

Модели описания дигитализированного письменного наследия: состояние проблемы

Models for the Description of Digitized Literary Heritage: State of the Art

Автор: Н. В. Ганчева, кандидат наук, главный ассистент

E-mail: nelly_vasileva@abv.bg

Название организации: Кирилло-Мефодиевский научный центр
Болгарской академии наук, София, Болгария

Ключевые слова: рукописи, современные печатные издания, модели
электронного описания, TEI, XML

Summary

The main task of the models for the description is to accurately reflect the specificities of different types of documents. In the report are indicated the main problems with respect to the electronic description of mediaeval manuscripts, on the one hand, and of contemporary publications, on the other. An important accent in it is the delineation of characteristic of our times two points of view about development of models for the description. According to one the perspectives are in convergence and integration of different models in a common scheme. The other point of view is completely opposite. It is based on understanding of the need to keep the entire range of models with the purpose to correct presentation of the characteristics of described materials.

Основной текст тезисов:

Проблема электронного описания дигитализированного письменного наследия в последние десятилетия является постоянным предметом обсуждений и исследований, в которых принимают участие как ученые, так и специалисты в области информационных технологий и библиотечные специалисты. В этих дискуссиях представлены различные точки зрения на проблему, в основе которых лежат различия между типами документов (рукописи, старопечатные книги, архивные документы, современные печатные издания), характеристики которых должны быть адекватно отражены в выработанных для них моделях описания. Объектом внимания в настоящем докладе являются вопросы электронного описания средневековых рукописей, с одной стороны, и современных печатных изданий, с другой. Они имеют различные параметры, поэтому необходимо создавать отдельные модели их описания, которые выполняет различные функции.

В начале укажем на несколько основных проектов с активным болгарским участием, направленные на разработку моделей электронного описания средневековых рукописей. Первые шаги были сделаны в 90-х годах XX в. в рамках Международного информационного центра по изучению источников балканской истории (CIBAL – International Information Centre on the Sources of Balkan History) в связи с решением о создании сводного каталога средневековых рукописей. В соответствии с этим решением был разработан первый формат для электронного описания рукописей, использующий базу данных в среде ISIS (Integrated Set of Information Systems), который, однако, не сумел ответить на требования специалистов, так как возникли проблемы, связанные с последовательностью полей для описания и их наименованиями, а и сам электронный продукт показал известную ограниченность. В конце последнего десятилетия XX в. международный проект MASTER (Manuscript Access Through Standards for Electronic Records) получил финансовую поддержку Европейской комиссии за разработку стандарта

электронного описания средневековых рукописей, который должен быть универсальным, т. е. предоставлять возможность введения информации из различных по своей культурной и языковой принадлежности памятников – латинских, греческих, славянских, еврейских и других. В процессе работы, однако, появились некоторые проблемы в связи с описанием славянских рукописей. С 1994 г. Институт литературы Болгарской академии наук работает над крупным проектом создания информационного массива, объединяющего археографические, кодикологические, текстологические и литературноисторические данные об оригинальных и переводных средневековых текстах в балканских кириллических рукописях, который называется „Репертоар древней болгарской литературы и книжности через компьютерные средства“¹. Специально разработанная для целей проекта система за компьютерным аналитическим описанием TSM (Template for Slavonic Manuscripts) соответствует специфике средневековых славянских текстов. Она создана по стандарту XML и оборудована поисковой программой – совместной продукт болгарских ученых и профессора Дейвида Бирнбаума. Форматы для описания средневековых рукописей MASTER и TSM основаны на рекомендациях консорциума TEI (Text Encoding Initiative).

Как видно, разработки моделей электронного описания рукописей начинается задолго до начала их дигитализации и подготовки их наборных электронных изданий. Этот факт следует оценить положительно, потому что успешное решение проблем описания через компьютерные средства и последующее ускоренное накопление информации о рукописях и рукописных коллекциях будет способствовать созданию правильной и эффективной стратегии для дигитализации, которая основана на знании об объеме, составе, физическом состоянии и ценности фондов. Так функции моделей описания рукописей

¹ Repertorium of Old Bulgarian Literature and Letters. Location: <http://clover.slavic.pitt.edu/~repertorium/index.html>. Last revised: 10.12.2005 by David J. Birnbaum.

расширяются – кроме изучения рукописного наследия они служат также и для регулирования процесса дигитализации.

Несмотря на достижения в изучаемой области еще много вопросов решены только частично. Поэтому на современном этапе продолжается работа по следующим проблемам: форматы и стандарты, по которым следует структурировать описание; полнота и характер данных; возможности для представления специфики славянской рукописной традиции; возможности для использования соответствующих шрифтов; средства и подходы для поиска и обработки данных; обеспечение совместимости между форматами.

Создание моделей описания печатных изданий связано с решением аналогических проблем. Однако окончательный результат различный, потому что в этом случае объектом анализа являются документы, которые имеют в принципе различные параметры от средневековых рукописей. Модель (модели) описания печатных изданий вырабатывается на основе международных стандартов библиографического описания ISBD. Для введения и хранения данных используются два наиболее распространенных международных формата для машиночитаемых библиографических записей – MARC 21 и UNIMARC. В результате достигается сравнительно высокий уровень стандартизации и унификации данных в библиографических записях. Однако это не значит, что в этой сфере все проблемы решены. Наоборот, в области библиографического описания и каталогизации тоже есть ряд нерешенных вопросов. Самые главные среди них связаны с: выбором между MARC форматами и схемами метаданных, применяемые в сетевых ресурсах; возможностями применяемых софтуерных продуктов для управления, хранения и предоставления доступа до дигитальных ресурсов; возможностями для эффективного обмена записей между базами данных, разработанных на основе различных софтуеров. Модели библиографического описания служат для идентификации и доступа до дигитализированных материалов, но также выполняют и контрольные функции по отношению к

процессу дигитализации в национальном и международном масштабе, которые могут осуществляться путем создания национальных сводных каталогов дигитализированных документов.

В последние годы наблюдается тенденция к сближению моделей описания различных видов документов, которая имеет как сторонников, так и немало противников. Одним из характерных проявлений этой тенденции является разработанная TEI схема описания рукописей TEI P5, в которой ясно проявляется стремление к объединению требований TEI и MARC. Противники такого сближения утверждают мнение, что принятие модели TEI P5 в ее нынешнем виде без каких-либо изменений невозможно, потому что она не учитывает традиция электронного описания средневековых славянских рукописей, которая отличается от традиций каталогизации латинских и греческих рукописей. По их мнению предложенная модель является опытом объединения различных точек зрения для современной каталогизации в области средневекового исследования в одной схеме. Т.е. они утверждают позиция, что соединение элементов различных моделей в одной схеме создает условия для некорректного предоставления данных о первичных материалах. Однако, все специалисты единодушны, что развитие моделей описания документов связано со стандартом XML.

Литература

Баранов, В. А. От оцифрованных коллекций средневековых рукописей к электронным многофункциональным интернет-библиотекам. // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам. Материалы международной научной конференции, Ижевск, 13-17 июля 2006 г. Ижевск, 2006, 3–9.

Баранов, В. Полнотекстовые базы данных как основа для электронных изданий средневековых рукописей в Интернете: требования, реализация, перспективы. // Scripta & e-Scripta, 6, 2008, 47–64.

Кнол, А, Зд. Ухлирж, Т. Псохлавец, Ст. Псохлавец. Изграждане на виртуална европейска библиотека за ръкописи: проектът ENRICH. // Библиотека, XVI, 2009, № 2, 7–12.

Милтенов, Я. Разработване на база данни за копия от средновековни ръкописи в електронния каталог на Централна библиотека на БАН. // Бюлетин на СБИР, 2006, № 1 (за 2006), 51–52.

Мусакова, Е. Ръкописи и описи, или опит да се опише днешната българска кодикология. // Проблеми на изкуството, XXXVIII-A, 2005, № 1, 50–57.

Ярукова, Б. Библиографски контрол и дигитализация в MARC среда: Опитът на Университетската библиотека „Св. Климент Охридски“. // Музеи, архиви и библиотеки в современном информационном обществе. V Международные музейные чтения „Современные проблемы музееведения“. Международная Интернет-конференция. <http://war.ogiik.forum24.ru/?1-5-0-00000007-000-0-0-1303235960>.

Bojadžiev, A. Constructing Repertorium 3.0. Another View of the TEI P5 Model for Medieval Manuscript Descriptions. // Scripta & e-Scripta, 7, 2009, 13–48.

Computational Approaches to the Study of Early and Modern Slavic Languages and Texts: Proceedings of the „Electronic Description and Edition of Slavic Sources“ conference, 24–26 September 2002, Pomorie, Bulgaria. Sofia, 2003. 328 p.

Computer Applications in Slavic Studies: Proceedings of Azbuky.NET International Conference and Workshop 24–27 October 2005, Sofia, Bulgaria. Sofia, 2006. 303 p.

Gorman, M. Bibliographic control or chaos: an agenda for national bibliographic services in the 21st century. // IFLA Journal, XXVII, 2001, № 5–6, 307–313; ERIC Full Text: <http://www.eric.ed.gov/PDFS/ED459796.pdf>.

Janakieva, T., Al. Dipchikova. Bibliographic control and digitization of the cultural heritage. // Globalization, digitization, access and preservation of cultural heritage: Papers from the intern. Conf. Sofia, Bulgaria, Nov. 2006. Sofia, 2007, 275–282. Также в: Библиотека, XIV, 2007, № 1, 5–15.

Medieval Slavic Manuscripts and SGML: Problems and Perspectives. Sofia, 2000. 372 p.

Miltenova, A. Computer Supported Processing of Slavic Manuscripts in Bulgaria: Repertorium Initiative. // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам. Материалы международной научной конференции, Ижевск, 13-17 июля 2006 г. Ижевск, 2006, 118–122.