

**ИНФОРМАЦИОННЫЙ БЮЛЛЕТЕНЬ  
АССОЦИАЦИИ**

**“ИСТОРИЯ И КОМПЬЮТЕР”**

**№ 31, СЕНТЯБРЬ 2003**

**МОСКВА**

**2003**

**ББК 73  
И74**

**Редакционный Совет:**

ст.преп. **Е.Н. Балыкина** (Минск), д.и.н. **Л.И. Бородкин** (Москва),  
к.и.н. **В.Н. Владимиров** (Барнаул), к.и.н. **И.М. Гарскова** (Москва),  
к.и.н. **С.А. Жакишева** (Алматы), к.и.н. **Е.В. Злобин** (Москва),  
д.и.н. **В.В. Канищев** (Тамбов), д.и.н. **С.Г. Кащенко** (С.-Петербург),  
доц. **А.В. Кириллов** (С.-Петербург), к.и.н. **И.Н. Киселев** (Москва),  
к.и.н. **А.Л. Кобринский** (Москва), д.и.н. **С.И. Корниенко** (Пермь),  
к.и.н. **Г.В. Можаяева** (Томск), к.т.н. **А.Ф. Оськин** (Полоцк),  
д.и.н. **В.В. Подгаецкий** (Днепропетровск), **А.Н. Полевая** (Москва),  
к.и.н. **И.Е. Сафонов** (Москва), д.и.н. **В.Н. Сидорцов** (Минск),  
д.и.н. **Т.И. Славко** (Тверь), д.и.н. **А.К. Соколов** (Москва),  
к.и.н. **И.Ф. Юшин** (Москва).

Ответственный редактор: **Л.И. Бородкин**

Компьютерная верстка: **А.Н. Полевая**, при участии **Т.Я. Валетова**

**Информационный бюллетень Ассоциации "История и компьютер".**

**№ 31. Сентябрь 2003 г. М., 2003. – 280 с.**

**ISBN 5-7904-0241-1**

Материалы для следующих номеров Информационного Бюллетеня можно присылать по адресу:

**119992, Москва, Воробьевы горы, МГУ, 1-й корпус гуманитарных факультетов, Исторический факультет, Лаборатория исторической информатики, к. 654.**

**Тел./факс: (095) 939-11-65**

**E-mail: borodkin@hist.msu.ru**

# АЛГОРИТМЫ ИСТОРИЧЕСКОЙ ИНФОРМАТИКИ

*От редактора*

Одним из перспективных направлений развития исторической информатики является разработка специализированных алгоритмов и программных средств, создающихся с учетом специфики данных исторических источников, представления их в электронном формате, формализации и компьютеризации аналитических процедур исторического исследования. Эффективность таких разработок во многом определяет перспективы исторической информатики как развивающегося направления современной исторической науки.

Начиная с данного номера Информационного бюллетеня мы приступаем к публикации материалов, отражающих новые разработки алгоритмического и программного обеспечения исследований, ведущихся в русле исторической информатики.

Три представленных здесь материала содержат описания новых алгоритмов и программ, созданных членами АИК для работы с историческими текстами, а также с данными о пространственной структуре археологических объектов.

*Д.фил.н. Баранов В.А., Вотинцев А.А.,  
Гнутиков Р.М., Миронов А.Н., Романенко В.А.  
(Удмуртский госуниверситет, Ижевск)*

## **СПЕЦИАЛИЗИРОВАННЫЙ ТЕКСТОВЫЙ РЕДАКТОР «МАНУСКРИПТ» СИСТЕМЫ ОБРАБОТКИ ДРЕВНИХ РУКОПИСЕЙ <sup>1</sup>**

1. При работе с уникальными по своему культурному значению древними текстами большое значение приобретает их комплексное изучение. Любое из предпринятых исследований, будь то текстологическое, палео-

<sup>1</sup> Работа написана в рамках проектов, поддержанных Российским фондом фундаментальных исследований ("Специализированный редактор для обработки и исследования древнейших славянских рукописных памятников", грант № 02-07-90318-в; "Многотекстовая информационно-поисковая система по древнерусским письменным памятникам "MANUSCRIPT", № 02-07-90424-в.) и Российским гуманитарным научным фондом («Информационно-поисковая система по древнерусским текстам»), грант РГНФ №01-04-12027в).

графическое, фонетическое, грамматическое, лексическое, литературоведческое, культурологическое или иное, в значительной мере нуждается в использовании выводов ученых разных специальностей. Одним из условий комплексного анализа, который проводится на протяжении длительного времени, становится обеспечение преемственности результатов работы с текстом в рамках одного проекта. Преемственность в данном случае заключается не только в учете выводов, полученных исследователями ранее, но и в возможности использования практических наработок, например, картотек, упорядоченных перечней, выборок, для анализа вводимого в научный оборот нового материала.

Так, при лингвистическом анализе древнейших и средневековых памятников для получения объективных выводов оказывается важным учитывать состав и структуру рукописей и входящих в них текстов, авторство и жанр последних, переводность и оригинальность разделов и многие другие факторы. Текстологические исследования являются в данном случае необходимым этапом при изучении рукописных памятников. Практически же применение каждым следующим исследователем уже имеющихся сведений начинается с членения рукописи и выборки анализируемых единиц в соответствии с имеющимися текстологическими сведениями, а каждый новый поворот исследования неизбежно ведет к перегруппировке данных в зависимости от новых задач работы. В любом случае практическое применение имеющихся текстологических данных достаточно трудоемко, например, в отношении компилятивных по составу памятников.

Выход из ситуации видится в предоставлении ученым некоего прообраза рукописи, в котором создается и сохраняется многообразная и разнопланово структурированная информация, которую можно оперативно использовать в различных исследованиях.

Основой хранимой информации должны быть формально вычлняемые исследователями компоненты рукописей/текстов и сведения о них. Любой из компонентов является в определенной степени универсальным, так как может существовать в любых рукописных памятниках. Так, любая рукопись может быть структурирована с точки зрения расположения текста – это листы, страницы, колонки, блоки текста; любой письменный текст мы членим на словоформы, многие рукописи должны быть фрагментированы с точки зрения авторства входящих в них текстов и т.п. Вся подобная информация может быть сохранена в виде единиц текста для дальнейшего использования в различных комбинациях.

Для обеспечения комплексного, разностороннего исследования уникальных рукописных памятников и преемственности в работе текст/рукопись должен существовать в максимально приближенной к оригиналу форме в виде набора минимальных структурных составляющих – знаков, которые группируются в более крупные единицы – словоформы, фрагменты, разделы и т.п. (иногда нелинейного состава), которым при-

сваиваются свойственные им значения. По аналогии с выделенными единицами могут быть созданы другие, отсутствующие в самом тексте, но связанные с фактически существующими (например, некие протофрагменты, по отношению к которым текстовые фрагменты могут быть исследованы с точки зрения разночтений, нормализованный грамматический или современный эквивалент текста и т. п.). Все сказанное может быть реализовано с помощью полнотекстовых баз данных.

Понятно, что на всем протяжении работы к электронному образу рукописи должен быть обеспечен доступ, в какое бы время оно ни осуществлялось и где бы электронная копия текста ни хранилась. Кроме того, работа пользователя с полнотекстовой базой данных может быть полноценной и эффективной в том случае, если набор, редактирование, фрагментирование, присвоение значений, выборка, группировка, упорядочивание текстовой информации осуществляются с помощью дружественного интерфейса, который вне зависимости от сложности структуры текста и связей между единицами представляет пользователю достаточно привычный образ рукописи.

II. Традиционный (близкий к печатному) вид текста сохраняют текстовые процессоры (редакторы), которые ориентированы в первую очередь на набор, редактирование и макетирование текста, не имеют штатных средств выделения в нем объектов и в которых отсутствуют возможности работы со свойствами выделенных объектов. Кроме того, сохранение текста осуществляется в локальных файлах, при этом коллективная работа с текстом затруднена. Обособленно стоят текстовые процессоры типа Microsoft Word, которые, имея богатые возможности расширения функций, мощную объектную модель, средства взаимодействия с базами данных, могут быть настроены в качестве интерфейсного инструмента для работы со специализированными базами данных, в частности, полнотекстовыми базами древнейших славянских текстов. Однако анализ методов настройки, способов подключения внешних компонентов, анализ системных требований и способов взаимодействия редактора Microsoft Word с базой данных показал, что разработка собственного инструмента редактирования древнерусских текстов, хранимых в базе данных, значительно менее трудоемка, нежели настройка и расширение возможностей программных продуктов, ориентированных на другие цели.

III. Для работы с данными, хранящимися в полнотекстовых базах данных, целесообразно создание специализированных редакторов. Разработка и создание подобного редактора в настоящее время осуществляется творческим коллективом в Лаборатории по автоматизации филологических исследований Удмуртского государственного университета. Редактор является одним из

модулей системы «Манускрипт»<sup>2</sup> и обеспечивает интерактивный доступ к базам данных.

К редактору могут быть предъявлены следующие ключевые требования:

- набор и редактирование текста и дополнительной информации о нем при непосредственном взаимодействии с базой данных;
- представление текста рукописи в близком к оригиналу виде;
- выделение и создание единиц текста и работа с их свойствами и значениями;
- работа со связями единиц (создание, переподчинение, удаление, изменение свойств и типов связей, работа с иерархическими структурами);
- поддержка многопользовательского режима работы.

Указанные требования нашли отражение в создаваемом продукте. Сегодня в стадии опытной эксплуатации находится прототип редактора, в котором реализованы следующие функции и режимы (см. рис. 1-3):

- многопользовательская работа при непосредственном взаимодействии с базой данных и сохранение в ней результатов работы с документами;
- создание новых документов (набор, редактирование и др. стандартные операции текстовых процессоров) в близком к оригиналу виде;
- выделение единиц текста/рукописи, занесение единиц в базу данных и работа с их свойствами и значениями;
- создание и удаление связей разных типов между единицами;
- работа с различными иерархическими структурами (выделение единицы, создание связи с родительской единицей и т.п.) и отображение структуры связей единиц в виде дерева;
- работа одновременно с несколькими иерархиями;
- работа с единицами в контексте иерархий;
- автоматическая догрузка недостающих элементов иерархии;
- отображение перечня знаков и возможность вставки их в текст с помощью мыши;
- автоматическое объединение символов в словоформы;
- средства автоматической инсталляции.

---

<sup>2</sup> О системе обработки рукописных памятников «Манускрипт» см., например: Баранов В.А., Вотинцев А.А., Гнутиков Р.М., Зуга О.В., Миронов А.Н., Никифорова С.А., Ощепков С.В., Романенко В.А., Рябова Е.В. Электронные издания древних письменных памятников и технология создания полнотекстовых баз данных // Круг идей: электронные ресурсы исторической информатики / Ред. Л.И. Бородкин, В.Н. Владимиров. – М., 2003. – С. 234-270.

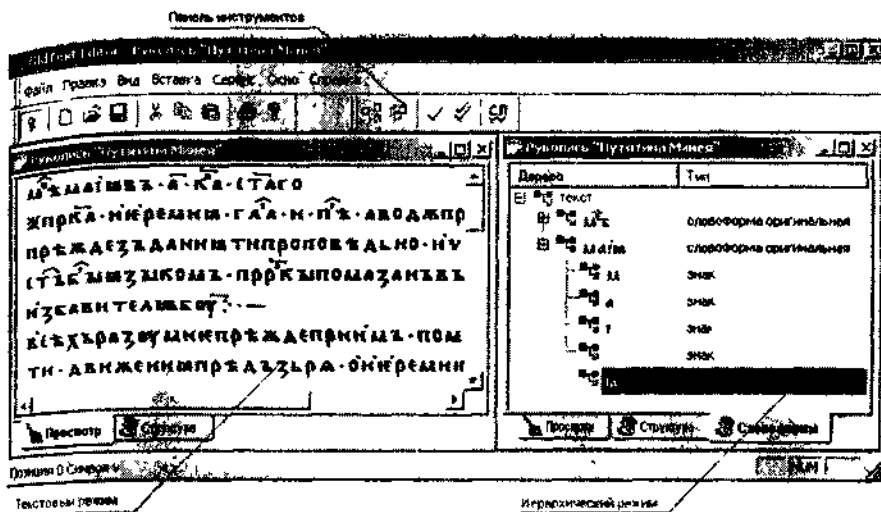


Рис. 1. Общий вид редактора.

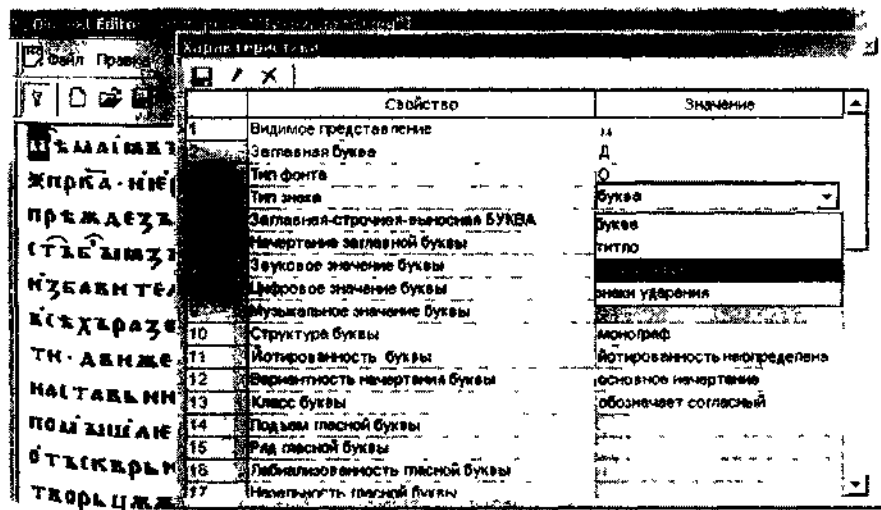
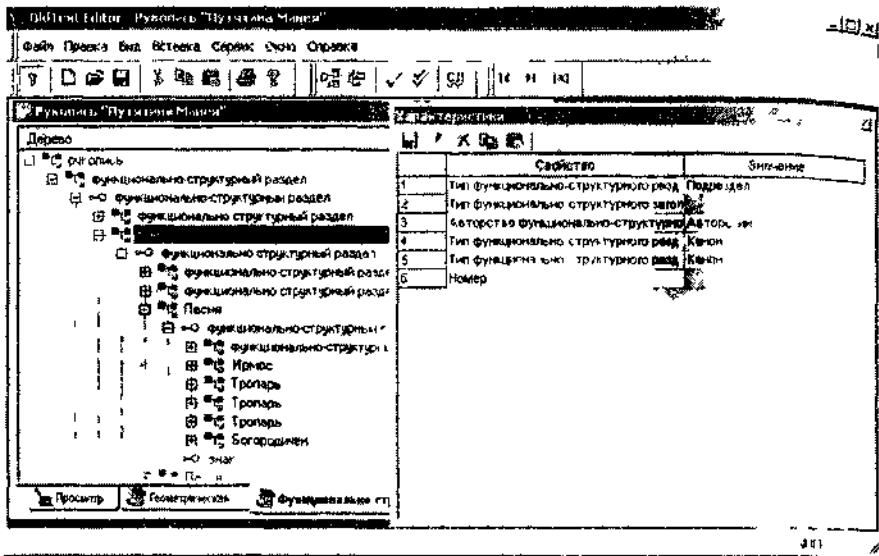


Рис. 2. Режим просмотра и изменения свойств единицы «Знак».



*Рис 3 Просмотр и редактирование свойств единицы «Канон» в рамках функционально-структурной иерархии*

IV. С технической точки зрения редактор состоит из ряда компонентов, реализующих набор следующих интерфейсов:

- IOldedReader – доступ к данным, чтение и запись объектов в базу данных. На данный момент реализована только непосредственная работа с СУБД Oracle через ADO (ActiveX Data Object), планируется организация работы и с другими источниками данных (XML документы, Web-сервисы, SOAP);
- IOldObject – представление единицы древнерусского текста, ее связей с родительскими и подчиненными единицами;
- IOldDocument – представление древнерусского текста как некоторого набора высокоуровневых единиц, таких, как рукопись и текст;
- IOldTOMViewer – интерфейс для отображения древнерусского документа в виде отформатированного текста для клиентов, поддерживающих Text Object Model (ТОМ, объектная модель текста).

V. Редактор реализован на языке C++, компоненты разработаны с помощью ATL (Active Template Library), клиентская часть выполнена с помощью MFC (Microsoft Foundation Classes). При работе клиентская часть



взаимодействует с базой данных системы «Манускрипт» через процедуры серверного API, реализованного в виде пакетов СУБД Oracle.

VI. Нарращивание функциональных возможностей редактора планируется вести в нескольких направлениях.

1. Поддержка одновременной работы с несколькими текстами. Это должно позволить проводить анализ нескольких текстов одновременно (например, разновременных рукописей, содержащих один и тот же текст).

2. Работа с различными выборками. Это должно позволить исследователю получать интересующие его выборки непосредственно в редакторе, должна быть обеспечена возможность сохранения и дальнейшей обработки.

3. Режим локальной работы. Поскольку работа с редактором в режиме взаимодействия с базой данных требует подключения к локальной сети (а это не всегда удобно), необходимо иметь возможность выгрузки определенной части документа и работы с ним в локальном режиме.

4. Доступ из Интернет. На сегодняшний день существующая реализация редактора позволяет пользоваться им из любой точки сети Интернет. Правда, прямое соединение неприемлемо по соображениям безопасности. Поэтому СУБД, с которой взаимодействует редактор, находится за Firewall и недоступна за пределами локальной сети. Такая ситуация должна быть изменена, поскольку ценность системы значительно возрастает при условии работы с ней большого количества исследователей, не только получающих информацию, но и привносящих в нее дополнительные сведения о тексте. В связи с этим должен быть создан дополнительный web-интерфейс, основанный на web-сервисах и на доступе редактора к данным по протоколу SOAP. Но уже сейчас информация о древних памятниках, хранящихся в БД, доступна через специальный web-интерфейс в режиме просмотра (см., например, сайт, посвященный рукописи «Путятина мина», по адресу <http://io.udsu.ru/ptm/>).

5. Интеграция с редакционно-издательской системой TeX. Несмотря на то, что редактор позволяет печатать текстовое представление документа, для подготовки письменных памятников к печатному изданию этого недостаточно. Непосредственное взаимодействие редактора с модулями TeX должно существенно облегчить подготовку текстов и справочных материалов в виде указателей, перечней, словарей к макетированию.

Члены коллектива, создающие редактор и систему «Манускрипт», надеются, что высказанные идеи и их практическая реализация вызовут интерес, и будут благодарны за отклики, рекомендации и замечания. Дополнительную информацию можно найти на сайте, посвященном редактору, по адресу: <http://io.udsu.ru/rd/>.