newspapers published in the United States between 1836 and 1922. Additionally, the program will make available bibliographic records and holdings information for some 140,000 newspaper

titles from the 1600s to the present. Further, it will include scope notes and encyclopedia-style entries discussing the historical significance of specific newspapers. The program is ambitious but to date has not produced a single set of text available to the public. As of 2006 not a single image from the project has been made available despite two years of work

5.2   Utah Digital Newspaper project.

In 2001 the J. Willard Marriott Library at the University of Utah was awarded a Library Services and Technology (LSTA) grant to research and demonstrate a newspaper digitization project. The Library succesfully digitized three weekly Utah newspapers for a total of 30,000 pages and developed an effective process that resulted in a second LSTA grant in 2003. That grant application was led by the Utah Academic Library Consortium and digitized another 106,000 pages and 17 new titles. The digitized newspapers in this website may be browsed by issue or searched by keywords, article titles, weddings, deaths, and births.

The Library Services and Technology Act is federal funding allocated by the Institute of Museum and Library Services and administered through the Utah State Library.

5.3   LAURIN project

To continue the tradition of collecting and providing access to newspaper clippings in the new-networked environment, the European Commission embarked upon the LAURIN project in 1998. Seventeen partners from seven European countries are co-operating in developing a model that addresses several issues, including shared indexing, shared and local access, new business models, and copyright. One of the main outputs of the project is an image processing software that replaces scissors, glue, and print files for maintaining newspaper-clipping collections. Other publicly available products include a multilingual thesaurus and a prototype of a network of clipping archives. A first release of the network will be available in Spring 2000. Within the scope of the LAURIN project, several special image processing software and tools have been developed for the digital reorganisation of the clipping activity. This will highlight the image processing software called libClip, which is based on standard software available on the market. Software developers and librarians have collaborated in the design and the customization of the tool for local use. This cooperation has been very beneficial for both sides. Librarians have been able to express their requirements concerning the functionality, handling, design, and workflow of the module.

5.4   British Library Online Newspaper Archive

This site presents five different historical newspaper publications shown in the form of their front page images from 1851, 1856, 1886, 1900, and 1918. It includes The Daily News, News of the World, Penny Illustrated, The Manchester Guardian, and the Weekly Dispatch. Users may select a year, month, and day to call up a particular page image. They may then click on a section of text to have it enlarged. Images may be viewed or downloaded in a PDF viewer for clearer viewing and printing.

5.5   TIDEN - A Nordic Digital Newspaper Library

The first historical "Digital Newspaper Library" in the world, which provides fulltext search to newspapers in Gothic script is a result of the TIDEN- The Nordic Digital Newspaper Project and the Finnish part of it, AURORA. These projects will result in The Nordic Digital Newspaper Library with Nordic newspapers from 1640-1860, open to the public in autumn in 2001.

5.6   NYPL Digital Gallery

NYPL Digital is your gateway to The Library's rare and unique collections in digitized form. NYPL Digital includes searchable databases like In Motion: The African American Migration Experience and NYPL Digital Gallery, online exhibitions such as Before Victoria, text from the Yizkor (Holocaust Memorial) Books, and more.

- Is your gateway to The New York Public Library's rare and unique international holdings in digitized form, including:
  - Historic maps
  - Rare prints and photographs
  - Illuminated manuscripts
  - Unusual printed ephemera
  - Sound files and moving images
  - Original art and more
- Continues to fulfill the Library's traditional mission in the Internet Age to collect, preserve and make its holdings available
- Connects you to digital versions of increasingly valuable, fragile and hard-to-use originals by offering a ever-growing selection of digital collections
- Offers many holdings in their entirety, ensuring the virtual equivalent of consulting them in person
- Organizes holdings via topical or thematic groupings into virtual collections, helping bridge physical (departmental or organizational) separation
- Maintains current links to related NYPL websites such as online exhibitions, archival finding aids, and e-books
- Reflects ongoing development and evolution of best practices in the digital environment
- Results from setting benchmarks and coordinating practices institution-wide for capture, description, storage, retrieval and delivery of digital surrogates

The United States Newspaper Program is a cooperative national effort among the states and the federal government to locate, catalog, and preserve on microfilm newspapers

published in the United States from the eighteenth century to the present. Funding is provided by the National Endowment for the Humanities. Technical assistance is furnished by the Library of Congress.

5.7 Missouri newspapers

Missouri newspapers are one of the most significant historical sources concerning the lives of Missourians, their culture and history.

The purpose of this project was to scan at least ten years of at least five historic Missouri newspapers and provide indexing for those issues. In an earlier project funded by a Library Services and Technology Act grant from the Missouri State Library, one year of several newspapers were scanned and index terms were assigned to significant articles in each of the newspapers. This proved to be very time and labor intensive. In this project, funded by an Institute of Museum and Library Services grant through the Missouri State Library, we originally intended to base the indexing on the headline and full text of the first paragraph of the article. This information was to have been entered by student workers.

About five months into the project, we discovered that Olive Software was marketing a product (Active Newspaper Archive) that would take scanned images from either hardcopy or microfilm, provide full text indexing, and automatically segment the newspaper issue into articles and advertisements. We purchased this software and contracted with Olive Software to process the images scanned in-house from hardcopy, those scanned from microfilm by OCLC, and from microfilm that we sent them.

The Optical Character Recognition (OCR) was not 100% accurate as it depended greatly on the condition of the image to be OCR'ed. However, it has provided full text searching capability at a fraction of the time and effort that manual indexing would have required. The software has tools that allow us to correct the OCR'ed text and re-index it when we feel that is necessary.

We hope that the results of this project will prove useful to students, researchers and the general public. Hopefully, we will be able to add to this collection in the future.

6. Digital Newspaper Archives :Indian scenario

Many newspaper publishers have already started the e-paper in their websites and make accessible all the past news papers, varying from 15 days to 3 months. It is a need to day to have an archive for all the newspapers in the digital form.

7. Conclusion

Technology is only as good as the use it is put to. From the above discussion one cannot compartmentalize whether Microfilm Archives are best suited over Digital Archives or vice versa. Though Digital Archives opens archived material to worldwide access, it

may not stand the test of time. Technology obsolescence is a reality and constant upgradation a necessity. Microfilm Archives are reliable, long-term preservation and retrieval system. In ideal conditions its life expectancy is 500 years, 100 times that of Digital Archives. Digital documents can be tampered or get corrupted whereas Microfilm documents are more or less sacrosanct. It would be ideal for organizations/libraries to adopt both forms of archiving should budgets permit. Microfilm Archives as the long term, safe and secure master true copy of the original and Digital Archives for ease of access, retrieval, dissemination, reuse and re-expression.

References

1. Reformatting: Digital Technology vs. Analog Technology, 18th Annual Preservation Conference, 2003. at http://www.archives.gov/preservation/conference/preservation_conference2003.html

2. http://www.newspaperarchive.com/

3. Digitisation of Newspaper Clippings: The LAURIN Project by Günter Mühlberger at http://laurin.uibk.ac.at/

4. An integrated system for creating a Digital Library from Newspaper Archives at http://www.iit.demokritos.gr/~bgat/DRH_v2.pdf

5. http://www.Newspaperarchive.com

6. http://www.newspaperdirect.com/

7. Newspaper at http://en.wikipedia.org/wiki/Newspaper

8. Newspaper Archives : a knowledge asset at http://eprints.rclis.org/archive/00007534/01/Newspaper_Archives_ICIM_2005.pdf.

BIOGRAPHY OF AUTHOR

Dr D Chandran, presently working as Professor and Head, Dept Library and Information Science, Sri Venkateswara University, Tirupati. He served as Chairman, Board of Studies. he has more than 80 papers and three books. He is closely associated with various Distance Education programmes in India. He participated in a number of conferences and seminars. He is Life member of many LIS professional bodies. He served as joint secretary IATLIS.

Email: chandrand2001@yahoo.co.in