

### Список литературы

- Волгирева 2002 — Волгирева, Г. П. Традиционная книжность Прикамья XVI — начала XX вв. как источник для изучения духовной культуры : автореф. дис. ... канд. истор. наук / Г. П. Волгирева. — М., 2002.
- Кириллические 2003 — Кириллические издания XVI–XVII вв. в Хранилищах Пермской области : каталог / под ред. И. В. Поздеевой. — Пермь, 2003.
- Клаус 1967 — Клаус, Г. Сила слова. Гносеологический и прагматический анализ языка / Г. Клаус. — М., 1967.
- Лотман 1970 — Лотман, Ю. М. Структура художественного текста / Ю. М. Лотман. — М., 1970.
- Поздеева 1995 — Поздеева, И. В. Русское старообрядчество и Москва в начале XX века / И. В. Поздеева // Москва старообрядческая. — М., 1995.
- Свинцов 1972 — Свинцов, В. И. О логической культуре редактора / В. И. Свинцов // Книга: Исследования и материалы. — М., 1972. — С. 31–49.

## Модуль выборки и запросов информационно-поисковой системы «Манускрипт»<sup>1</sup>

А. А. Вотинцев

Удмуртский государственный университет, Ижевск, Россия

*This paper presents the descriptions of the ideas, structures, characteristics, directions and development prospects of the module of queries and data sets that are components of the Information Retrieval System “Manuscript”. The module of queries and data sets are intended for a deep and versatile analysis of the information collected and saved in the system database.*

Информационно-поисковая система «Манускрипт» (далее — ИПС «Манускрипт») предназначена для работы с древними рукописями и представляет собой комплекс взаимодействующих программ, центральным звеном которого является база данных. В базе данных системы на сегодняшний день накоплена подробнейшая информация о славянских рукописных памятниках. Система сайтов ИПС «Манускрипт», хотя и является своеобразным «окном», предоставляющим доступ к базе данных системы, не может удовлетворить все потребности специалистов: ориентируясь на широкого пользователя, мы частично ограничиваем объем извлекаемой информации, тем самым сужая возможности по исследованию текстов. Отсутствие гибкого средства глубокого и разностороннего анализа накопленной в базе данных информации продиктовало необходимость создания отдельного модуля, призванного расширить возможности ИПС «Манускрипт» в этом направлении. Модуль выборки и запросов разрабатывается средствами, имеющими более гибкие возможности манипуляции данными, нежели средства для разработки сайтов. В то же время универсальность модуля

<sup>1</sup> Работа по созданию ИПС «Манускрипт» ведется при поддержке Российского фонда фундаментальных исследований (грант № 05-07-90217в).

требует от пользователя специальной подготовки при его использовании.

Модуль оперирует терминами ИПС «Манускрипт», определяющими элементы хранения в БД: единицы, связи, свойства и характеристики единиц, справочники, иерархии и др. [Электронные 2003, 234]. Например, единицы *текст* — *словоформа* — *знак* находятся в связи *родитель* — *потомок*, входящей в лингвистическую иерархию. Единицы типа *словоформа* и *знак* имеют свойство *видимое представление*.

Основной поисковой категорией модуля выборок и запросов является множество единиц определенного типа, результат поиска которых может быть сохранен и представлен в различных видах. Формирование запросов для получения необходимых выборок осуществляется с учетом иерархий единиц, определенных в ИПС «Манускрипт». Действия пользователя при формировании выборок включают в себя следующие основные этапы:

- определение границ поиска единиц;
- выбор типа искоемых единиц;
- задание ограничений для искоемых единиц;
- выполнение запроса и сохранение выборки;
- определение состава печатной формы.

Сформированная и сохраненная выборка может быть уточнена повторным запросом. Кроме того, для сохраненных выборок доступны операции работы с ними как с множествами, что позволяет разделить сложную операцию по извлечению данных на более простые и добиться нетривиальных результатов. Формы представления результата на сегодняшний день — это линейный указатель с выбираемым набором полей и сравнительный указатель найденных единиц. Единожды сформированную выборку можно неоднократно вывести в различных печатных представлениях.

Несмотря на то, что пока модуль имеет в своем арсенале небольшой набор операций по извлечению данных и представлению результатов, их комбинирование (последовательность, состав, уточнение) позволяет получать некоторые предопределенные выборки. Например, модуль позволяет построить указатель словоформ для указанного текста, имеющих в своем составе на заданной позиции символ, удовлетворяющий определенным свойствам. Еще одним примером предопределенной выборки является срав-

нительный указатель единиц нескольких текстов. Для основных предопределенных выборок имеются пошаговые инструкции. Проводится работа по расширению количества предопределенных выборок и описанию шагов, которые должен проделать исследователь, чтобы добиться желаемого результата.

Сложной задачей сегодняшнего этапа является типизация запросов и форм представления результатов, интересующих исследователей. Пожелания исследователей предварительно необходимо формализовать, чтобы их можно было согласовать с требованиями математического аппарата и строгостью запросов к базе данных, заложенных в модуль. Неучтенные потребности в представлении текстовой информации, интересующей исследователей древних рукописных памятников, могут значительно повлиять на развитие модуля, в частности, на состав операций, реализуемых модулем.

Планируется развитие модуля в направлении расширения множества поисковых операций, усложнения условий поиска, расширения состава искоемых объектов, интеграции выборок и редактора ИПС «Манускрипт», разработки методических рекомендаций по выполнению типовых запросов, разграничения полномочий для пользователей системы, оптимизации быстродействия выполнения запросов, совершенствования пользовательского интерфейса.

Модуль создается в технологии трехуровневой архитектуры, программная логика распределена между сервером базы данных и сервером приложений, клиентская часть системы представлена web-браузером. Модуль, являясь Web-приложением, доступен зарегистрированным пользователям из Интернета.

### Список литературы

- Baranov et al. 2004 — Victor Baranov, Andrey Votintsev, Roman Gnutikov, Aleksey Mironov, Sergey Oshchepkov, Vitaliy Romanenko. Old Slavic Manuscript Heritage: Electronic Publications and Full-Text Databases // EVA 2004 London (Electronic Imaging, the Visual Arts Conference & Beyond). Conference Proceedings. — University College London. Institute of Archaeology. Principal Editor: James Hemsley. — London, 2004. — 11.1–11.8.
- Баранов и др. 2004 — Баранов, В. А. Информационно-поисковая система «Манускрипт»: новые возможности электронного издания древнерусских рукописей / В. А. Баранов, А. А. Вотинцев, Р. М. Гнутиков, О. В. Зуга, А. Н. Миронов [и др.] // Информацион-

ный бюллетень Ассоциации «История и компьютер». № 32 : материалы IX конф. АИК (апр. 2004 г.) — Москва ; Томск : Изд-во Том. ун-та, 2004. — С. 7–9.

Баранов и др. 2003 — Баранов, В. А. Электронные издания древних письменных памятников и технология создания полнотекстовых баз данных / В. А. Баранов, А. А. Вотинцев, Р. М. Гнутиков, О. В. Зуга, А. Н. Миронов [и др.] // Круг идей: электронные ресурсы исторической информатики : тр. VIII конф. Ассоциации «История и компьютер» / под ред. Л. И. Бородкина, В. Н. Владимирова. — М. ; Барнаул : Изд-во Алт. ун-та, 2003. — С. 234–270.

## Использование формата TEI для обмена данными с полнотекстовой информационно-поисковой системой «Манускрипт»<sup>1</sup>

П. А. Вотинцев

Удмуртский государственный университет, Ижевск, Россия

*The project contributes to the development of the means for the exchange of documents and their meta- and analytical description (under the XML-TEI format, <http://www.tei-c.org/>) with the full-text databases giving means for a multipurpose processing of the document objects and ensuring creation of electronic publications for various purposes on the Internet (Information Retrieval System "Manuscript", <http://manuscripts.ru/>).*

Работа по созданию электронных изданий древних рукописных памятников в настоящее время ведется различными группами исследователей во многих странах мира. Использование компьютерных баз данных для анализа древних текстов представляется очень перспективным, особенно в связи с развитием Internet-технологий. Однако существует проблема преобразования данных разных форматов.

Проект предусматривает разработку средств обмена данными между форматами хранения документов и их мета- и аналитического описания (на основе формата XML-TEI<sup>2</sup>) и полнотекстовыми базами данных, предоставляющими средства для многофункциональной обработки объектов документа и обеспечивающими создание электронных публикаций различного назначения в Интернете (ИПС «Манускрипт», <http://manuscripts.ru/>).

<sup>1</sup> Работа по созданию ИПС «Манускрипт» ведется при поддержке Российского фонда фундаментальных исследований (грант № 05-07-90217в).

<sup>2</sup> The Text Encoding Initiative [Электронный ресурс]. — Режим доступа: <http://www.tei-c.org/>, <http://www.tei-c.org/release/doc/tei-p5-doc/html/>, свободный. — Загл. с экрана.