An Introduction to Digitization Projects Conducted by Public Libraries: Digitization and Optimization Techniques

Gürkan Tuna[a,*], Recep Zogo[b] and Bahattin Demirelli[b]

[a]Department of Computer Programming, Edirne Vocational School of Technical Sciences, Trakya University, Edirne, Turkey
[b]Directorate of Library and Documentation, Trakya University, Edirne, Turkey
* Corresponding author. Tel.: +90-284-224-0283; fax: +90-284-224-0287; e-mail: gurkantuna@trakya.edu.tr

Short Communication

ARTICLE INFORMATION

ABSTRACT

It is no longer debatable that the Internet has been changing the way that we use to obtain information. Nowadays, it is really hard for researchers to write a paper without relying on electronic resources such as online databases, electronic journal and libraries due to their limited time. Therefore recent years have witnessed massive digitization projects conducted by public libraries. However, this digitalization trend necessitates addressing policy issues related to digitization and the management of electronic content. In addition, though digitization can be attractive at first, it is a costly transition in terms of staff training, equipment costs, user needs and sustainability. In this study, we propose optimization techniques to address various problems encountered during a digitization process and improve efficiency.

Keywords: Public libraries, digitization, software development, optimization techniques.

I. Introduction

With the introduction of the Internet, most public libraries have started to provide electronic resources such as electronic books, electronic journals and multimedia content. This trend has led to a transformation of them into digital libraries. Library automation systems have helped in the transformation by providing access to available collections through the use of digital catalogues (IFLA, 2013). However, this transformation requires a high-cost and complex digitization project.

Digitization can be defined as the process of converting traditional library materials to digital formats for electronic use and distribution (Witten and David, 2003). In this way, the traditional library materials can be stored in computers and manipulated easily. Most digital libraries are built in the following order (Cornell University Library/Research Departments, 2000), (Alhaji, 2009).

- Selecting documents to be added to the digital library collection and digitizing them to an appropriate digital form,
- Assigning the metadata to the documents being added to the collection,
- Indexing and storing the documents and metadata for efficient search and retrieval,
- Developing a web-based interface to enable browsing, searching, retrieving and viewing the contents of the digital library.

In this paper, we propose optimization techniques that can be used in the daily operation of digital libraries. The techniques can greatly improve the efficiency of a digitization process especially in massive digitization projects.

The rest of the paper is organized as follows. Section II reviews hardware and software components of digitization projects. Section III presents several optimization techniques for digitization projects. Finally, the paper is concluded in Section IV.

II. Digitization Projects

Setting up a digital library mainly requires sources of content in digital form, whether digitized or born digital content. In most applications, a digital library becomes an integral part of the services of a library. Different from the traditional library services, it provides access to digital collections and is built, managed and made accessible to make readily and economically available for use.

Digitization addresses three main needs of libraries as follows:

- Preserving the documents,
- Making the documents more accessible,

- Reusing the documents.

The planning and implementation phases of a digitization project are crucial to its eventual success (Ming, 2000). Because the decisions made during these phases play a key role in determining the sustainability and usefulness of the electronic resources created. Basically, the planning phase consists of tasks related to building digital library collections, defining strategies for performing these tasks, identifying required resources, and determining a timeline for carrying out these tasks. On the other hand, the implementation phase involves actual steps required to set up a digital library collection. The implementation phase is generally led by a project manager and involves several steps as follows (Sitts, 2000), (Smith, 2001):

- Establishing the digitization team,
- Setting up the Information Technology infrastructure,
- Procuring and installing digital library software packages,
- Finalizing policies and specifications,
- Completing arrangement of workflow for digitization
- Creating the online digital library collection,
- Obtaining copyright permissions,
- Providing access to the digital library collection.

Freely available open source digital library software packages can be used to build a digital library.

III.   Optimization Techniques

Digitization of documents for a digital library is generally accomplished in six stages: Registering, Scanning, Optical Character Recognition (OCR), Proofreading and formatting and producing the Final Version (Smith, 2001).   The following list presents optimization techniques and methodologies for the components of digital library software packages developed by public libraries.

- Though most smoothing algorithms reduce background noise and improve the appearance of scanned documents, they are destructive to text and other data on a document image. Therefore, color document images need to be handled using specific smoothing filters for both lossy and lossless compression.
- OCR translates images to a document format that indexers already know, but it causes to lose the layout, images and color of the original image. In a searchable PDF, the textual content extracted via OCR is put behind the image so search indexers can see it and Acrobat Reader allows selecting it as text. Mass adoption of PDF in most digital software packages make searchable PDF's the ideal format to store digitized paper.
- Automatic classification of scanned images reduces the time spent for each image.
- RAW formats retain image information that is normally lost when capturing to common lossy image formats. They allow post-processing with minimal loss of quality. The problem with them is

that they cannot be viewed by most imaging applications and require much space than the common image formats, making them difficult to transmit over networks or e-mail. They can be preferred for rare and valuable documents.
- For multi-site distributed digital libraries, encrypted file sharing solutions that use digital certificates can be used to provide security.

IV.   Conclusion

Digital libraries allow users to work with the electronic versions of full text documents and their associated images. They bring significant benefits to the users. However, setting up a digital library requires a complex digitization project.

This study reviews the steps involved in digitization projects and proposes solutions to improve the efficiency of digital library software packages. Considering the fact that digitization projects take a lot of time, effort and money, the solutions proposed in this paper may prove to be useful source of reference for both the librarians and the practitioners.

References

IFLA (2013) 'Bridging the Digital Divide: making the world's cultural and scientific heritage accessible to all', IFLA/UNESCO Manifesto for Digital Libraries. Available at: http://www.ifla.org/publications/iflaunesco-manifesto-for-digital-libraries [Accessed 24 August 2013]

Witten, I. H. and Brainbridge, D. (2003) *How to Build a Digital Library*, London, Morgan Kaufman Publishers.

Cornell University Library/Research Departments (2000) 'Moving theory into practice: digital Image for libraries and archives'. Available at: http://www.library.cornell.edu/preservation/tutorial [Accessed 25 August 2013]

Alhaji, I. U. (2009) 'Digitization of Library Resources and the Formation of Digital Libraries: A Practical Approach'. Available at: http://www.ais.up.ac.za [Accessed 26 August 2013]

Ming, D. C. (2000) 'Access to Digital Information: Some Breakthrough and Obstacles', *Journal of Librarianship and Information Science*, 32(1), 26-32.

Sitts, M. K. (2000) 'Handbook for Digital Projects: A Management Tool for Preservation and Access', Northeast   Document Conservation Center, Andover, Massachusetts.   Available   at: http://www.nedcc.org/digital/dman.pdf   [Accessed   25 August 2013]

Smith, A. (2001) 'Strategies for Building Digitized Collection', Washington, D.C. Digital Library Federation, Council   on   Library   and   Information   Resources. Available from http://www.clir.org [Accessed 26 August 2013]

Authors' Biographies

Gürkan Tuna serves as an Asst. Prof. at Trakya University, Edirne, Turkey. He received his B.S. degree in computer engineering from Yildiz Technical University, Istanbul, Turkey, in 1998, and his M.S. degree in computer engineering from Trakya University, Edirne, Turkey, in 2009. He received his Ph.D. degree in electrical engineering from Yildiz Technical University, Istanbul, Turkey, in 2012. Tuna has authored several papers in international conference proceedings and refereed journals, and three book chapters. He has been serving as a reviewer for international journals and conferences. His current research interests include smart grid, ad hoc and sensor networks, robotic sensor networks, multisensor fusion, energy harvesting, and energy-aware routing.

Recep Zogo received his master's degree in public administration from Trakya University, Edirne, Turkey, in 2009. He currently serves as the Head of Library and Documentation Directorate, Trakya University, Edirne, Turkey. E-mail: recepzogo@trakya.edu.tr

Bahattin Demirelli received his bachelor's degree in library science from Ankara University, Ankara, Turkey, in 1994. He currently serves as an expert at Library and Documentation Directorate, Trakya University, Edirne, Turkey. E-mail: bahattind@trakya.edu.tr