

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

МОДЕЛИ ОПИСАНИЯ ДИГИТАЛИЗИРОВАННОГО ПИСЬМЕННОГО НАСЛЕДИЯ: СОСТОЯНИЕ ПРОБЛЕМЫ

Н. В. Ганчева

Кирилло-Методиевский научный центр Болгарской академии наук, София, Болгария

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.

Проблема электронного описания дигитализированного письменного на-следия 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 ко-торых 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 элек-тронного описания средневековых рукописей, с одной стороны, и современ-ных 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 средне-вековых 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, да и сам электронный продукт показал известную ограниченность. В конце последне-го 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. пре-доставлять возможность введения информации из различных по своей куль-турной и языковой принадлежности памятников — латинских, греческих, славянских, еврейских и других.

В процессе работы, однако, появились не-которые проблемы в связи с описанием славянских рукописей. С 1994 г. Ин-ститут литературы Болгарской академии наук работает над крупным проек-том 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 лите-ратуры и книжности через компьютерные средства“¹. Специально разрабо-танная для целей проекта система для компьютерного аналитического опи-сания TSM (Template for Slavonic Manuscripts) соответствует специфике средневековых славянских текстов. Она создана по стандарту XML и обору-дована поисковой программой — совместным продуктом 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 на-чинаются 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, соста-ве, физическом состоянии и ценности фондов. Так функции моделей описа-ния рукописей расширяются — кроме изучения рукописного наследия они служат также и для регулирования процесса дигитализации.

Несмотря на достижения в изучаемой области, еще много вопросов ре-шены только частично. Поэтому на современном этапе продолжается работа по следующим проблемам: форматы и стандарты, по которым следует струк-турировать описание; полнота и характер данных; возможности для пред-ставления 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 ана-логичных проблем. Однако окончательный результат иной, потому что в этом случае объектом анализа являются документы, которые имеют параметры, в принципе отличные от параметров средневековых рукописей. Модель (моде-ли) описания печатных изданий вырабатывается на основе международных стандартов библиографического описания ISBD. Для введения и хранения данных используются два наиболее распространенных международных фор-мата для машиночитаемых библиографических записей — MARC 21 и UNIMARC. В результате достигается сравнительно высокий уровень стандар-тизации и унификации данных в библиографических записях. Однако это не значит, что в этой сфере все проблемы решены. Наоборот, в области библио-графического описания и каталогизации тоже есть ряд нерешенных вопросов. Самые главные среди них связаны с: выбором между MARC форматами и схемами метаданных, применяемых в сетевых ресурсах; возможностями при-меняемых программных продуктов для управления, хранения и предоставле-ния доступа до дигитальных ресурсов; возможностями для эффективного об-мена записей между базами данных, разработанных на основе различных про-грамм.

Модели библиографического описания служат для идентификации и доступа до дигитализированных материалов, но также выполняют и кон-трольные 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 не-мало 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 высказывают мнение, что принятие модели TEI P5 в ее нынешнем виде без каких-либо изменений невозможно, потому что она не учитывает традицию 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, что со-единение 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.

Список литературы

Баранов, 2006 — 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, 2006. С. 3–9.

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

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

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

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

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

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

Computational, 2002 — 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.

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

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

Janakieva, 2007 — 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. Pp. 275–282.

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

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