ANALYSIS OF THE OPPORTUNITIES FOR OBTAINING NEW KNOWLEDGE FROM ANCIENT MANUSCRIPTS

Katia Rasheva-Yordanova, Inna Dimitrova, Georgi Dimitrov

University of Library Studies and Information Technology (BULGARIA)

Abstract

In the last few decades, the approaches, methods and tools of studying ancient manuscripts have changed significantly. Technological advances have enabled the extraction of new knowledge from manuscripts to be realized through various manuscript-friendly techniques. Solutions such as spectroscopy, reflectography, and multispectral capturing, as well as various computer algorithms, machine learning, and computational methods are just some of the techniques and approaches within the reach of today's researchers. The assumption is that medieval manuscripts contain a wide variety of metadata, difficult to be extracted by using traditional means.

This paper focuses on the implementation of interdisciplinary methods and tools for the study of ancient manuscripts and their capacity to extract new data for comprehensive research of these monuments. The paper is organized into four sections: The Introduction outlines some general issues connected to storage, preservation, digitalization and research of parchment manuscripts. In the second chapter we discuss the methodology for manuscript study and the specific aspects of their research, mostly connected to computer imaging and visualization of palimpsests. The next chapter focuses on the challenges and limitations to manuscript analysis encountered during the research of ancient manuscripts and palimpsests. The Fourth chapter exhibits the results of our investigation of the topic – a graphic representation of the Process for New Knowledge Extraction and an Algorithm obtained through application of automated solutions and machine learning. The Conclusion supports the thesis that the future development of new knowledge extraction from already researched historical monuments (manuscript codices) depends greatly on interdisciplinary tools and methods which have a bigger capacity to build digital databases and support new and multi-aspect scientific approach.

Keywords: ancient manuscripts, manuscript analysis, new knowledge, palimpsest.

1 INTRODUCTION

Many historical documents have been digitized over the past two decades. The main objective of the digitalization process is motivated both by the need for these cultural monuments to be accessible and widely used, on the one hand, and by the existing capacity to provide their adequate digital copies, on the other. In the general case, the benefits of digitalization are related to (1) preservation and conservation of document funds; (2) providing access to historical document collections through computer networks; (3) construction of digital library databases; (4) collections mobility (so-called libraries without borders); (5) easy navigation and full-text use of the digitized documents.

Ancient manuscripts present a specific case of objects that need to be digitally exhibited. For their storage a constant temperature of 18 degrees Celsius is required, 55% relative humidity and light up to 50 lux. Undoubtedly, the research tasks of manuscript digitization and analysis need to be carried out in strict compliance with all measures protecting the material body of these cultural monuments. Therefore, the interdisciplinary study of ancient manuscripts brings together experts from various fields of science, including chemists, philologists, librarians, computer scientists and others, working in collaboration to expand their knowledge and by implementation of advanced non-invasive research methodology and technical tools, to obtain new data capable of being subjected to various analyses.

This paper focuses on the implementation of interdisciplinary methods and tools for the study of ancient manuscripts and their capacity to extract new data for comprehensive research of these monuments. The paper is organized in three relatively separate sections: The first section examines the specifics of ancient manuscripts and their study. The second section presents an analysis of the existing challenges in ancient documents investigation. The third section proposes an algorithmization of the knowledge extraction process from manuscript data and discusses the possibilities for generating new knowledge through application of automated methods and technologies.

2 METHODOLOGY

2.1 Specific aspects of ancient manuscripts

The word "manuscript" comes from the Latin "manu scriptum", which means "handwritten". With the development of the first scripts and alphabets in various parts of the world, a variety of raw materials and writing surfaces were used, such as stone, metal, baked and unbaked clay, wax tiles, palm leaves, wood, bark, bamboo, cotton, linen or silk fabrics, bone and ivory, animal skins, parchment, papyrus and paper, etc.(Creydt & Fischer, 2021).

Manuscripts fall into the category of the so-called "Rare and Valuable Editions", where the process of digitization requires a special approach, measures and treatment. However, digitalization has a number of benefits for the researchers. On the one hand, digitization can improve the objects' accessibility to a wider audience and protect rare and fragile documents from damages during their study (Bourgeois & Kaileh, 2004). On the other hand, digitalization makes it possible to study the hidden information in the manuscripts.

Over the last twenty years, various techniques and methods of analysis have emerged in this context, through which today we can obtain additional information about the history of the ancient manuscript and the living conditions at that time (Creydt & Fischer, 2021). Ancient manuscripts may possess a large amount of supplementary information contained in the underlying textual strata and the material body composition, whose imperceptibility to the naked eye, poses a challenge to their analysts.

In the past, writing materials were so scarce that they were often used more than once. Thus, a parchment may have been used several times. This phenomenon creates a new, multi-layered artifact called the palimpsest, in which the process of erasing the original manuscript is not entirely effective, and the original can often be distinguished below the superficial inscriptions. This gives researchers the opportunity to get a glimpse of some previously undisclosed details of manuscript history.

2.2 Specific aspects of palimpsest manuscripts

Etymologically, the word palimpsest comes from the Greek palin ("again") and psaein ("scrape") (Moss & Schreiber, 2006). Generally speaking, the term palimpsest refers to a specific type of reuse or recycling of a manuscript, in which the original content of the manuscript is partially or completely erased by scraping or washing and is replaced by new content at a later time (Okello & Duran, 2021) (Moss & Schreiber, 2006).

Palimpsests creation is due to one of the following reasons (Lorenz, 2021): (1) expensive and difficult to find manuscript material; (2) isolation of the monastic centers; (3) ageing of certain written monuments or if their content is assessed as outdated or incomprehensible by the users. There are two distinctive subtypes of palimpsests (Lorenz, 2021):

(1) Palimpsest, created as a result of reuse of the material (the most common type of palimpsest) - when the parchment of a manuscript is used for a new manuscript with unrelated content. A palimpsest belonging to this type consists of two main elements: *scriptio inferior*, which derives from the old manuscript and the newly added written text - *scriptio superior*.

(2) Palimpsest, created by adapting the artifact - a less common type of palimpsest, includes fragments of two or more manuscripts, integrated into one newly created body. As a result, the manuscript incorporates two main constituent parts - older and newer one.

Both techniques were especially popular in the Middle Ages when this type of recycling was carried out by scribes - mainly due to lack of material. In the recycling process water, or chemical agents were used to delete the existing text, after which the new text was applied to the cleaned base. However, washing and scraping were usually not complete, as they often left traces of previous writings (Okello & Duran, 2021). Over time, traces of the old script reappeared, leading to the creation of the palimpsest.

Nowadays, the word palimpsest can refer to anything that has several layers. For scholars in education, the notion of the palimpsest refers to the "discursive, material, and spiritual connection of history and present based on the notion that, in a palimpsest, despite the obliteration, remainder of the previous text survives". (Coloma, Means, & Kim, 2009) (Okello & Duran, 2021) In a philosophical sense, the palimpsest is "a useful way of understanding the existential complexity of culture, as previous 'inscriptions' are erased and overwritten, yet remain as traces within present consciousness" (Ashcroft, Griffiths, & Tiffin, 2000), (Okello & Duran, 2021). However, the current scientific investigation of ancient palimpsests is more focused on the developing of new methods and techniques, capable of extracting new data from old monuments and the potential they offer for knowledge accumulation.

3 Challenges to manuscript analysis

Ancient manuscripts pose significant research challenges to the scholars involved in their analysis (Asi, Cohen, Kedem, El-Sana, & Dinstein, 2014), (Antonacopoulos & Downton, 2007). These challenges often depend on the research objectives and the specific approach chosen to addressing them.

Manuscript analysis on many occasions require the use of various analytical platforms. At the same time, in order to obtain correct interpretations and conclusions, a number of additional sampling techniques and procedures need to be introduced (Newton, Ramage, & Gadegaard, 2018), which must be as sparing and minimally invasive as possible and incapable of inflicting damages to the manuscript.

Research tasks aiming at ancient manuscripts study often include: (1) analysis of the age of the document; (2) material analysis of the ink and parchment; (3) paleographic analysis and determination of the number of scribes; (4) derivation of metadata from manuscripts; (5) examination for the presence of palimpsests in the document; (6) improving the readability of the texts in the various layers of the manuscript, and many more.

Researchers of ancient manuscripts may specialize in different scientific fields and work on different types of tasks, and still encounter similar limitations and challenges (Table 1).

Table 1. Captions should be placed above the tables.

Research task	Challenge
Manuscript age analysis	It is achieved through physical and chemical analysis, which are not often sparing and can lead to partial or complete destruction of manuscript parts.
Identification of the number of scribes (authors) of the document	The text is partially in bad condition, since characters are faded-out and the images are degraded by background noise (Fiel, Hollaus, Gau, & Sablatnig, 2014).
Extraction of metadata from manuscripts	- Medieval manuscripts contain a wide variety of metadata, which makes it difficult to extract them by using traditional means. - Irregular text layout of documents - many manuscripts include text in the fields of the page (so-called side notes) which serve as notes to the main text and are located in the frame (fields) of the page. While the main text is mostly horizontal, the side notes can have different orientations and be placed in the page margins. (Antonacopoulos, Clausner, Papadopoulos, & Pletschacher, 2011) (Antonacopoulos, Clausner, Papadopoulos, & Pletschacher, 2013)
Examination for the presence of palimpsests in the document	The older text that has been cleaned can be restored and read by chemical means, which, however, risk damaging the manuscript.

Nowadays most of the above tasks can be accomplished through the intervention of computer algorithms, machine learning and computational methods. To solve the set of tasks, the researchers focus mainly on automated solutions and various manuscript-sparing techniques. Some of them will be presented in the next section of the report.

4 RESULTS

4.1 Process of extracting new knowledge from manuscripts

The process of studying manuscripts is a multi-aspect one and can be presented as a linear algorithm with four main stages - (1) preparation of the manuscript, (2) preparation of the equipment (3) digitization and processing (4) analysis and extraction of new knowledge (Fig. 1).
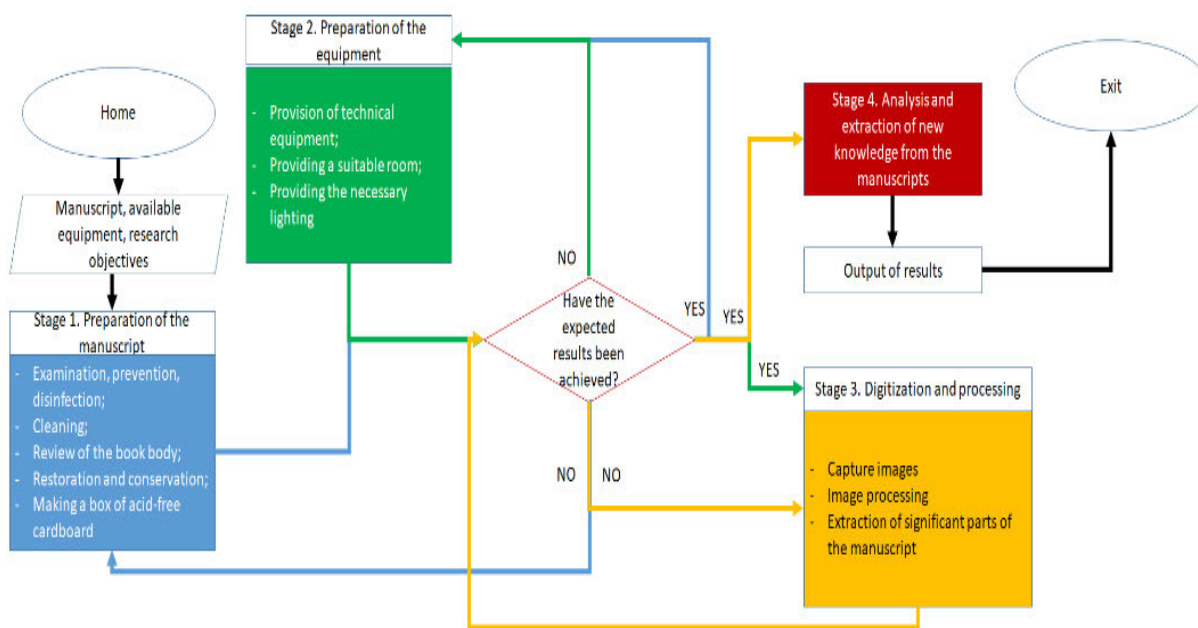
Figure 1. Algorithm of new knowledge extraction from ancient manuscripts

The stage related to proper new knowledge acquisition is Stage 4 (number 4). However, its realization is impossible if performed independently or in case it precedes any of the other 3 stages. Stage 1 (manuscript preparation) is usually the most time-consuming and incorporates overcoming more challenges during implementation than the other stages. It involves only manual operations which are impossible to automate due to the specifics of the material and the age of the object. Process automation can be performed only in the third and fourth stages of the algorithm. In general, the introduction of technological solutions could amplify the knowledge gained from the studied objects. For example, the digitization of the manuscript using specific UV lighting after post-processing of the frames provides more opportunities to bring forth the lower layer of the palimpsest and extract new knowledge from it. Moreover, with the development of image analysis and imaging technology, physical manuscript investigation can be replaced by non-destructive and non-invasive image-based methodology for obtaining scientific data, in order to prevent the fragile parchment codices from being damaged (Lu, Psarrou, Licata, Konstantinou, & Kokla, 2008).

4.2 Automated solutions to manuscript study

Returning to the tasks performed by the researchers, which we discussed in the previous sections, we can identify several significant benefits of process automation in the last two stages of the algorithm (Table 2).

Process automation could increase productivity of digitized images processing. In addition, automation is applicable to retrieving metadata and determining the number of scribal hands in a manuscript.

On the other hand, if not properly handled, automation may lead to accumulation of incorrect data and thus formulation of inexact research hypotheses. Therefore, digital data examination and verification using various analytical techniques during Stage 4 can provide means for correct data study and interpretation. Considering the multiple benefits of developing an interdisciplinary approach to study manuscript monuments, implication of the above research algorithm will contribute to preservation of valuable manuscripts and provide novel means of knowledge acquisition.

Table 2. Automated solutions to manuscript study.

Research task	Solution through automation	Methods and techniques
Extraction of metadata from manuscripts	Fine indexing is possible by applying image analysis (Bourgeois & Kaileh, 2004), through which information useful for indexing documents can be extracted.	Extraction by using computer vision systems
Identification of the number of scribes	Once automated, the task of scribal hands identification can be applied to a huge amount of historical documents and thus become a valuable tool for paleographers. (Fiel, Hollaus, Gau, & Sablatnig, 2014).	Study of graphemic representation (Bensefia, Paquet, & Heutte, 2003). - Kinetic and allographic feature analysis (Bulacu & Schomaker, 2007). - Detecting ink trail width (Brink, Smit, Bulacu, & Schomaker, 2012)
Improving the readability of the manuscript	With appropriate techniques for capturing and processing of captured images barely visible or deleted characters can be partially or completely restored and text damages due to poor storage conditions.	Multispectral imaging (Hollaus, Gau, & Sablatnig, 2012)
Study of palimsests	Increases productivity and protects the manuscript.	Through the application of modern photo and digital methods and machine learning.

5   CONCLUSIONS

Despite the fact, that most of the world's ancient manuscripts have already been digitized, we still cannot be sure that modern science has succeeded in obtaining all the knowledge from them. We are witnessing the constant discovery of new opportunities for analysis, favored by the development of technology.

This paper analysed the existing practices and potentials for extracting knowledge from manuscripts and presented an algorithm that visualized the main stages required for the study of manuscripts. The analysis performed here gives us reason to believe that the main goal of manuscript researchers in the future could be directed to implication of state-of-art technology such as spectroscopy, reflectography and multispectral capturing of images, capable of giving an insight into research aspects, unavailable to the naked eye. This assumption is based on the greater capacity of these scientific methods to obtain new knowledge from already researched historical monuments and build digital databases creating an opportunity for a new multi-aspect and comprehensive scientific approach and simultaneously affording protection to valuable historical artefacts.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Antonacopoulos, A., & Downton, A. (2007). Special issue on the analysis of historical documents. *International Journal on*, 75-77.

[2]   Antonacopoulos, A., Clausner, C., Papadopoulos, C., & Pletschacher, S. (2011). Historical document layout analysis competition. *International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1516–1520). IEEE.

[3]   Antonacopoulos, A., Clausner, C., Papadopoulos, C., & Pletschacher, S. (2013). (pp. 1454–1458). IEEE.

[4]   Ashcroft, B., Griffiths, G., & Tiffin, H. (2000). Post-colonial studies: The key concepts. *Psychology Press*.

[5]   Asi, A., Cohen, R., Kedem, K., El-Sana, J., & Dinstein, I. (2014). A Coarse-to-Fine Approach for Layout Analysis of Ancient Manuscripts.

[6]   Bensefia, A., Paquet, T., & Heutte, L. (2003). Information retrieval based writer identification. *ICDAR*, 946-950.

[7]   Bourgeois, F. L., & Kaileh, H. (2004). Automatic Metadata Retrieval from Ancient Manuscripts. *Document Analysis Systems.* Florence: DBLP.

[8]   Brink, A., Smit, J., Bulacu, M., & Schomaker, L. (2012). Writer identification using directional ink-trace width measurements. *Pattern Recognition*, 162-171.

[9]   Bulacu, M., & Schomaker, L. (2007). Automatic handwriting identification on medieval documents. *ICIAP*, 279-284.

[10]   Coloma, R., Means, A., & Kim, A. (2009). Palimpsest histories and catachrestic interventions. *Counterpoints*, 3-22.

[11]   Creydt, M., & Fischer, M. (2021). Mass Spectrometry-Based Proteomics and Metaproteomics Analysis of Ancient Manuscripts. In *Exploring Written Artefacts* (pp. 183-212). De Gruyter. doi:https://doi.org/10.1515/9783110753301-010

[12]   Fiel, S., Hollaus, F., Gau, M., & Sablatnig, R. (2014). Writer Identification on Historical Glagolitic Documents. *Society of Photo-Optical Instrumentation Engineers (SPIE).* doi:https://doi.org/10.1117/12.2042338

[13]   Hollaus, F., Gau, M., & Sablatnig, R. (2012). Multispectral Image Acquisition of Ancient Manuscripts. *EuroMed 2012: Progress in Cultural Heritage Preservation* (pp. 30-39). LNISA. doi:10.1007/978-3-642-34234-9_4

[14]   Johannessen, L. (н.д.). Folklore and cultural memory: promises and pitfalls. *Performing wisdom: Proverbial lore in modern Ugandan society*, 2013.

[15]   Lorenz, T. (2021). Liturgical Books in Icelandic Palimpsests. *Háskóli Íslands Student Conference on the Medieval North.*

[16]   Lu, T., Psarrou, A., Licata, A., Konstantinou, V., & Kokla, V. (2008). Unsupervised Classification for Ancient Manuscript Analysis. *International MultiConference of Engineers and Computer Scientists.* IMECS.

[17]   Moss, C., & Schreiber, J. (2006). The palimpsest: A conceptual framework for understanding teacher beliefs. *The American Educational Research Association Annual Meeting.* San Diego.

[18]   Newton, J., Ramage, G., & Gadegaard, N. (2018). Scientific Reports.

[19]   Okello, W., & Duran, A. (2021). Here and There, Then and Now: Envisioning a Palimpsest Methodology. *International Journal of Qualitative Methods*. doi:10.1177/16094069211042233