

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

ОБЗОР ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ «РУССКАЯ ЛИТЕРАТУРА XVIII ВЕКА»¹

М. В. Пономарева

Санкт-Петербургский государственный университет, Санкт-Петербург

The information retrieval system «Russian literature of XVIIIth century» allows to find texts and their parts according to a set of parameters, such as bibliographic elements, versification, mentioned proper names etc. Such approach allows for multidimensional analysis of texts, from different point of views. The texts for the system are prepared according to TEI P4 guidelines. The set of parameters is open and is determined by the TEI markup. As far as we know, there are no analogous systems for Russian fiction literature.

Доклад посвящен описанию информационно-поисковой системы «Русская литература XVIII века» (antology-xviii.spb.ru).

Использование информационных технологий в литературоведческих исследованиях не соответствует возможностям информационных систем в их применении к гуманитарным наукам. До сих пор преобладают электронные собрания текстов, представляющие собой просто цифровые версии книжных изданий. При этом такие электронные издания являются замкнутыми системами, между отдельными текстами которых, как правило, не обнаруживается взаимосвязей и отношений. Вместе с тем, современные компьютерные технологии предлагают принципиально иные подходы к работе с художественными текстами и в прямом смысле слова открывают новые возможности для фундаментальных научных исследований. Примерами использования такого рода технологий и новых подходов являются, в частности, такие проекты, как: English Poetry Full-Text Database (http://il.proquest.com/products_pg/descriptions/english_poetry_2.shtml), Emblem Project Utrecht (<http://www.let.uu.nl/emblems>), Dante's Lemmatized Works (<http://dante.di.unipi.it/ricerca/dante.html>)

На данный момент подобные проекты осуществляются западноевропейскими и американскими исследователями на материале зарубежной литературной культуры. Однако возможности собственно информационного поиска (во всяком случае, доступные публично) в этих проектах остаются достаточно ограниченными.

В качестве основы для нашей информационно-поисковой системы используется корпус текстов русской литературы XVIII века. Выбор данного материала обусловлен его актуальностью для современной культурной парадигмы, обзорностью текстов, четкой структурированностью жанровой системы, динамикой литературного развития.

¹ Работа выполнена при финансовой поддержке РФФИ, грант № 07-00493а.

Главные задачи проекта:

1. создать модель представления текстов художественной литературы, которая бы позволяла рассматривать материал с различных точек зрения;
2. создать инструмент для учебно-методической и научно-исследовательской работы.

Создавая ИПС «Русская литература XVIII века» в качестве инструмента для описания истории литературы, мы исходим из возможностей современных компьютерных технологий в работе с художественными текстами. В первую очередь, речь идет о многомерном анализе данных по набору параметров. Применительно к литературоведческим задачам в качестве таких параметров могут выступать, например: принадлежность произведения определенному автору, жанру; время написания; время и место публикации; формальные признаки (рифмовка, строфика и проч.); наличие определенных образов; тематика и т. д. Пользователь имеет возможность, согласно заданным параметрам (одному или нескольким), выбирать определенную группу текстов для изучения. Таким образом, весь корпус текстов трансформируется в соответствии с теми или иными установками, в результате чего литературный процесс XVIII века окажется представленным с разных точек зрения.

Важнейшая особенность данной ИПС состоит в том, что набор параметров, описывающих текст, потенциально неограничен. Описав текст один раз, впоследствии можно усложнять это описание, добавляя к его характеристике какие-либо новые параметры, необходимые для продолжения исследования. Подготовка текстов в электронном формате осуществляется в соответствии с рекомендациями TEI (Text Encoding Initiative, <http://www.tei-c.org>).

Подобная информационно-поисковая система позволяет максимально сократить элемент субъективной интерпретации, что в высшей степени актуально для гуманитарных наук в целом.

Применение открытой кодировки позволит пользователям настраивать систему на те или иные специфические задачи, добавляя новые параметры (теги). Особое внимание при разработке ИПС будет уделено вопросам ее интеграции с существующими инструментами анализа данных и организации научного процесса.

В перспективе будет возможно обобщить опыт модели, разрабатываемой на материале словесности XVIII века, на другие периоды истории русской литературы.