

Computer Supported Processing of Slavic Manuscripts in Bulgaria: Repertorium Initiative

Автоматизирана инициатива
18.07.2008 г.
Инициатива инициатива 20.07.2008 г.

Описание на инициативата PDF

Computer-supported research and teaching in medieval studies in Bulgaria has been growing up at an increasing pace over the past decades with implementation of new methods in this area. The beginning was the Bulgarian-American project "Computer Supported Processing of Old Slavic Manuscripts" funded by IREX – Washington (1994–1995). At that time a new type of software was built. It was based on the Standard Generalized Markup Language (SGML), accepted by the International Standards Organization (ISO), and, especially, in its TEI implementation. This undertaking was built on the framework developed within the TEI by creating a set of modifications for manuscript description.

The major template (Template for Slavic Manuscripts, TSM) was developed in the process of the teamwork of Prof. David Birnbaum of University in Pittsburgh, USA, Assoc. Prof. Andrej Boyadzhiev from Sofia University, Bulgaria and Prof. Anissava Miltenova from Institute of Literature BAS. The initial system for encoding of medieval Slavic texts (TSM) was discussed in an international conference that took place in Blagoevgrad (24th–28th July, 1995). The philosophy of SGML helped to settle some well-known misunderstandings among paleoslavists concerning philological questions of terminology, inventory of units, character sets and data structure.

The movement from a relational database management system (RDBMS) framework to SGML marked a significant reorientation in the conceptualization of computer-assisted manuscript description. More importantly, though, our SGML-based undertaking was oriented towards preparing manuscript descriptions that might be suitable for printing, electronic rendering, and searching, as was the case with the RDBMS approach.

Within the framework of Repertorium project, over three hundred fifty manuscripts were processed in the Institute of Literature BAS by using TSM system in the SGML and, after that (2001), in XML environment. Scientific papers and indices of the pilot project were published under the title Medieval Slavic manuscripts and SGML: Problems and perspectives (2000) and in the book Computational Approaches to the Study of Early and Modern Slavic Languages and Texts (2003). The Internet presence of the Repertorium Initiative is located at <http://clover.slavic.pitt.edu/~repertorium/>.

We consider our prior close collaboration with specialists in Slavic and general humanities computing (e.g., Institute for Computational Linguistics, Pisa, Italy; and Central European University, Budapest, Hungary) to be one of the strongest features of our both evaluative feedback on our proposals and means for ensuring that our results will reach authoritative figures and institutions. In 2003 a joint contract was signed between the British Library and the Central Library BAS, having as a major target the processing by our group of professionals a certain collection of Slavic manuscripts from the BL. The same type of project is going on with Sweden (official partner Royal Academy of Sciences and as sub-partners all major Swedish libraries with Slavic manuscript collections). Recently a similar contract was signed with the Library of the Russian Academy of Sciences in St Petersburg, Russia.

The Repertorium project activities nowadays are concentrated on the following main fields:

The first of these is to continue enlarging the corpus of analytical descriptions of manuscripts in Bulgaria and, ultimately, elsewhere (an “electronic catalogue”) both according to chronological and to thematic principle.

The second is to add facsimiles in the form of computerized picture files, linked to the relevant entries in the catalogue database.

Quite an important field is the development of auxiliary materials and databases (“electronic reference books”) for the study of Slavonic manuscripts, in many cases by extrapolation of the data assembled in the other phases of the project. Part of this field consists of bibliographic database for the described sources.

As a necessary part of the manuscript description the system for cataloguing of microforms is developed in CL BAS.

The computer processing of Slavic manuscripts has been discussed at the 12th International Congress of Slavists, Krakow, 1998. Participants from Bulgaria, Byelorussia, Czech Republic, Finland, Italy, Macedonia, Great Britain, the US, etc. put on discussion some mainstream questions in the field. One of the results from this discussion was the establishment of a Commission to the Executive Council of the Congress for Computer Supported Processing of Slavic Manuscripts and Early Printed Books. The Commission organized a special panel at the next congress in Lubljana (2003).

The last phase of this process is characterized not only by the accumulation of still more manuscript descriptions, but also by the conversion of our materials from SGML to XML. The transition to XML was dictated by the remarkably broad acceptance of XML within the electronic-text community, and particularly by its adoption by the TEI, initially as an alternative to SGML, but ultimately as a replacement for it. We have currently converted over one hundred manuscript descriptions from our initial corpus of three hundred; the rest will be converted in time, and all new descriptions are being created directly in XML.

Because the Repertorium Initiative goes beyond manuscript studies in seeking to provide a broad and encyclopaedic source of information about the Slavic medieval heritage, it also incorporates such auxiliary materials as bibliographic information and other authority files. In this capacity the Repertorium Initiative is closely coordinated with three other projects: the project for Authority Files, which defines the terms and ontology necessary for medieval Slavic manuscript studies; Libri Slavici, a joint undertaking of the Bulgarian Academy of Sciences and the University of Sofia in the field of bibliography on medieval written heritage; and identifying the typology of the content of manuscripts and texts with the aid of computational tools. All three of these share the common structure of the TEI documents and use a common XSLT (Extensible Stylesheet Language for Transformations) library for transforming documents to a variety of formats (including XML, HTML [Hypertext Markup Language], and SVG [Scalable Vector Graphics]) thus providing a sound base for the exchange of information and for electronic publishing.

