

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

ИНФОРМАЦИОННАЯ СИСТЕМА «ПАМЯТНИКИ РУССКОЙ РУКОПИСНОЙ КНИЖНОСТИ ВОПРОСНО-ОТВЕТНОГО ЖАНРА»

М. Г. Бабалык¹, А. В. Пигин², А. Г. Варфоломеев²

1Карельская государственная педагогическая академия;

2Петрозаводский государственный университет, Петрозаводск

The article deals with the problems of textological study and publication of Russian medieval written documents in the form of questions and answers. The authors suggest the project of a Web information system for comparison and investigation of such documents by means of mathematical models and methods.

Жанр диалога (беседы, в том числе вопросно-ответной) является очень древним в мировой литературе. В древнерусской оригинальной (непереводной) литературе эта форма впервые была использована в Речи Философа в Повести временных лет. В такой форме написаны и многие апокрифы: Беседа трех святителей, Вопросы Иоанна Богослова Господу на горе Фаворской, Вопросы Иоанна Богослова Аврааму на Елеонской горе, Луцидариус, Вопросы о Тивериадском море и другие. В XVII–XX вв. вопросно-ответная форма активно использовалась старообрядческими книжниками в полемических целях. В рукописных сборниках этого времени встречается также большое число вопросно-ответных компиляций, которые еще не обращали на себя внимания исследователей. В ходе исследований апокрифа «Беседа трех святителей» традиционными методами было установлено, что эти компиляции включают заимствования из «Беседы...» и других ранних вопросно-ответных памятников [Бабалык, 2012]. Следовательно, между произведениями этого жанра существует генетическая связь, которая может быть установлена путем тексто-логического анализа. Выявление этой связи, определение круга источников, включенных книжниками в вопросно-ответные компиляции, позволит объяснить цели (литературные, полемические и прочие) создания такого рода текстов и соответственно обогатит наши знания о литературе и культуре позднего русского средневековья и нового времени.

Текстологическое изучение вопросно-ответных произведений имеет свою специфику, которая определяется в первую очередь высокой степенью вариативности этих текстов. Разные списки одного вопросно-ответного памятника могут различаться как количеством вопросно-ответных пар, так и их последовательностью [Бабалык, 2012]. Для сравнения списков друг с другом, для построения вероятных стемм списков [Бородкин, 1983], а также для идентификации фрагментов известных произведений во вновь найденных списках вопросно-ответных компиляций необходимо применять компьютерные технологии. Более того, необходимо организовать работу сетевого научного сообщества и виртуальной среды для исследования списков, так как они находятся в самых разных рукописных хранилищах [Варфоломеев и др., 2010]. Такие задачи требуют использования общепринятого, универсального формата для представления текстов, который был бы удобен для многократного текстологического анализа с помощью различных методик. В качестве формата представления текстов мы предлагаем использовать уже существующий формат на основе стандарта TEI, который был создан нами ранее [Бабалык и др., 2010]. Существенные свойства предлагаемой нами информационной системы состоят в следующем:

1) круг используемых моделей текстов и алгоритмов их сравнения и классификации будет широким, так как тексты памятников, на которые ориентировано Веб-приложение, имеют разнообразную структуру.

2) будут разработаны новые алгоритмы и программные модули, предназначенные для быстрого поиска аналогичных текстовых фрагментов в других произведениях. Так как полное совпадение текстов вопросно-ответных пар встречается довольно редко, а отдельные слова или словосочетания часто входят в разные вопросы и ответы, корректное установление совпадения оказывается трудной задачей. Мы предполагаем использовать для этого принципы семантического поиска, предварительно формализуя смысл вопросов и ответов с помощью графовых структур или других моделей.

3) Информационная система будет не только электронной библиотекой литературных памятников, но и виртуальной средой исследований текстов. Это значит, что размеченные тексты могут исследоваться в нашей среде повторно, при этом ученые могут видеть методики и результаты их предшественников.

Разрабатываемый в рамках нашего проекта инструментарий должен состоять из:

- модуля регистрации, авторизации и персонализации работы с виртуальной средой,
- специализированного XML-редактора, позволяющего размечать тексты в соответствии со стандартной схемой разметки TEI P5 [Wittern et al., 2009],
- модуля создания и редактирования описаний методик и результатов исследований текстов,
- модуля, позволяющего искать описания исследований по различным критериям,
- модуля, состоящего из различных программ сравнения, классификации, семантического поиска текстов, которые могут использоваться в ходе исследований.

Для построения предлагаемого Интернет-ресурса будет использована клиент-серверная архитектура, при этом основные вычисления и обработка данных будут происходить на сервере. Данная концепция позволит довольно эффективно работать в сети и не потребует от пользователя дополнительной установки программного обеспечения, поскольку на стороне клиента будут использоваться возможности стандартного Web-браузера.

Работа выполняется при финансовой поддержке Программы стратегического развития ПетрГУ в рамках реализации комплекса мероприятий по развитию научно-исследовательской деятельности, а также в рамках проекта, поддержанного грантом РГНФ (проект № 12-04-12027в).

Список литературы

Бабалык, 2012 — Бабалык М.Г. Апокриф «Беседа трех святителей» в русской рукописной книжности: исследование и тексты. LAP Lambert Academic Publishing, Saarbrücken, 2012.

Бабалык и др., 2011 — Бабалык М.Г., Варфоломеев А.Г., Пигин А.В. Использование формата TEI для публикации и анализа списков произведений вопросно-ответного жанра // Информационные технологии и письменное наследие: Материалы междунар. науч. конф. (Уфа, 28–31 октября 2010 г.). Уфа;Ижевск, 2010. С.17–20.