Anissava Miltenova, Ivan I. Iliev
Institute for Literature, Bulgarian Academy of Sciences

COMPUTER CORPORA AND PUBLIC CHALLENGES

Abstract. The report aims to show the process of all the work on building projects designed to meet societal challenges with new approaches to history, culture, the need for digital corpora of terminology and its functioning in foreign languages in non-Slavic countries. Regarding the project "Southern and Eastern Slavs: Diversity and Interaction of Written Cultures 11th-20th Century" are presented several stages of work: structuring the virtual museum, research of cultural heritage collections and analysis of sources, website design (https://sesdiva.eu/) and the search engine. The project "Scripta Bulgarica: Digital Library for Medieval Bulgarian Literature" gained popularity and applicability in education. It contains 506 terms in Paleoslavistics with an attached dictionary (in several Western European and Slavic languages: English, German, Russian and Bulgarian, as well as their respective terms in Old Bulgarian, Greek and Latin). The benefits of the usage the project and its perception among different users (students, teachers, museum specialists, journalists, etc.) are undoubted. The effectiveness of the projects depends on their sustainability and their continuation over time.

Keywords. digital corpora, virtual museum, cooperation, terminology, education, sustainability

Recently, the use of web-based corpora and easier access to digital reference sources in the field of humanities have become extremely topical. The present situation in society and education unquestionably showed the role and applicability of already existing projects, as well as the need to develop and improve them further. We will make a review of the activity in this direction at the Department of Old Bulgarian Literature of the Institute for Literature with the Bulgarian Academy of Sciences (BAS), bearing in mind the effect of the application of the projects and particularly their answer to contemporary public challenges.

Most generally, we could divide the implemented and ongoing projects into two groups:

(a) Projects that visualize historical literary facts, summarizing the development of science by certain topics;

(b) Projects that support education by compendia, dictionaries, and corpora of texts accompanied with analysis.

A case in point, typical of the first group (a), is the joint Bulgarian-Russian-Belgian project "South and East Slavs: Diversity and Interaction of Written Cultures 11th – 20th century" (https://sesdiva.eu/), funded under the ERA NET Rus Plus programme 2018–2020 (Bulgarian National Science Fund, BNSF (Bulgaria), contract No. DO 02/2/12.06.2018; Russian Foundation for Basic Research, RFBR (Russia), contract No. 18-512-76004/01.02.2018; Research Foundation – Flanders, FWO (Belgium)). ERA NET Rus Plus is a competition for a selection of countries from the European Union (*European Research Area*: https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/era_en) and Russia (in this concrete session) that have agreed to support projects of mutual interest through their research funds, uncovering new aspects of history and culture through joint research

teams. The head of the project was Institute for literature, Bulgarian Academy of Sciences and participants – Institute for Slavic studies, Russia and Department of Languages and Cultures, Ghent University, Belgium. The aim is to interpret the diversity and interaction between the written cultures of South and East Slavs applying the professional experience of the specialists from three research institutions with different traditions in the accumulation of knowledge and their interpretation using a variety of methodological approaches to create a freely accessible *virtual museum*. The project steps on the present increased exchange of knowledge between the countries of the European Union (including also Bulgaria, Slovenia, Croatia) and Russia, on the one hand, and between the Slav countries in the Western Balkans (Serbia, Montenegro, the Republic of North Macedonia, Bosnia and Herzegovina) and Russia – on the other. The project is a multidisciplinary one, the outcome is presented not only in traditional form (an anthology of research papers and documents) but also as a virtual museum based on digital archives of sources, photos, manuscripts and images, as well as illustrating texts.

The project is based on the understanding of a *common European heritage* of Southern and Eastern Slavs. It has the following objectives:

1) To establish a network of institutions in Belgium, Bulgaria and Russia engaged in academic research of Slavic literatures;

2) To endorse the application of new methods at the analysis of transborder phenomena and ideas reflected in cultural heritage;

3) To create a virtual museum – a format with easily accessible, visible facts to present the variety and interaction of written heritage and cultures;

4) To establish a basis for future cooperation between researchers and encourage future joint research through working meetings and a scientific conference;

5) To popularize in an accessible way objective knowledge about Eastern and Southern Slavs, particularly historical sources, among European citizens.

The significance of the project is determined by the contemporary processes of displacing humanities from the field of priority branches of science and, more specifically, the visible trends of marginalization and isolation of Slavic written cultures due to the lack of sufficient knowledge, as well as not understanding their history and cultural heritage. Thus world culture loses one of its integral parts, in many respects rewarding for the overall history of Europe and the world. Another aspect of the project is to outline the trend of acceptance or denial in the interpretations of the sources, of agreement or opposition in the overall concept of the literary process and its specific features in different periods.

The research teams from the Institute for Literature at the Bulgarian Academy of Sciences, the Institute of Slavic Studies at the Russian Academy of Sciences and the Department of Languages and Cultures at Ghent University, Belgium, were extremely successful in their work. Throughout the project, a total of 187 articles in Bulgarian, Russian and English, each supplied with 5 to 12 illustrations, were uploaded to the digital platform. This exceeds the number of the initially planned 140 articles. There are both reviews and concrete analyses and portraits of scholars. Users can also find important facts of interaction between the cultural heritage of Southern and Eastern Slavs, ones that were not well known or were wrongly interpreted. Essential processes in the development of literature and the contacts between its representatives have been revealed. New sources about written culture and its expressions have been analyzed, ones that are an irrefutable document of the variety, the interaction between cultures, and ones through which the most essential rises and falls in communication can be seen.
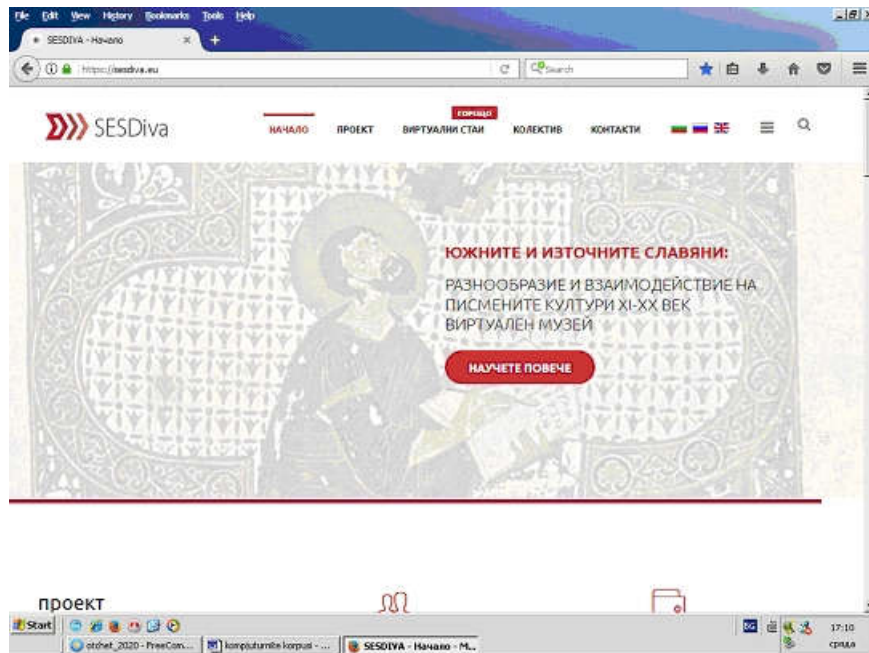
Figure 1. Initial page of the project South and East Slavs

Valuable conclusions in revealing new aspects of cultural heritage were made in several directions:

- new aspects of the essence and distribution of literary works that are examples of medieval Slavic writing for both the Eastern and the Western Slavs (Uspensky miscellany, Miscellany of Priest Dragol, Berlin Miscellany, the Synodikon of Tsar Boril, *Chrysorrhoas*, Elijah's Book, *Codex Suprasliensis*, and others);

- the migration of manuscripts and old printed books (11th-18th century) was tracked (works of Clement of Ohrid, of John the Exarch in Russian tradition, the liturgical book Prologue with Orthodox Slavs; Story about the transfer of St. Nicholas' relics from Myra to Bari; Russian redactions of the lives of St. John of Rila (15th-17th century), the works of Michael, Metropolitan of Kolasija, and others); as well as new data about the cultural exchange between the Southern and the Eastern Slavs and vice versa in the 19th and the 20th century (Ruthenian devotional songs among Southern Slavs; the 'Belgrade Circle': Pushkin, Lermontov, Gogol and Tolstoy in Serbian interwar comics; the early reception of Lev Tolstoy in Bulgaria; Russian reflections on Bulgarian literature in the 20th century, and others).

- valuable first-time observations are made about the centres for communication between the Southern and Eastern Slavs – the monasteries on Mount Athos, the St. Catherine monastery on Mount Sinai, the cities of Kyiv, Moscow, Sofia, and others.

- there are analyses of the collections of travellers on the Balkans and their contribution to preserving cultural heritage, as well as the fate of these collections.

- one exceptionally interesting item is the virtual "room" intended for the beginning of individual reading and featuring manuscripts used for the education of ruler families or commissioned by them.

- enriching the subject of the Russian émigrés on the Balkans with new materials is an unquestionable success of the project; this "room" contains 23 articles with illustrations, each introducing in academic circulation little known or unknown facts and interpretations. Until recently, this subject was completely untouched in both Russia and Bulgaria.
- the "room" dedicated to the National Revival period contains an impressive number of articles (40) with the results of projects and research of many years. Here, too, one can find conceptual articles on the origin and development of Pan-Slavism, as well as of the theory of the Illyrian movement in the Balkans.
- there is an analytic and systematic presentation of philological and historical science in the "room" set aside for scholars and writers who contributed to the elaboration of the problems of interaction of Slavic cultures in the 19th-20th century (38 articles). The virtual museum is configured in such a way that additions to it can and should continue to be made.
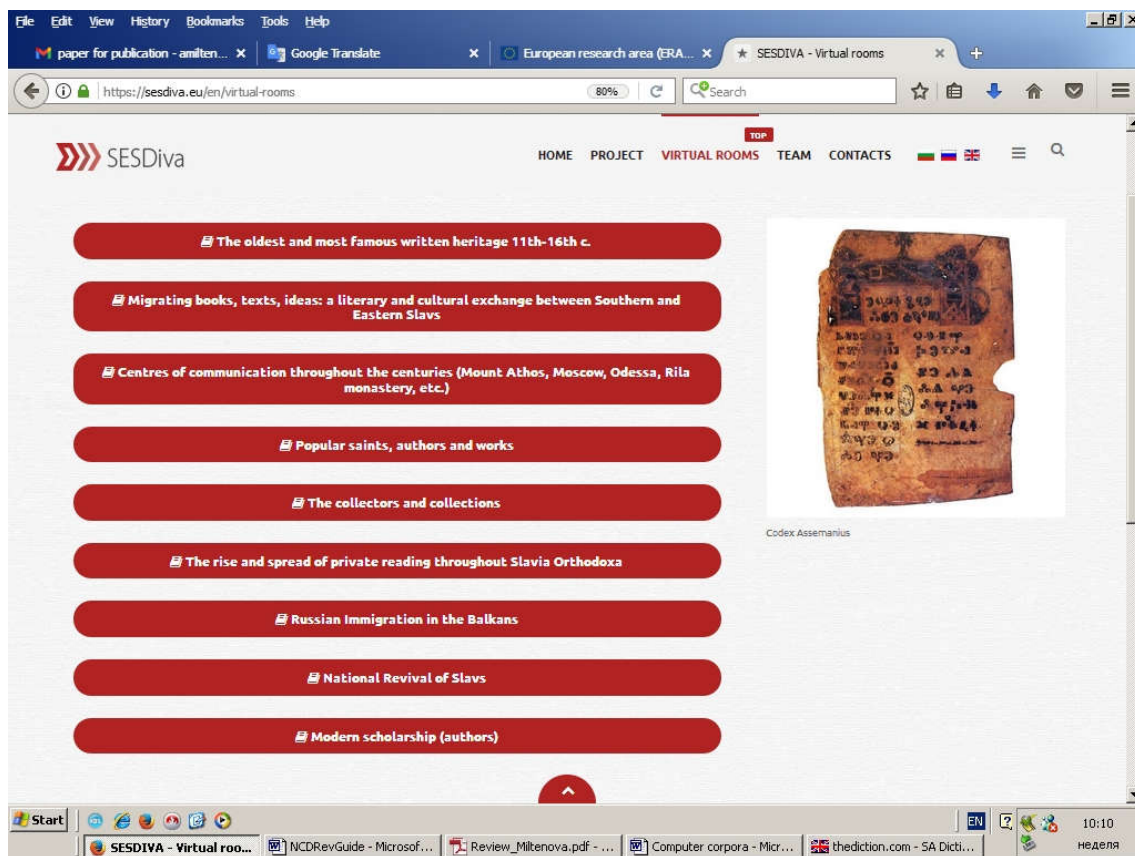

Figure 2. Virtual rooms in the project South and East Slavs

An international scientific conference was organized as part of the project on the subject of *Routes of Literary Communication between Eastern and Southern Slavs (11th-20th century)*, November 21-22, 2019 (Sofia, BAS), with participants from Belgium, Bulgaria, Russia, together with invited scholars from Serbia and Ukraine – over 40 in all. The reports showed and discussed important phenomena and processes, unstudied manuscripts, old printed books and archival documents. The proceedings

were published in 2020 (in Bulgarian, Russian and English, with summaries and tags in English)[1].

As to the second group mentioned above (b), the project "SCRIPTA BULGARICA: Digital Library of Medieval Bulgarian Literature" (http://www.scripta-bulgarica.eu/bg), financed by the National Science Fund (NSF) (contract No. ДФНИ-КО2/5 of 12.12.2014), acquired popularity and applicability in education. The objective of the project is *to create an electronic thesaurus of terms and a library of Old Bulgarian texts* (accessible through a special portal) integrating three main types of information:

– apparatus (terminology) in the field of Palaeo-Slavistic studies – both in the existing scientific literature and in the successful functioning digital corpora. The project responds to the need for an easily accessible reference of standardized terminology (*glossary*) in several West European and Slavic languages: English, German, Russian and Bulgarian, as well as their relevant terms in Old Bulgarian, Greek and Latin;

– articles of encyclopaedic type about translated and original works outlining the most popular genres and genre forms in hagiography, homiletics, hymnography, chronicles, narrative works, apocrypha and others;

– transcribed texts from Slavic manuscripts representing the translated literature from Byzantine Greek, as well as some original Bulgarian works;

– articles about the most frequently translated Byzantine authors (e.g. Gregory the Theologian, St. John Chrysostom, St. Basil the Great, Athanasius of Alexandria, Anastasios of Sinai, and Hesychius of Jerusalem, among others), whose works (genres and genre forms) modelled and formed Orthodox Christian written culture.
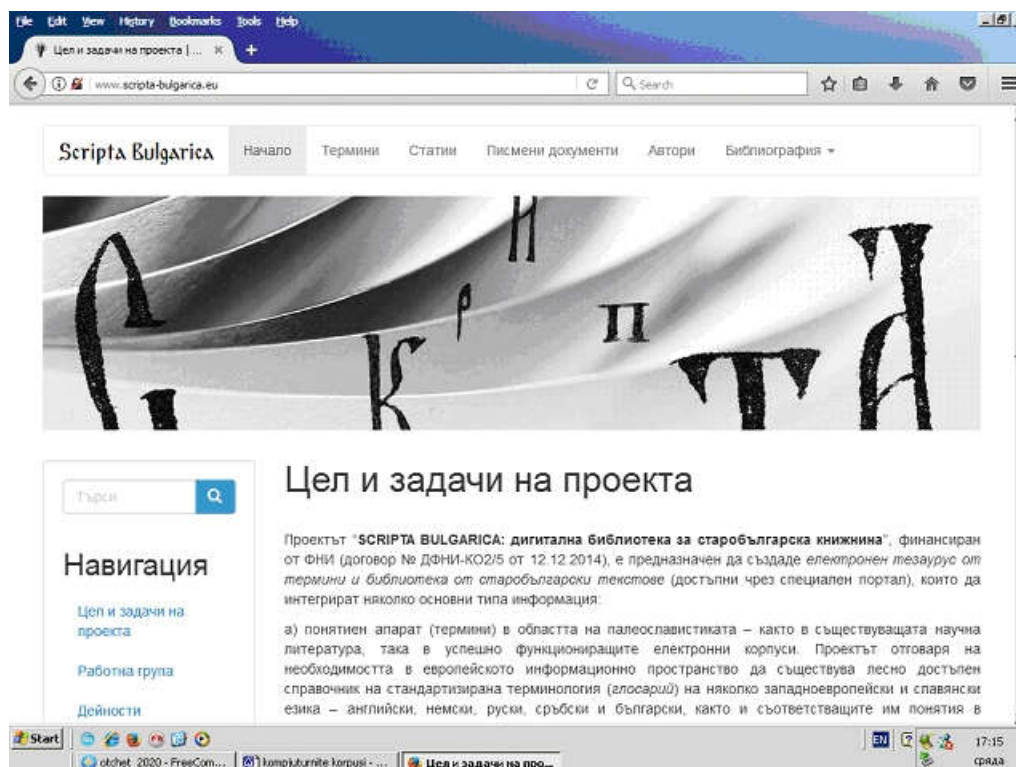

Figure 3. Initial page of the project Scripta Bulgarica

The portal was developed with the joint efforts of specialists from the Institute for Literature, BAS, and the Faculty of Mathematics and Informatics at Sofia University. It steps on the experience and materials accumulated as a result of several national and international projects of the Department of Medieval Bulgarian Literature at the Institute for Literature since 1995. These projects include a *Repertorium of Old Bulgarian Literature and Letters* (1995–2015) – within the framework of the international *Repertorium* initiative, launched by the Department which is its basic organisation. The international projects of the Department to create digital resources in Palaeo-Slavonic studies were implemented in partnership with the British Library (London) (1995-2012), the Library of the Russian Academy of Sciences in St. Petersburg (2004-2010), the universities in Gothenburg and Uppsala (Sweden) (2008-2015), the Slovenian Academy of Sciences and Arts, and the Institute of the Lithuanian Language in Vilnius, among others. The Department has a traditional partnership with Sofia University, both in the case of the *Repertorium* initiative and in the development of the *Slavica Sanctorum* digital encyclopaedia (http://www.eslavsanct.net/), as well as the *Libri Slavici* project (http://slav.uni-sofia.bg/naum/librislavici). As a result of these projects, digital resources were created and materials were collected which constitute *a full-text information corpus* uniting linguistic, archeographic, paleographic, codicological and textological data.

The first task of the project was to develop the architecture of the portal and the structure of the terminological articles, the articles presenting written documents, as well as those containing metadata about the documents and articles about authors. In the course of its work, the team was guided by the target groups of the users of the digital product, more specifically: students (bachelor, master and doctoral students) who have to receive adequate information about the terms in the process of their education; specialists (lecturers, researchers and librarians) who use medieval Slavic manuscripts and archival material in digital form in their work; a wide cultural audience interested in Slavic literary tradition and the fields of cultural heritage related to it. It is important to note that the usual approach of a preliminary outline of taxonomy and definition of hierarchical relations between the various terms is not used in this project. Instead, there is a digital platform that allows the definition of the relations and links in the process of loading the material and the subsequent generation of different types of classifications. This forms "nests" of terms mutually connected by certain features, which in turn are connected with the transcribed texts and the metadata attached.

At creating the prototype of terminological articles the team was guided by the expert assistance of Prof. Andrei Boyadzhiev (Sofia University) and Prof. David Birnbaum (University of Pittsburg, Pennsylvania, USA). Some foreign libraries and research institutions had identically directed projects and we contacted them and exchanged experience. In discussion with them, as well as with several other colleagues (e.g. Prof. Margaret Dimitrova, Sofia University; PhD Adelina Angusheva-Tihanov, University of Manchester, UK; PhD Vera Podkovirova, Library of the Russian Academy, St. Petersburg; PhD Marina Krutova, Russian State Library, Moscow; Prof. Ralph Cleminson, University of Oxford, UK, and others) we have studied and discussed some issues relevant to the terminology. I will try to mention some of them here, formulated by A. Boyadzhiev at the establishment of the computer model to describe a term:

Name of text/Term
- What does a text call itself in the manuscripts and old printed books? What is the academic tradition (term) in naming this text? On what is that based? What is the connection with other medieval traditions (Byzantine, Latin or other)?

- What is the function of the text and what is the sphere of its perception? Is it a sermon, enkomion, prayer a hymnographic form or a part of didactic literature (story, exemplum, apophthegma, wise sentence)?
- To what degree can the designation in the manuscripts also be accepted as an academic name (term)? How is the academic name (term) formed and what criteria should it meet?
- In what way are this text cited in the Middle Ages and today, how is it quoted?

Text/term author
- Who created the medieval text? What are the peculiarities of the author's name? In what way was it formed? By what rules is it transferred from one language to another, from Byzantine Greek to Old Bulgarian or vice versa?

Text birth and history
- In what way was a text compiled? What are the concepts of its translation, redaction, compilation and revising according to existing research?

Structure, composition and content
- What are the separate elements of the text? What in the text can orient us about the initial composition of the work? What are the various rhetorical tropes and figures in the text and how are they called? What is their role?

Text language
- What is the language of a text? Where does the language stand in comparison with the history of the individual Slavic languages? Is there a difference between the language of the individual text and the language of the manuscript as a whole?

Manuscript and book
- How did a text reach us? How can we describe the book or scroll that preserves the text in digital form? What is the story of the codex until today?

Text and medieval man
- Who reads and pronounced this text? What were the people who officiated at the service in the church, chanted or read those medieval words? What is the audience before which a text is pronounced (performed) called?

Text and act
- What happened while a text was being read? What were the individual acts during the reading of a text (for example during divine service)?

Text and time
- When was a text read? Did it change depending on the time when it was pronounced? Did its name and functions change with time?

Text and other forms of medieval art
- How was medieval text transformed into an icon, a mural, an illustration, an illumination or a musical work? Which terms can express this transformation?

To date, the project includes 506 terms in the field of paleoslavic studies (codicology, palaeography, textology, history of literature). In some cases, we have clarification of terms in the field of genres and genre forms in medieval literature covering both hagiography, hymnography and homiletics, and the genres out of the divine service. Attention has been paid to the types of manuscript books and the types of texts, keeping the principle of following an approach at which the terms of medieval literature and culture are explained through the information contained in the original sources, but along with that, their translations are tracked, as well as the academic tradition in contemporary European countries. Reference to an authoritative source, from which the material is taken and which serves as a reference, is an obligatory element of the translation of the term into the contemporary languages in the glossary of terms. As we know, the specifications attached to standards serve as a terminological

reference in technical fields. The subjects which the humanities study use sources of an encyclopaedic character, library systematizations, as well as definitions of individual authors as sources of reference.

The structuring of the computer model to present texts from medieval manuscripts and their comments were easier to achieve. So far, the section of the project for transcribing texts from Old Bulgarian literature has come to include 119 articles, and that with encyclopaedic articles about source texts – 99 articles. The Byzantine authors included in the project with articles number 39. Additions of terms and current bibliography continue actively in all sections. We use the Unicode font *Bukyvede* adopted by the consortium. Since 2008 were have been in connection with the commission that approves the use of fonts including Pld Cyrillic and Glagolitic in both traditional publications and web-based platforms (The Medieval Unicode Font Initiative: https://mufi.info/m.php?p=mufi).

The *Scripta Bulgarica* project also includes the unique for such initiatives in the humanities connection between the standardized apparatus for description, research, publication and translation of medieval texts with the keywords and the bibliographic descriptions. This allows the integration of the apparatus of scientific meta-data and the reference bibliographic apparatus in an integral whole, the advanced phase of the project seeing a developed ontology of terms. The portal and the search programme were created through the *Drupal* (https://www.drupal.org/) platform for digital platform development. The work of the specialists in informatics from Sofia University was exceptional: Prof. Krassen Stefanov, Assistant Prof. Mihail Peltekov, Assoc. Prof. Alexander Grigorov and doctoral students from the Faculty of Mathematics and Informatics.

In conclusion, it should be pointed out that work on the development of computer corpora has made considerable over the past decade. Cooperation between scholars, exchange of knowledge, developing technological approaches and the opportunities for the public to have access to the results quickly are unquestionable success. The institutions and the working groups in them still experience an extreme shortage of systemic project funding to ensure their sustainability and continuation in time. The need of training young specialists with knowledge of both philology and history, hand in hand with classical languages and interest in digitization and information technologies is pressing. The last competitions for doctoral students and young scientists showed that such specialists are becoming increasingly rare. This probably calls for a future clear-cut strategy for action in this direction.

References

[1] Birnbaum, D. Standardizing characters, glyphs, and SGML entities for encoding early Cyrillic writing. – *Computer Standards & Interfaces,* 18:3(1996), 201–252.
[2] Birnbaum, D. Computer-Assisted Analysis and Study of the Structure of Mixed Content Miscellanies. – *Scripta& e-Scripta*, 1(2003), 15–54.
[3] Birnbaum, D. Analyzing and visualizing the structure of medieval encyclopedic works with XML-related technologies.
[4] Birnbaum, D. Macro-Level and Micro-Level Structural Analysis of Mixed Content Miscellanies. – In *Studia Caroliensia: Papers in Linguistics and Folklore in Honor of Charles E. Gribble.* Ed. Robert A. Rothstein, Ernest Scatton, and Charles E. Townsend. Slavica: Bloomington, Indiana, 2006, 41–56.
[5] Burnard, Gougerot, Lalou, Robinson 1999: Burnard L., M. Gougerot, E. Lalou, P. Robinson. Toward a European Standard for Manuscript Description: The MASTER project. [1999] (http://www.tei-c.org.uk/Master/Hermes/).

[6] Cleminson, R. Research Project on Computer Supported Processing of Slavonic Manuscripts and Early Printed Books. – In *Annual of Medieval studies at CEU,* 1997–1998. Budapest, 1999, 186–187.

[7] Computer Processing of Medieval Slavic Manuscripts. First International Conference, Blagoevgrad, 24–28 July 1995. *Proceedings.* Joint with D. Birnbaum, A. Boyadzhiev, M. Dobreva. Sofia, 1995.

[8] Computational Approaches to the Study of Early and Modern Slavic Languages and Texts. Editors: D. Birnbaum, A. Miltenova, and S. Slevinski. Sofia, 2003.

[9] Computer Applications in Slavic Studies**. Proceedings of Azbuky.Net International Conference and Workshop, 24–27 October 2005. Ed. A. Miltenova, D. Radoslavova, E. Pancheva. Sofia, Bulgaria. Sofia 2006.

[10] Medieval Slavic Manuscripts and SGML: Problems and Perspecti*ves*. Ed. A. Miltenova, D. Birnbaum. Sofia, "Marin Drinov" Publishing House, 2000.

[11] Miltenova, A. Sofia Corpus of Data of Slavonic Manuscripts. – *Slavic & East European Information Resources*, 3, 2002, 2/3 (Libraries in Open Societies: Proceedings of the Fifth International conference ICCEES in Talin, Estonia, July 2000), 81–86.

[12] Miltenova, A., A. Boyadzhiev, D. Radoslavova. A Unified Model for the Description of the Medieval Manuscripts? – In *Computational Approaches to the Study of Early and Modern Slavic Languages and Texts*. Sofia, 2003, 113–125.

[13] Miltenova, A., A. Boyadzhiev, St. Velev. *Repertorium* of Medieval Slavic Literature: Computer and Philological Standards. – *Palaeobulgarica*, XXII, 1998, N 2, 50–69.

[14] Miltenova, A., A. Boyadzhiev. An Electronic Repertory of Medieval Slavic Literature and Letters: a Suite of Guidelines. – In *Medieval Slavic Manuscripts and SGML: Problems and Perspectives*. Sofia, 2000, 44–68.

[15] Proceedings of Extreme Markup Languages[®] – Master Bibliography: http://conferences.idealliance.org/extreme/biblio.html

[16] Slovo: Towards a Digital Library of South Slavic manuscripts. Proceedings of the international conference, 21–26 February 2008, Sofia, Bulgaria. Ed. H. Miklas, A. Miltenova. София, 2008.

[17] Дигитални и аналитични подходи към писменото наследство. Материали от 7-мата международна конференция El'Manuscript „Писменото наследство и информационните технологии“, 2018. Ред. Х. Миклас, В. Баранов, К. Хоукинс, Ю. Фуксбауер, А. Милтенова. София: Гутенберг, 2019.

[18] Маршрути на книжовно общуване между източните и южните славяни (XI-XX век). Международна научна конференция, 21-22 ноември 2019 г. София. Съст. и ред. А. Милтенова, И. Калиганов. София: ИЦ „Боян Пенев“, 2020. ISBN 978-619-7372-24-3.

[19] Преоткриване: Супрасълски сборник, старобългарски паметник от X век. Изток-Запад, Съст. и ред. А. Милтенова. София, 2012.

amiltenova@gmail.com
ian_ilye@abv.bg