Преглед НЦД 5 (2004), 49–58

Pavel Pavlov
(Faculty of Mathematics and Computer Science, Sofia)

XEDITMAN: A XML EDITOR FOR MANUSCRIPT DESCRIPTIONS AND ITS IMPLEMENTATION FOR CATALOGING OF BULGARIAN MANUSCRIPTS

Abstract. We present the specialized editor XEditMan *(XML Editor for Manuscript Data)* which is an XML-oriented tool for editing and browsing catalogue descriptions of mediaeval manuscripts. It offers a friendly interface for entering data on mediaeval manuscripts; visualisation and queries to the descriptions already available. The descriptions are compatible with the document type definition (DTD) structure suggested by the project MASTER (Manuscript Access through Standards for Electronic Records) and adopted by the Text Encoding Initiative. During the data entry the elements which are filled in appear in a sequence which is adopted in the manuscript cataloguing practice. The interface is in Bulgarian and this facilitates preparing electronic descriptions by people who are not acquainted in detail with the DTD structure. The tool can be used also for visualization of single descriptions in two modes: complete descriptions or user-selected group of elements. Comparative study of multiple descriptions is achieved through database queries. Currently, 300 descriptions are available. By December 2004 a collection of 800 descriptions on Mediaeval Bulgarain manuscripts stored in Bulgaria will be prepared.

Keywords: XML, manuscript cataloguing, visualisation

1. Introduction

In the recent years we witness the growing interest to the presentation of cultural heritage and access to it in electronic form. One of the essential matters in this field of work is related to organizing the workflow of data entry in a way which guarantees high speed, accuracy and manageability of the efforts.

Despite digitisation itself, a special care should be taken for the pre-digitisation works, the most typical of them being the electronic cataloguing of collections.

The models underlying mediaeval manuscript data were in the focus of the scholarly and librarian community for the last three decades. In this direction an approach which attracted great number of followers is the use of mark-up languages which nowadays is most typically done in practice through XML.

In this paper we present the development of an integrated tool which makes possible entry of data, visualisation and execution of queries on catalogue data on Old Bulgarian manuscripts. Data are entered in XML format, but in order to offer adequate processing capabilities are also stored in a database.

2. Bulgarian Experience in Electronic Cataloguing of Manuscripts

The idea to use a markup language for manuscript descriptions goes back to the 1990es. With the advent of mark-up languages, a team in Bulgaria suggested in 1994-95 a structured description of manuscript data built as an extension of Text Encoding Initiative[1]. A project called *The Repertorium of Old Bulgarian Literature and Letters* was started as "…an archival repository capable of encoding and preserving in SGML (and, subsequently, XML) format of archeographic, palæographic, codicological, textological, and literary-historical data concerning original and translated medieval texts represented in Balkan Cyrillic manuscripts"[2]. For the needs of the project, a computer model based on on SGML was developed. Currently there are 300 manuscript descriptions which should be made available on the project website[3].

In the late 90es, the National Library "St. Cyril and St. Methodius" and the Institute of Mathematics and Informatics became associated members of the MASTER project *(Manuscript Access through Standards for Electronic Records)* supported by the EC[4]. Within this project, a TEI-conformant DTD for mediæval manuscripts was developed with the ambition to answer the needs of European repositories, and software for making and visualising records on manuscripts. The document type definitions developed for the MASTER project were adopted by the TEI in May 2003 with minor revisions.

In the Repertorium project, data were entered through Author/Editor software product of SoftQuad company, a predecessor of HoTMetaL and currently available XMetaL editors. In the data entry process, users were seeing all elements from the description on the screen (surrounded by the SGML delimiters, e.g. <P> </P>) which formed long list spread on several screens. This was not very convenient, if we also add that the appearance of elements followed the structure of the DTD which is not the same as the sequence of elements natural for the people working with mediaeval manuscripts. The organization of work was oriented towards one specialist working on one description, which produced results of different quality in the group of almost 10 specialists working on the descriptions [Dobreva, Jordanova, 2000]. The description data were entered in English which made them usable by English language speakers. To enter fragments of Old and Middle Bulgarian texts a designated font was created, and in data entry the LANG attribute was assigned to elements containing text in Old or Middle Bulgarian while for all other languages was supposed that they contain texts in English.

The experience of the pilot catalogue descriptions within the MASTER project was different in two directions: the data were entered in both Bulgarian and English with the idea that this will serve larger research community, and the editor used for the tests was NoteTabLight[5]

---

[1] http://www.tei-c.org/—*Text Encoding Initiative Website*

[2] *Repertorium*, http://clover.slavic.pitt.edu/~repertorium/index.html — website of the *Repertorium of Old Bulgarian Literature and Letters*

[3] On April 25, 2004 there was still a message that link is disabled for file update.

[4] MASTER, http://www.cta.dmu.ac.uk/projects/master/, website of the MASTER project.

[5] http://www.notetab.com/ntl.php, NoteTabLight – free editor offered as an alternative to the commercial professional editor NoteTab Pro.

This previous Bulgarian experience made clear that the software used in the previous cases was not convenient enough for the staff which took part in data entry. In both cases the tools were allowing data entry and visualization of one particular manuscript, but were not capable of processing sets of catalogue descriptions. Thus we came to the conclusion that there is a necessity of building a new tool, which would combine the most typical tasks related to manuscript cataloguing. Our solution is presented below.

3. XEditMan: an Integrated Tool for Entering Data, Visualisation and Executing Queries

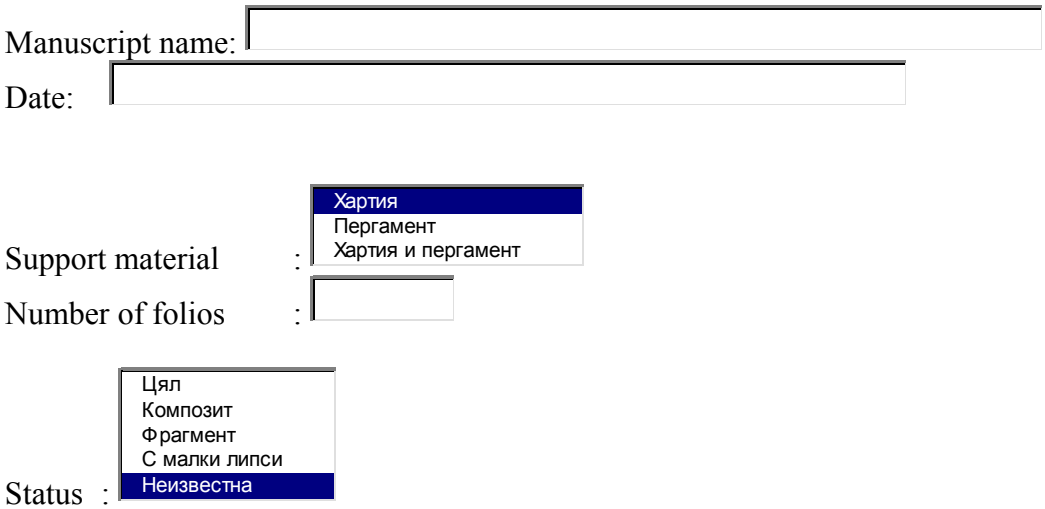
XEditMan, XML Editor for Manuscripts, is an integrated tools which offers support for the following activities:

- Editing a new catalogue description
- Editing an existing catalogue description
- Visualisation of one complete description
- Visualisation of user-selected data from a description
- Executing queries over a group of catalogue descriptions.

The editor is currently oriented towards the use of the MASTER document type definition (DTD) for manuscript descriptions adopted by TEI[6]. The editor has an interface in Bulgarian language which is organized as a librarian worksheet (see Fig. 1), the names of elements are translated for reader's convenience to English. The data chosen form the MASTER DTD represent a minimum set of data. We are aware that this form is not allowing entering sophisticated descriptions, but it is convenient and efficient for entering basic manuscript data.

To facilitate data entry, we included three types of fields:

- Fields where content is entered by the user
- Fields with supplied values in the form of a list of elements
- Fields where a value is entered in advance, but could be edited by the user (this is usually the value 'no information').



[6] The relevant materials can be found on http://www.tei-c.org.uk/Master/Reference/, last accessed on April 25, 2004.

Watermarks : Няма сведения

Collation: Няма сведения

Quire structure: Няма сведения

Vertical size (mm):

Horizontal size (mm):

Number of Scribes:

Name: Неизвестно

Script: Няма сведения

Ink: Няма сведения

Extent : Единствен
Основен
Допълнителен

Description of the script: Няма сведения

Place of origin: Няма сведения

Provenance: Няма сведения

Acquisition: Няма сведения

Linguistic and Orthographic Features:

Decoration:

Binding: Няма сведения

Current condition: Няма сведения

Content (general description):

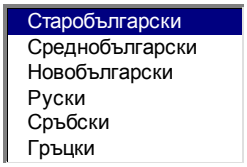
Number of texts:

Scope:

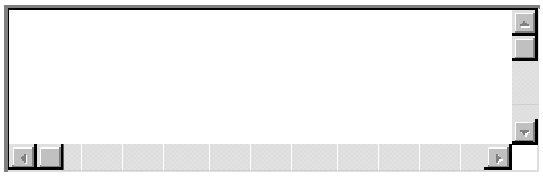
Author:

Title:

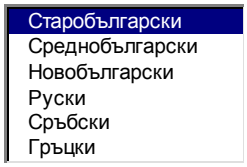
Written in:



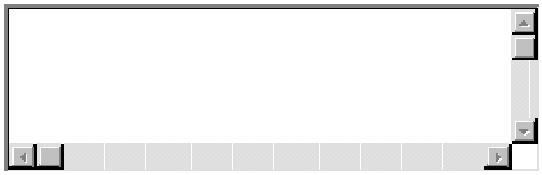
Incipit:



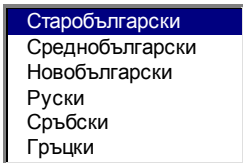
Written in:



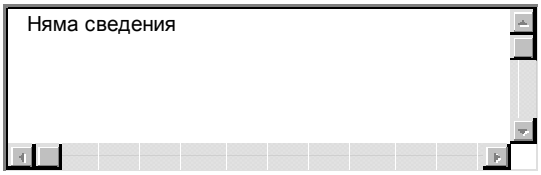
Explicit:



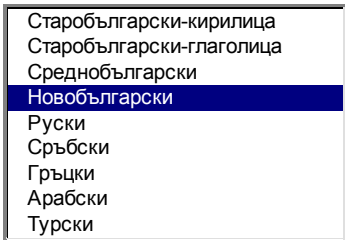
Written in:



Marginal notes:

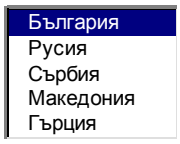


Written in:

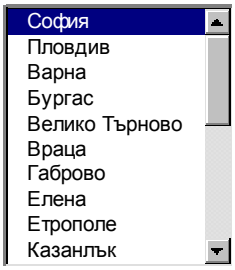


Repository data

State:



City:



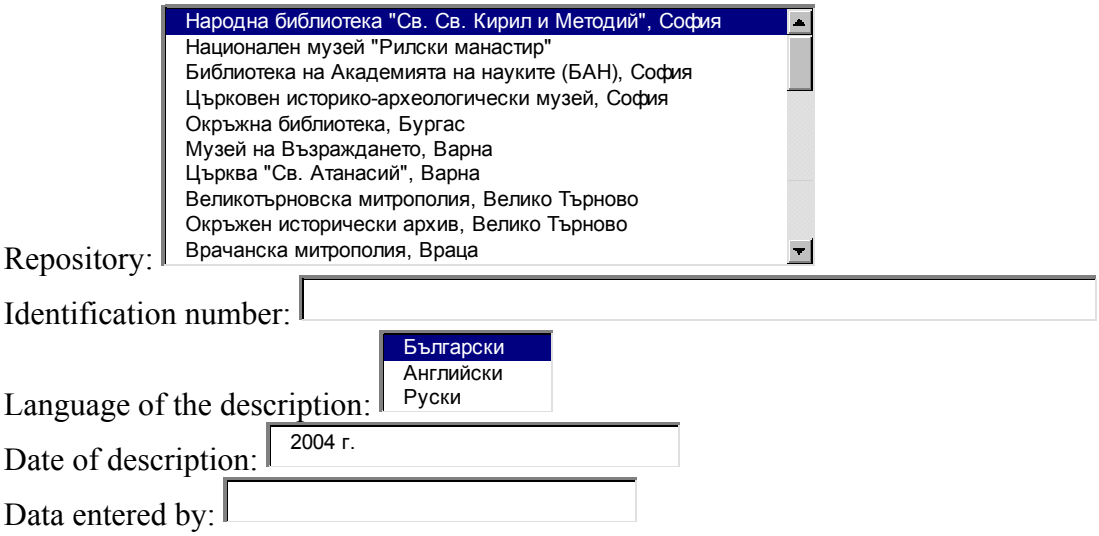


Fig. 1. Data Entry Worksheet

The user enters data and presses a button which produces the description according to the DTD structure. For elements which can be used repetitively, such as scribes and texts, during the initial data entry the user supplies the number of repetitive elements needed and during the editing of an existing document he/she can enter all values. These elements are preceded by a red bar in the interface in order to be easily distinguished. For **visualisation** of a particular manuscript, XEditMan offers a view similar to the one of data entry. Fig. 2 shows an example of a visualised fragment. A special care was taken to visualise Old Cyrillic texts properly.

Since scrolling through a description is not always of benefit to the users interested in particular details, XEditMan offers a possibility to see selected data from a particular description. The user can choose the data he is interested in (see Fig. 3). These data appear following the pattern which is already familiar to the user – the only difference is that in this case not all available data are shown, but only the selected ones (see Fig. 4).

For greater flexibility of manuscript description use, the data from the XML description are also stored in a database (in the particular case, this is Microsoft Access). This allows executing queries, for example, of all manuscripts from a specific repository; sorted by date of origin or location, etc. On Fig. 5 we present the result of a query showing works written by different authors in the manuscript electronic catalogue collection.

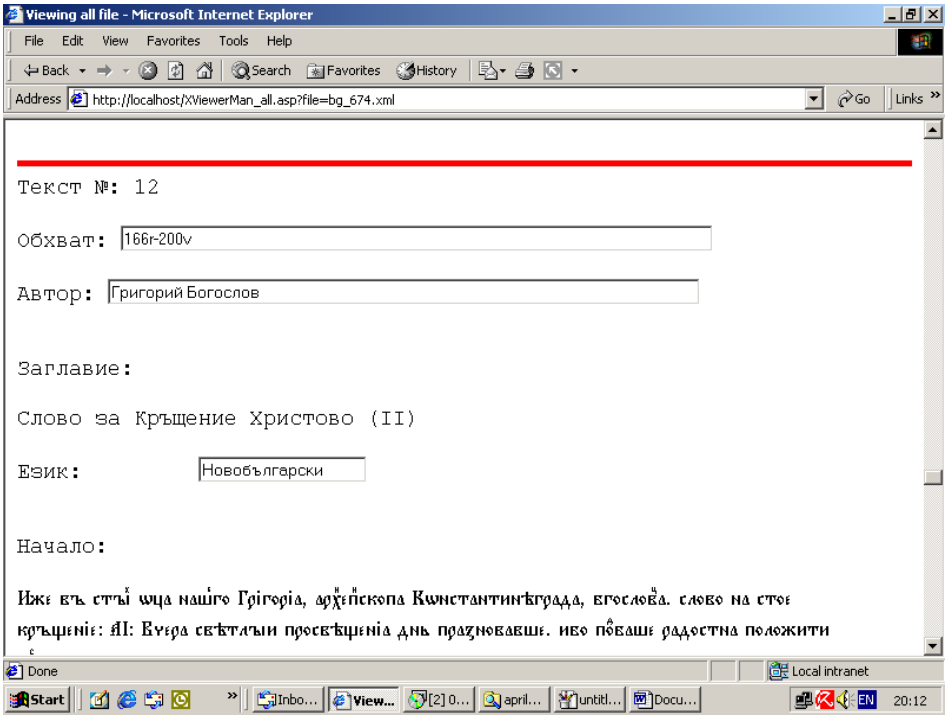


Fig. 2. Example of a visualised fragment in XEditMan

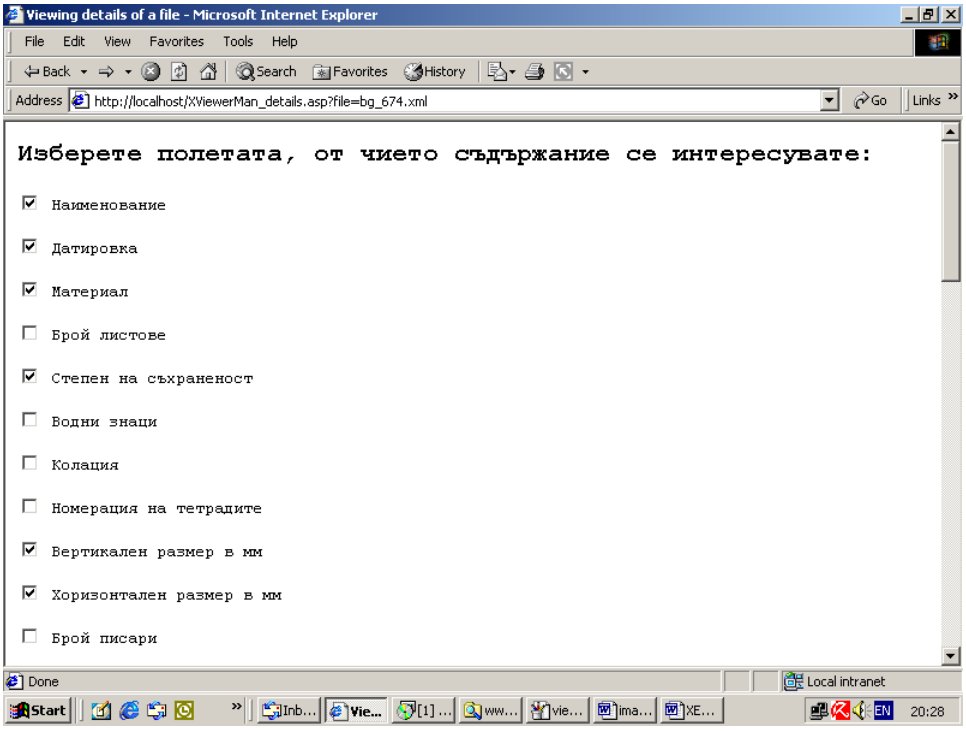


Fig. 3. An example of choosing data for visualisation

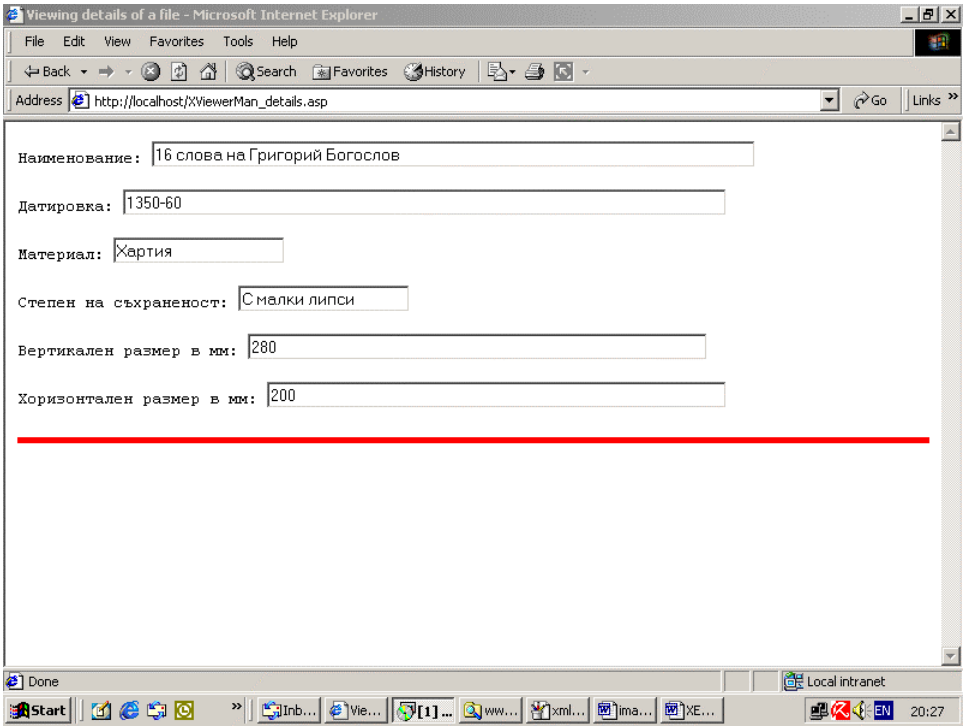


Fig. 4. Visualisation of data selected by the user

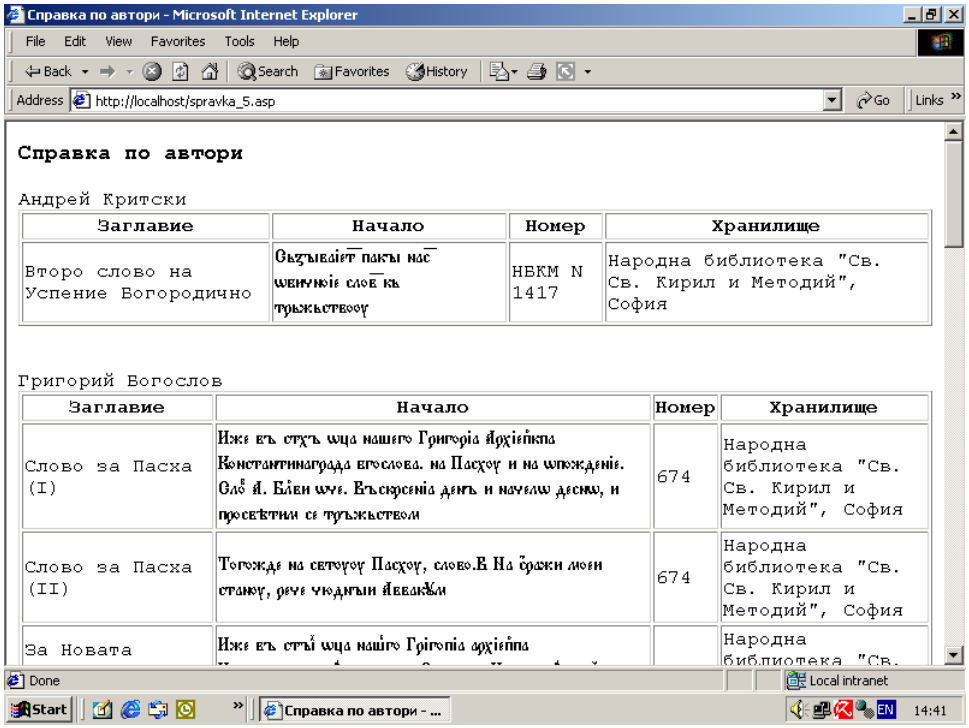


Fig. 5. An example of a query on all available manuscripts in the electronic catalogue

4. Conclusions

We presented briefly the functionality of XEditMan: an integrated tool which allows users to enter data on mediaeval manuscripts; to edit them; to visualise complete descriptions, as well as parts of them; and to execute queries on the collection of catalogue descriptions as a whole. This tool is currently used in the Institute of Mathematics and Informatics for the first mass data entry on Old Bulgarian manuscripts. We hope that this will lead to better access to the manuscript heritage of Bulgaria.

This works was done with the support of the project Knowledge Transfer for the Digitisation of Scientific and Cultural Heritage to Bulgaria, FP6 contract N MTKD-CT-2004-509754.

References

[1] M. Dobreva, M. Jordanova, *Some Psychological Aspects of Computer Modeling of Complex Objects,* In: A. Miltenova, D. Birnbaum (eds.), *Medieval Slavic Manuscripts and SGML, Problems and Perspectives,* Prof. M. Drinov Academic Publishing House, Sofia, 2000, pp. 295–310.

mailto:pavlovp@fmi.uni-sofia.bg