

Towards the Preservation and Availability of Historical Books and Manuscripts: A Case Study

Eleni Galiotou[†]

[†] TEI of Athens, Department of Informatics. GR-12210, Egaleo, Greece.
egali(at)teiath.gr

Abstract: We present a case study on the use of digitization for the preservation and availability of the books and manuscripts in a historical library. At first, we give a brief description of the collection. Then, we describe the process of digitization and our approach to the availability and awareness of the digitized documents.

Keywords: Historical documents, Digitization, preservation, Availability

I. INTRODUCTION

Collections of manuscripts and early printed books of great philological and historical interest are often kept in libraries of monasteries and other remote historical sites. Access to such collections is a quite difficult task due factors like: the deteriorated state of the documents, the conditions under which the collections are kept, the limited accessibility to the location, etc. The digitization of the documents and their availability to the public via electronic media contributes to their awareness and in the same time their preservation since, it ensures full access to their content without frequent use of the original which would contribute to a further deterioration of their bad state. Recently, mass digitization of historical documents was made possible in the course of research projects such as IMPACT which was funded by the European Commission and aimed at massively digitizing European historical texts and at proposing methods for improving the access to their content (Balk 2009).

In this paper, we discuss the problem of preservation and awareness of the collection of books and manuscripts in the library of the “Evangelistria” holy monastery (The holy monastery of the Annunciation of the Virgin) on the island of Skiathos, Greece. The paper is organized as follows: In section 2 we give a brief description of the monastery and its library while in section 3 we describe the digitization of the documents and problems we came across during this process. In section 4 we discuss our approach to the availability and awareness of the digitized documents. Finally, in section 5 we draw conclusions and point to future work.

II. THE LIBRARY

A. The Monastery

The “Evangelistria” Monastery (the holy monastery of the Annunciation of the Virgin) was founded on

the island of Skiathos, Greece in 1794 by a group of monks, members of the “Kollyvades” spiritual movement who were forced to leave the Holy Mount of Athos in the face of disturbances resulting from disputes over the celebration of Canonical Requiem Masses Saturdays rather than Sundays. The “Kollyvades” movement had as an objective the return to the traditions of the ancient orthodox church and greatly influenced the spiritual life of the island. In addition, the Monastery provided significant moral and material assistance to the Greek revolution of 1821 as well as to the pre-revolutionary movements. The cultural sites of the Monastery include the religious and national museum, the folk museum, the exhibition where byzantine and post-byzantine icons, wooden and silver crosses, liturgical vessels, and other ecclesiastical items are kept.

B. The Content of the Library

Of particular interest is the library where a significant number of rare manuscripts and early printed books are kept and was formed by learned and bibliophile monks and other scholars who donated their books and manuscripts to the Monastery. The collection comprises books printed from the 16th to the 19th centuries, manuscripts from the 13th to the 19th centuries, and patriarchal documents. Obviously, most documents in the collection are religious ones. Indicatively, there are works by St. John Chrysostom, Saint Basil the Great, Athanasius Parios, Nikodemos the Hagiorite, Patriarch Photius etc. Besides religious documents there is also a significant number of books of general interest such as Geography, Mathematics, Grammar and works by Homer, Thucydides, Aristotle etc. The oldest printed book is a hand-bound edition of the Gospel with gold, silver and precious stones which was printed in Venice in 1539 (Fig. 1).



Figure 1. The Gospel (published in 1539).

III. DIGITIZATION

Recently, the task of digitization the library was under-taken in the course of the effort for the maintenance and reconstruction of the Monastery. The main reason for this undertaking was the increasing interest of philologists, theologians and historians in the books and manuscripts of the collection and at the same time a very bad state of the documents. Therefore, our goal was twofold:

- prevent any further damage of the documents
- facilitate the work of the researchers

For the initial recording and classification of the documents we used the “Catalog of Manuscripts and Books of the Holy Monastery of the Annunciation of the Virgin in Skiathos” which was created 1916 and published in 1961-62 by Anthony former Bishop of Elia (1961, 1962a, 1962b). Following the catalog, the documents were located, classified and arranged on the shelves.

An appropriate database was created which would contain all the elements of the catalog and help the location of the documents on the shelves. So, each record of the database would contain the following fields: serial number, title, author/editor, language, publisher, publication year, publication country, volume number, number of pages, comments, library/shelf number. Then, the documents were digitized and linked to the database. The digitization was performed using a digital camera and a tripod bringing to light rare documents as well as their extremely bad condition due to:

- damage (Fig. 2),

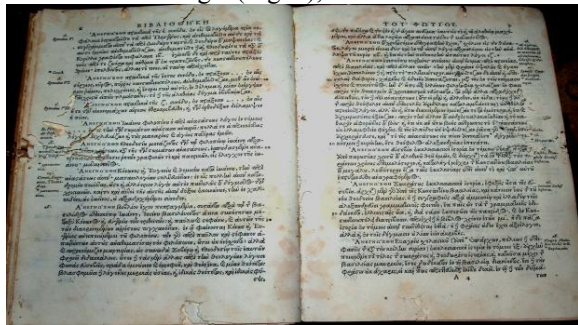


Figure 2. Damaged document.

- stains (Fig. 3)



Figure 3. Stained document.

- bleed-through distortion (Fig.4)

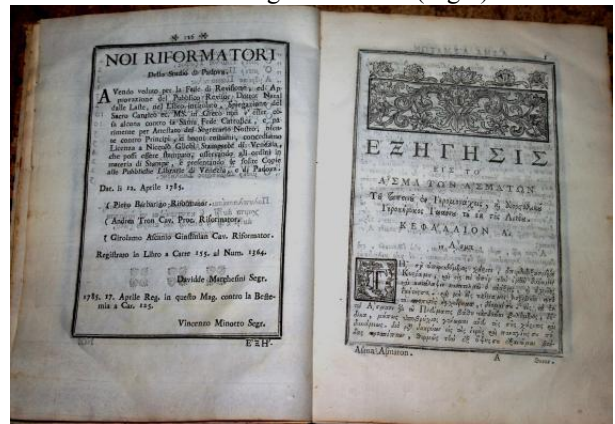


Figure 4. Bleed-through distortion.

- annotations (Fig.5), etc



Figure 5. Annotated pages.

Kesidis et al. (2009) and Kesidis et al. (2011) have proposed methods for dealing with the abovementioned problems along with solutions to the problem of accessing the content of the digitized documents in the collection without the use of an optical character recognition system.

Significant results of an application of these methods are reported in (Kesidis et al., 2011). The experiment was performed on 110 images corresponding to single pages of documents printed during the 17th and 18th centuries and using a query set consisting of 32 key-words. In the proposed word spotting framework, image processing techniques were used to improve the quality of the document images and word segmentation locating words in the document was performed. Then word images were created from keywords and compared to all the words in the digitized documents. Research procedures were refined with the help of a user feedback process. In order to improve the efficiency of the retrieval procedure, a set of natural language processing tools was created, comprising:

- a morphological generator for early Modern Greek that provides the user with the ability search the documents using only a base form and locate all the inflected word forms
- a synonym dictionary which facilitates access to the semantic context of the documents and enriches the results of the search process.

The experimental results have shown that the use of generated word-forms in the word spotting procedure along with the user's feedback has led to significant improvement of the retrieval performance compared to the case when only the keywords were considered.

Upon completion of the digitization procedure of all the documents of the collection, we plan to use the abovementioned techniques in order to improve document images and provide access to the content of the totality of the documents thus facilitating the work of the researchers.

IV. AVAILABILITY AND AWARENESS

With the digitization of the documents and their linking with the database, the first challenge of this undertaking was met; access to the documents became possible without using the originals, thus avoiding further deterioration of their state. Yet, the availability remained an open issue since the system was less than user-friendly.

In order to facilitate the availability and awareness of the digitized documents their publication on a set of DVDs was decided. A user-friendly navigation software to the digitized material was created and 59 printed books (corresponding to about 28000 pages) were processed and published on 3 DVDs along with the software (Fig. 6). The subset of the digitized collection consisting of these 59 books is a representative example of the documents in the library taking into account their philological, theological and/or historical interest, their rarity, their thematic diversity etc.

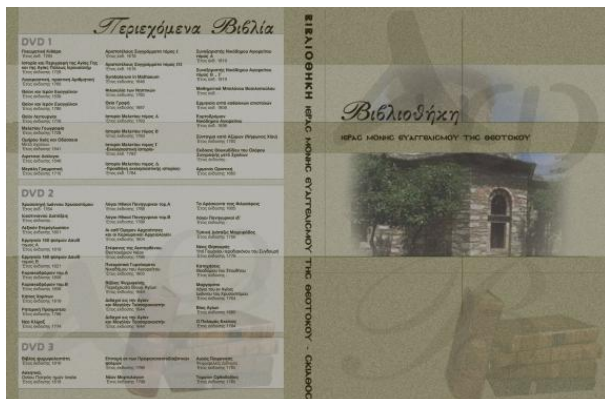


Figure 6. The 59 books available on DVDs.

Special importance was placed on the user-friendliness of the navigation software since the target group would consist mainly from experts in the Humanities who were not particularly familiar with advanced computer use.

With the use of the navigation software, the user is able to access a book (fig. 7),

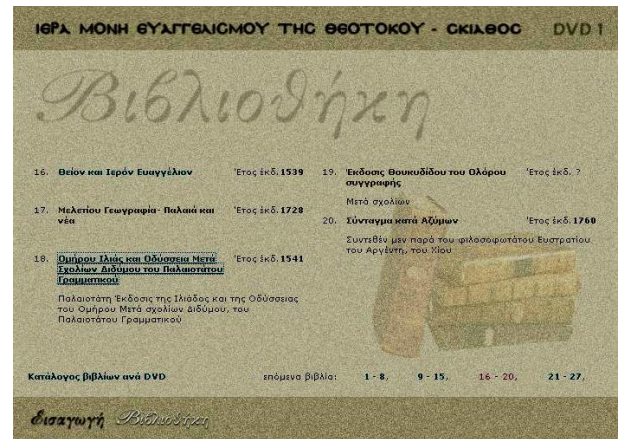


Figure 7. Excerpt of a list of books on DVD .

inspect each page (fig. 8) and print it for further study.

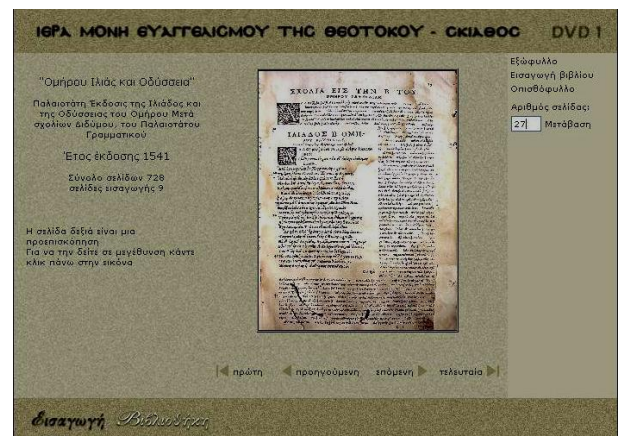


Figure 8. Inspection of an annotated edition of Homer's Iliad and Odyssey (published in 1541) .

The abovementioned undertaking of digitizing the documents and their publication on a set of DVDs was implemented by the Department of Informatics of the Technological Educational Institute of Athens and the company Data & Control Systems Ltd with the support of the Municipality of Skiathos.

V. CONCLUSIONS

In this paper, we have described the first phase of an ongoing project for the preservation and awareness of the historical books and manuscripts in the library of the Evangelistria Monastery on the island of Skiathos, Greece. It comprised:

- the creation of the infrastructure for the inventory and digitization of the documents,

- the digitization of a significant subset of the documents and
- the publication of a representative number of digitized documents on a set of DVDs.

Future work involves the completion of the digitization procedure, the complete cataloguing of the collection, and the availability of the digitized documents on the Web.

REFERENCES

Anthony former Bishop of Elia, “The Catalog of Manu-scripts and Printed Books of the Holy Monastery of the Annunciation of the Virgin in Skiathos (Part I)”, *Theologia*, 31(4) (1961) (in Greek).

Anthony former Bishop of Elia, “The Catalog of Manu-scripts and Printed Books of the Holy Monastery of the Annunciation of the Virgin in Skiathos (Part II)”, *Theologia*, 32(2) (1962a) (in Greek).

Anthony former Bishop of Elia, “The Catalog of Manu-scripts and Printed Books of the Holy Monastery of the Annunciation of the Virgin in Skiathos (Part III)”, *Theologia*, 32(4) (1962b) (in Greek).

[Balk, H., “Poor Access to Digitized Historical Texts: The Solutions of the IMPACT Project”, Proc. 3rd Workshop on Analytics for Noisy Unstructured Da-ta \(AND’09\), Barcelona, Spain, ACM Press \(2009\)](#)

[Kesidis, A.L., Galiotou, E., Gatos, B., Lampropoulos, A., Pratikakis, I., Manolessou, I., Ralli, A., “Accessing the Content of Greek Historical Documents”, Proc. 3rd Workshop on Analytics for Noisy Unstructured Data \(AND’09\), Barcelona, Spain, ACM Press, 55-62 \(2009\)](#)

[Kesidis, A.L., Galiotou, E., Gatos, B., Pratikakis, I., “A Word Spotting Framework for Historical Machine Printed Documents”, *Int. J. on Document Analysis and Recognition*, 14\(2\), Springer, 131-144 \(2011\).](#)