Digital Tools for Medieval South Slavic Manuscripts Research

Detelin Luchev
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, bl. 8, Acad. Georgi Bonchev, Str., 1113 Sofia, Bulgaria
dml@math.bas.bg

Maxim Goynov
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, bl. 8, Acad. Georgi Bonchev, Str., 1113 Sofia, Bulgaria
goynov@gmail.com

Desislava Paneva-Marinova
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, bl. 8, Acad. Georgi Bonchev, Str., 1113 Sofia, Bulgaria
dessi@cc.bas.bg

Radoslav Pavlov
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, bl. 8, Acad. Georgi Bonchev, Str., 1113 Sofia, Bulgaria
radko@cc.bas.bg

Konstantin Rangochev
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, bl. 8, Acad. Georgi Bonchev, Str., 1113 Sofia, Bulgaria
krangochev@yahoo.com

Lubomir Zlatkov
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, bl. 8, Acad. Georgi Bonchev, Str., 1113 Sofia, Bulgaria
lyubcho@gmail.com

ABSTRACT

This paper presents a web-based environment, designed to provide researchers of Old Church Slavonic manuscripts with an intuitive, semantically sound framework and the respective tools for collection, description, categorization, analysis, and presentation of their respective findings. And while the digital *repertorium* described here concentrates mostly on the paleographic features of the South Slavic manuscripts from the fourteenth century, a period of considerable literary achievements in the Balkans, we feel that the incorporation of the state-of-the-art methods form the field of digital presentation and preservation of cultural heritage will contribute to the better understanding of a still under-researched topic.

CCS CONCEPTS

• **Information systems**; • **Information systems applications**; • **Digital libraries and archives**; • **Information retrieval**; • **Specialized information retrieval**; • **Applied computing**; • **Document management and text processing**; • **Arts and humanities**;

KEYWORDS

Digital palaeography, South Slavonic manuscripts, Semantic description, Manuscript typologies, Data access and analysis, Script analysis

ACM Reference Format:
Detelin Luchev, Maxim Goynov, Desislava Paneva-Marinova, Radoslav Pavlov, Konstantin Rangochev, and Lubomir Zlatkov. 2024. Digital Tools for Medieval South Slavic Manuscripts Research. In *International Conference on Computer Systems and Technologies 2024 (CompSysTech '24), June 14, 15, 2024, Ruse, Bulgaria.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3674912.3674946

CompSysTech '24, June 14, 15, 2024, Ruse, Bulgaria

ACM ISBN 979-8-4007-1684-3/24/06
https://doi.org/10.1145/3674912.3674946

1 INTRODUCTION

The fourteenth century was a period of cultural and spiritual advancement in the entire Balkan peninsula, distinctly manifested throughout the lands of the Southern Slavs. The exploration of the writings on the Balkan peninsula during the period has been a long-standing goal for the researchers in the Slavic languages field. The collection of data from the primary medieval written sources, the specification of the origin, the systematization of the manuscripts, collected in diverse repositories, according to their paleographic characteristics, have all been challenging research tasks [1].

Since the end of the twentieth century, there have been efforts from various libraries on projects on digitalization and online categorization of the Slavic manuscripts from the fourteenth century, however currently there is currently no single high-quality method for their description. The codices are presented in most cases with only a brief description together with digital images on microfilms or in graphic file formats, but this way of presentation only provides bibliographical references to the descriptions, rather than comprehensive information on the manuscripts [2].

The research work presented in this paper aims to create IT tools for exploration and study of handwriting, copying and writing in general. It provides the means for structurization, processing, management, visualization, and analysis of these data, significantly facilitating the research process. An online accessible digital *repertorium* (a specialized web-based software environment) was developed to store and preserve information about valuable manuscripts and copyists (scribes) with significant contribution to the development of the South Slavic languages, writings and culture. The goal is the attainment and efficient utilization of the vast amounts of knowledge on the digitalized written heritage, and to provide tools for higher adaptivity and efficient interactions between the researcher and the digital *repertorium*.

In part 2 of the article, the goals and directions of work on South Slavic copyist and scriptoria from the fourteenth century are presented. Part 3 sets out in detail the developed Semantic structures

for representing knowledge of South Slavic manuscripts, copyists and writing. The results regarding the created digital *repertorium* for medieval South Slavic manuscripts and scribes, based on an established theoretical framework and a working scientific methodological approach, are presented in part 4.

2 SOUTH SLAVIC COPYIST AND SCRIPTORIA FROM THE FOURTEENTH CENTURY

The research project "Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)" aims to collect data from medieval primary written sources and providing that data to the scientific community as a reliable tool for origin resolution, definition, and systematization of the manuscripts, preserved currently in the book repositories, from a paleographic perspective [3]. For the implementation of that aim, the project is concentrated on the exploration of new sources and their consolidation with existing sources, thus contributing to the methods for reliable identification of specific medieval scribes, based on the information on South Slavic copyists from the fourteenth century – a particularly significant period for the cultural and historical development on the Balkans, influential on the cultural processes in Europe as a whole.

The sizeable literary production, related to translation and editing of various liturgical texts, replenishment of the new sacramental collections (*i.e.*, newly sourced fixed-content collections) and works on ascetic and monastic topics, the introduction of the Jerusalem (Sabbaite) Typikon required considerable copyist efforts [4]. Concurrently, the development of calligraphy in the Bulgarian and Serbian lands during the fourteenth century was influenced to a great extent by the Greek minuscule writing and the use of Cyrillic became less stereotyped, which affected the possibilities for manifestation of the personal qualities of the copyists, along with the tendencies for development of the main calligraphic schools [5].

Despite a number of monikers of medieval copyists from the fourteenth century is known and there being attempts to differentiate groups of handwriting, based on their respective penmanship [4], there is still a lack of global research and synchronization of their literary work in the Slavic medieval studies. The development of a methodology for identification and classification of the nature and types of writing, as well as the work of the individual copyists, through which each particular case could be consistently conformed, and not left to the intuition of the specialists, which is still the most frequently used approach, is an emerging, wide field for collaboration between the Slavic paleography and the computer technologies in the framework of digital paleography.

3 SEMANTIC STRUCTURES, REPRESENTING KNOWLEDGE OF SOUTH SLAVIC MANUSCRIPTS, COPYISTS AND WRITING

In order to assist the research and analysis of the manuscript works of the South Slavic scholars of the fourteenth century, the project includes machine-readable and -processable semantic structures development (*i.e., Descriptive model of the Medieval South Slavic literary heritage from the fourteenth century*), allowing efficient digital access, semantic description, structuring, automated processing, management, visualization, and analysis of immense volumes of scientific knowledge, data, and metadata of the invaluable literary heritage, accessible and presented in the digital *repertorium*. The model is mainly focused on the specific manuscript and its components, copyists, writing and its paleographic features.

A key role in the descriptive model of the South Slavic medieval heritage has the literary resource, presented in the class "Manuscript", characterized with descriptors of the three knowledge levels" "Identification", "Description", and "Paleography" (Figure 1). The "Identification" level represents general identification aspects of the "Manuscript" object", outlined by the descriptors "Conventional nomenclature", "Origin (Scriptorium)", "Location", "Dating (year)", "Signature", "Typology" and the respective descriptors for the relevant additional notes. The "Description" level comprises descriptors for "Orthography", "Material (paper, parchment, mixed, *etc.*)", "Size (in millimeters)", "Number of pages (manuscript body, notes, end page (header), end page (footer), *etc.*)", "Bibliography". The most important for the identification of the paleographic characteristics of the manuscripts is the "Paleography" level. It includes descriptors such as "Typology of the writing", "Decorative elements of the writing", "Slant of the letters", "Size", "Illuminations", "Copyists", *etc.* Figure 1b indicates the primary descriptors, characterizing a "Copyist" (author, scribe, copyist) class, such as "Name", "Title", "Period of activity", "Bibliography".

The „Paleography" level presents a short description of the handwriting, as well as an analysis of the separate characters, outlining specific characteristic of the copyist's handwriting style, including:

- exterior glyph elements, e.g. punctuation, diacritics, and illuminations;
- interior glyph elements, e.g. dots and hyphens;
- modifications of glyphs, e.g. ligatures and special symbols [6].

The paleographic description in the proposed model is based on the glyph canon of the Old Church Slavonic alphabet, described in detail in the P.P. Chobitko manual [7]. The choice of this particular work is due to the necessity to examine *the manner of inscription of each individual letter*, to study, and to describe the number of strokes and the writing order used by each copyist. This aids the determination of the specific characteristics of the writings of a particular copyist and allows comparisons between different handwriting styles.

In the context of the development of the model and its efficient application by researchers in the paleographic field, the input (or annotation) of sufficient number of manuscripts in the digital *repertorium* enables the computation of glyph (*eg.* punctuation) frequencies. This kind of computation provides the opportunities for not only analyzing the copyist's handwriting, but also determining whether the discovered features are typical for the historical period and the scriptorium to which the manuscript pertains. The data comparison allows for a more detailed categorization of the codices and their dating, assignment to a particular scholarly center, or even for a basic distinction between unical and semi-unical scripts [7].

The descriptive model includes a complete set of relations between concept classes, typical for a conventional relational database (*eg.* 1:N, N:M), but is designed in a flexible and expandable manner

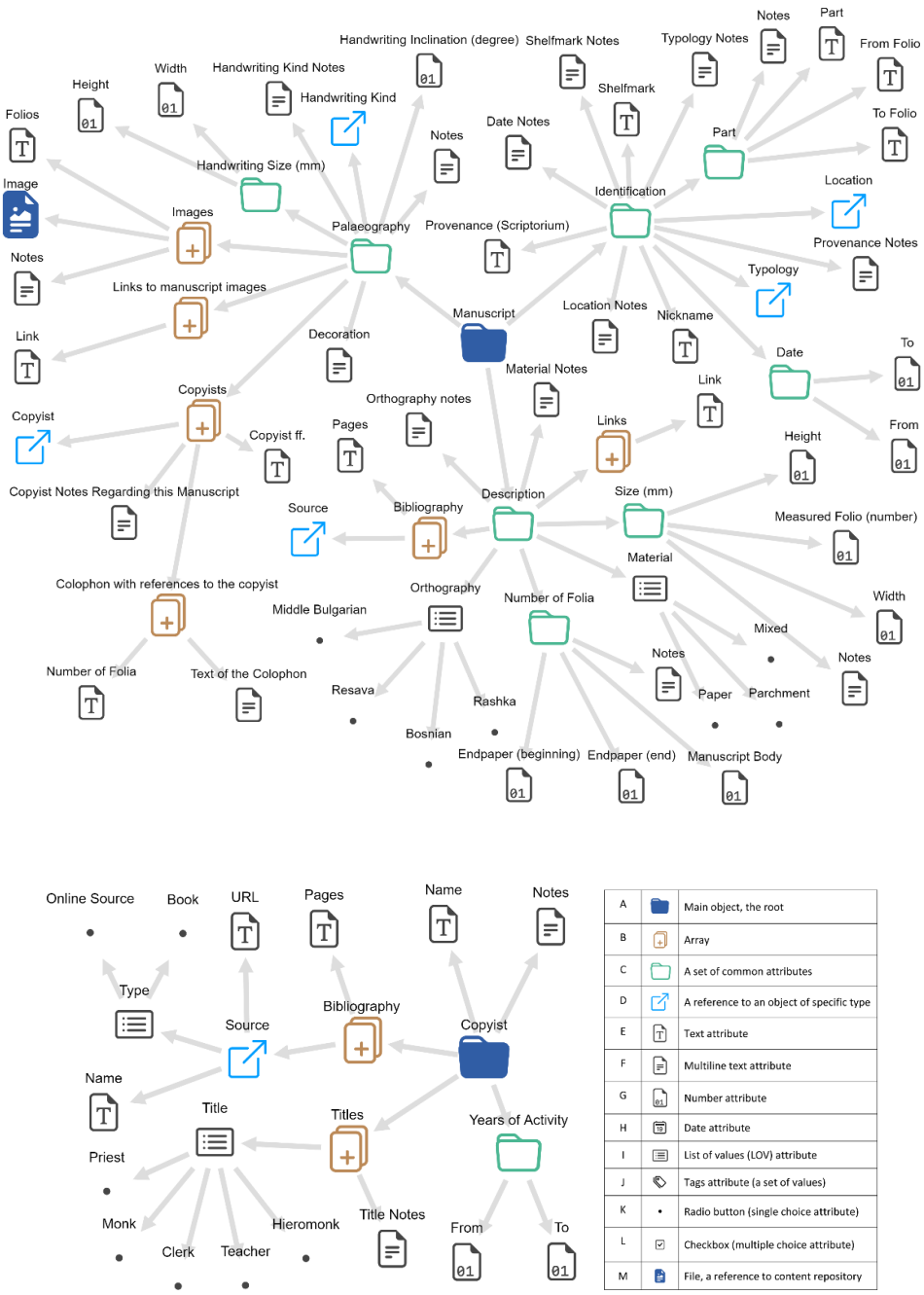

Figure 1: A descriptive model of the South Slavic medieval heritage: a) class "Manuscript" (top), b) class "Copyist" (bottom).

by utilizing a non-relational database [8]. This enables the augmentation of the model with new descriptors and new object classes. The one-to-many relationship (1:N) is implemented with the "Array" element (B) (Fig. 1). The many-to-many relationship (N:M) utilizes a combination of arrays (B) and references (D) (Figure 1). The "Reference" element is a simple reference to the actual object (in this way the redundancy in the database is minimized). Sets of common elements (C) are utilized for sustain the adequate semantic organization of the descriptors. The function of the simple data elements (E, F, G, H, I, J, K, L) is to guarantee the purity of data

in the model, to ensure the convenience of the data input editor, and to contribute to the optimal structuring of the information in the database with the purpose of subsequent efficient structured searching. The arrays (B) and the references (D) ensure the realization of complex relations between the objects' features and other objects in the model.

The descriptive model allows:

- A detailed description of a particular Slavic medieval manuscript in a diverse context (eg. identification, description of the paleographic characteristics content);
- Compilation and structured presentation of data from primary medieval literary sources, preserved in various manuscript repositories (*e.g.* in the Balkan countries, Russia and Ukraine, in monastery libraries in the Holy Mountain, in the monastery of St. Catherine in Sinai, *etc.*, as well as in various libraries in Western Europe);
- Efficient use of the accumulated knowledge of the digitalized literary heritage by designing of a digital *repertorium* for online available manuscripts, and providing collection, searching, browsing, systematization, synthesis, data management and intelligent curation;
- Presentation/discovery of new (hidden) knowledge about the manuscripts – examination of variations, influences and trends, elaboration on so-far-unidentified paleographic features, determination of origin, school and presumed copyists, association of manuscripts from different locations, but with common author, school, *etc.*, determination of handwriting styles, illumination elements, *etc.*
- Presentation/discovery of new (hidden) knowledge about the authors (copyists, scribes, calligraphers) – relation of manuscripts to unidentified authors, specification of writing styles/handwriting, *etc.*, the authors' relation with known scribal schools, discovery of new calligraphic schools, *etc.*

4 DIGITAL *REPERTORIUM* FOR MEDIEVAL SOUTH SLAVIC MANUSCRIPTS AND COPYISTS

The information about the Slavic manuscripts and copyists from the fourteenth century is processed, analyzed and published in the specially dedicated online *repertorium* (https://kopisti.kmnc.bg/bg), based on an established theoretical framework and a working scientific methodological approach. From a research perspective, the collection and classification of the required data for reliable identification of genuine medieval calligraphers aims at the derivation of new knowledge about the organization and activities of the calligraphic schools and local workshops, where they were educated and produced their works. Each literary resource is digitized and described according to the descriptive model based on the available metadata. Once added to the digital *repertorium*, it could be found through a variety of context-based services for browsing and intelligent curation. The tools provided offer the means for detailed research and analysis of each glyph and its elements, detection of similarities between manuscripts (*e.g.* inscription of letters and their elements), support for collaborative, simultaneous work by researchers in the system, and management of the process of description and structurization of data.

The digitized literary resources are accessible and intelligently curated (reviewed, discovered, examined for similarities, *etc.*) by the automated services. The *repertorium* is a web-based software environment and evolved from the team's experience in creating digital cultural content management systems (*i.e.*, CultIS, [10]), such as virtual museums, digital libraries, *etc.* [9]. Functional components provided by the *repertorium* are metadata management, presentation function modules, advanced services and administrative services related to media storage and user data storage.

Currently, the system contains metadata on over six hundred copyists, over fourteen hundred manuscripts, and more than eighty locations. Over eighty interdependencies between manuscripts are displayed. For completion of the database, over three hundred bibliographic sources are cited, representing scientific studies on the South Slavic medieval literary heritage. The metadata are input in two languages – Bulgarian and English.

The primary challenges, associated with the creation and administration of the objects metadata (digitized manuscripts, copyists, typologies, relations, sources, locations, *etc.*) are approached with several functional modules. The data management module is designed to aid the expansion, editing, storage, and post-processing of the digitized manuscripts and the respective metadata. The module provides the opportunities for efficient data input, including the reutilization of already present metadata, suggestions of predefined input options, autocomplete, augmentation of additional digital objects, *etc.*

The basic control panel of the object management module enables simultaneous metadata input compliant with the descriptive scheme in Bulgarian and English. The auxiliary panel offers the editors the possibilities to reevaluate already input values for the relevant descriptors and, if necessary, to reselect and reuse the information. This panel also offers options, related to translation and transliteration of texts, review of an object modification history, with backup recovery options, review of all references to the object, object export to JSON format, *etc.*

The object presentation module offers various presentation options for object sets, selected by the search and arrangement module. The search variables (in collection or presentations) could be saved, allowing the users of the system to predefine their preferred selection options and reuse them without the need to respecify the browsing criteria. All data and presentation are available for export in an appropriate format for further processing. The users have the opportunities to compose advanced search request in the database, without necessarily having proficiency in the field. The search forms could simultaneously combine multiple search criteria with diverse search operators (numerical (*eg.* >, <. =), text (full-text, fuzzy, dates, *etc.*). The search form functionalities are also supported by lists of predefined values for each field, enabling the users to easily observe data values already present for a certain descriptor and reuse them in further search requests. The search results could be presented in list (or table) format, as wells as in grid format, with thumbnails for each object (if available).

The basic control panel of the object presentation module comprises all object metadata (Figure 2), input in compliance with the respective descriptive scheme. In the auxiliary panel, the users could review all created references to the particular object. The

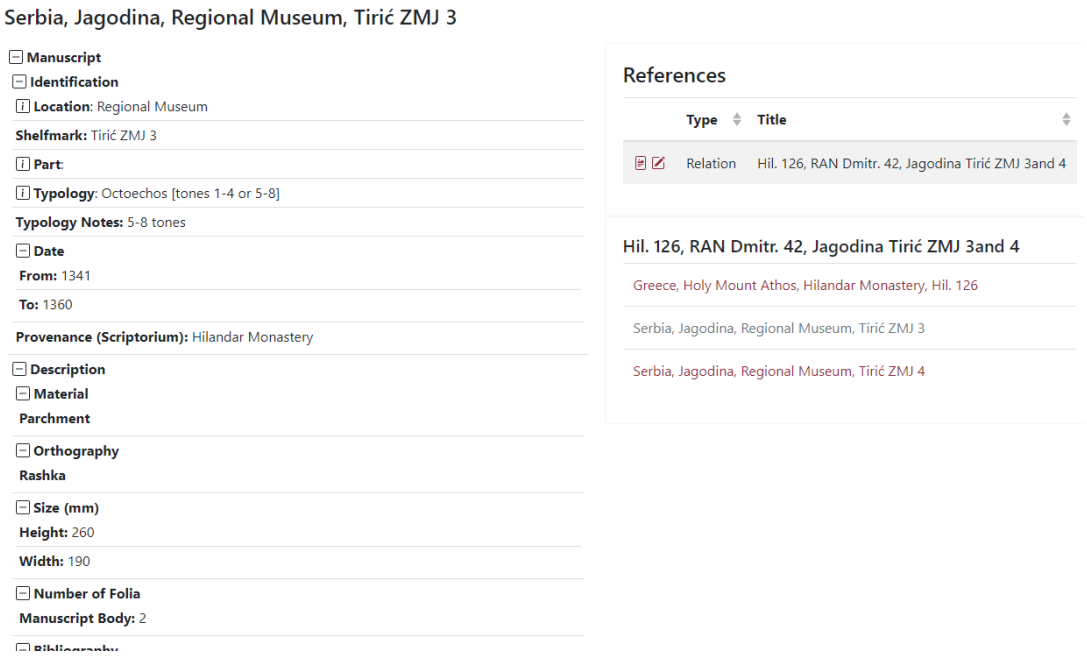

Figure 2: Object presentation module.

basic preview options in the system could be altered and redesigned to optimally fulfill the users' requirements.

5 CONCLUSION

The digital *repertorium* for medieval South Slavic manuscripts and copyists is available online and stores information about the works, handwriting styles, and biographies of the fourteenth century South Slavic scholars and the monastic scriptoria with significant contribution to the development of the South Slavic languages, literary tradition and culture. The digitization, systematization, compilation and curation of the literary artifacts of their work could benefit researchers in appreciating in a new light the knowledge gained from the medieval South Slavic manuscripts, contributing to the studies of the cultural and scientific policy of Bulgaria and the Balkan countries as whole. The use of structured data achieved in the project also presupposes work on potential machine learning models as a more practical contribution to support the research community in its analytical developments. At the same time, the development of the project also contributes to a better theoretical understanding and solving of the research questions arising in the field of digitized collections.

ACKNOWLEDGMENTS

This research is supported by the Bulgarian Scientific Fund under the research project No KP-06-N50/4/30.11.2020 "Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)" (web page: https://kopisti14.kmnc.bg/).

HISTORY DATES

Received . . .; revised . . .; accepted . . ..

REFERENCES

[1] Marko Skarpa, Stilyana Batalova, and Marie-Hélène Blanchet (Eds.). 2021. Religious Bloom Bulgaria 13th – 15th c. Proceedings of the International Scholars Conference, Sofia, 12 – 13 July 2019. Cyrillo-Methodian Research Centre.

[2] Marco Scarpa, Marta Riparante, Desislava Paneva-Marinova, Maksim Goynov, and Detelin Luchev 2022. Online Database for South Slavonic Manuscripts from the 14th c.. Palaeobulgarica XLVI, 1, 181-202. https://palaeobulgarica.eu/en/560/online-database-south-slavonic-manuscripts-14th-c [In Bugarian]

[3] Marco Scarpa, Constanta Burlacu, and Desislava Paneva-Marinova. 2021. New Digital Life for Fourteenth Century South Slavonic Scriptoria. Digital Presentation and Preservation of Cultural and Scientific Heritage, 11, 57–64. https://doi.org/10.55630/dipp.2021.11.5

[4] Elisaveta Musakova. 2015. Writing and writers in Bulgarian manuscripts until the end of the 17th century. Sofia: National Academy of Arts. [in Bulgarian]

[5] Elena Kotseva. 1985. Development of the Bulgarian Cyrillic script in the XIV-XV centuries. In Boryana Velcheva, and Krasimir Stanchev (Eds.). Slavic Paleography and Diplomacy: lectures from the Second Seminar on Slavic Paleography and Diplomacy and reports from a scientific conference "The Decoration of Balkan Manuscripts up to the XVIII Century", Sofia, August 1983. Sofia: CIBAL, 23 – 47. [in Bulgarian]

[6] Marta Riparante, and Milena Davidović. 2023. The Case Study of the Serbian Copyist Ioan. On the Development of a Scientific Method for South Slavonic Palaeography. Digital Presentation and Preservation of Cultural and Scientific Heritage, 13, 127–136. https://doi.org/10.55630/dipp.2023.13.12

[7] Pyotr P. Chobitko. 2008. ABC book. The ABC of Old Russian writing. Saint Petersburg; Moscow: Gymnasium of St. Basil the Great. [in Russian]

[8] Desislava Paneva-Marinova, , Marco Scarpa, Marta Riparante, Maxim Goynov, Detelin Luchev, Lilia Pavlova, and Lubomir Zlatkov. 2022. Machine-readable Descriptive Structures to Study Medieval South Slavic Scriptoria and Scribes. TEM Journal 11, 2, 675-682

[9] Desislava Paneva-Marinova, Stoikov Stoikov, Lilia Pavlova, and Detelin Luchev. 2019. System Architecture and Intelligent Data Curation of Virtual Museum for Ancient History. SPIIRAS Proceedings 18, 2 (Apr. 2019), 444-470. https://doi.org/10.15622/sp.18.2.444-470.

[10] CiltIS. (n.d.). Web-based software platform for intelligent digital management and presentation of large data sets and knowledge from the field of culture, humanities and social sciences. Retrieved March 07, 2024, from https://cultis.math.bas.bg/bgAtul Adya, Paramvir Bahl, Jitendra Padhye, Alec Wolman, and Lidong Zhou. 2004. A multi-radio unification protocol for IEEE 802.11 wireless networks. In Proceedings of the IEEE 1st International Conference on Broadnets Networks (BroadNets'04) . IEEE, Los Alamitos, CA, 210–217. https://doi.org/10.1109/BROADNETS.2004.8