

Электронное критическое издание средневекового текста: постановка задачи, основные требования и инструментальная подготовка

В. А. Баранов, Р. М. Гнутиков

Ижевский государственный технический университет, Удмуртский государственный университет, Ижевск, Россия

1. Существует несколько типов изданий средневековых памятников письменности – дипломатические, эдиционные, критические [Лихачев 1983]. В основе различий – направленность и назначение публикации. Хотя эти типы в методике подготовки и в технике подачи материала очень существенно различаются, все же между ними есть одно сходство: любое издание – это всегда интерпретация исследуемого документа. В самом общем виде типы различаются количеством используемых дополнительных списков и методикой анализа текста, количеством и видом комментариев и приемами подачи (визуализации) материала. В то же время ни один из этих параметров не исключает другой, скорее – дополняет. Иначе говоря, можно предполагать, что различия между изданиями не столь принципиальны и непреодолимы, как это следует из практики подготовки печатных публикаций.

Ограничения, накладываемые конкретным типом печатного издания на состав и вид материалов, могут быть сняты при подготовке электронных публикаций, в частности для Интернета. Уже сейчас в сети существуют электронные публикации, содержащие (1) сканкопии страниц рукописей, (2) наборные копии текстов, (3) различного рода комментарии, (4) указатели и перечни лингвистических единиц текста, (5) конкордансы. Специально созданные программные средства обработки данных позволяют организовать навигацию по единицам коллекций или отдельного издания, позволяют сформировать запрос на основе существенных для пользователя параметров, упорядочить единицы выборки и организовать удобную форму их представления.

Несмотря на определенные успехи, достигнутые в работе по созданию электронных публикаций древнейших и средневековых письменных памятников, следует признать, что в интернет-изданиях еще достаточно

слабо разработана система подачи, представления материала. Речь идет, в частности, о средствах сравнения данных разных списков одного произведения между собой: о подаче информации об истории текста в виде истории списков, о визуализации различий в структуре и составе рукописей, о технике подачи разночтений, о показе соотношения оригинала и его перевода и некоторые другие.

2. Сформулируем некоторые требования, которые должны быть положены в основу модели данных электронного критического издания.

2.1. Количество и соотношение рукописей (списков):

- отношения рукописей должны описываться связями *антиграф – список*, где каждый антиграф в свою очередь является списком с предыдущего антиграфа;

- связи между антиграфом и списком могут быть вероятностными;

- связи между антиграфами и списками могут быть альтернативными;

- пользователь должен иметь возможность построить и сохранить в базе реконструированный текст, которым может быть как предполагаемый промежуточный список, так и протограф;

- каждый из реконструированных текстов в свою очередь может быть исходным для реконструкции предшествующего списка;

- при сравнении один из списков (в том числе и реконструированный текст) может быть выбран основным, относительно которого будут показываться различия;

- любая из рукописей может быть представлена в электронном издании не полным текстом, а только теми фрагментами (сочетаниями слов, словоформами), которые отличают ее от других.

2.2. Соотношение оригинал – перевод:

- между текстами на разных языках могут существовать отношения *оригинал – перевод*;

- в группе рукописей на каждом из языков может быть свое соотношение *антиграф – список*.

2.3. Выявление различий

При сравнении списков устанавливаются:

- различия в составе и структуре списков на уровне фрагментов,
- различия (разночтения) в лингвистических компонентах соответствующих друг другу фрагментов списков.

2.4. Соотношение рукопись – текст – фрагменты:

- сравнение единиц может быть осуществлено *в одном тексте (произведении)*, представленном разными рукописями и/или фрагментами одной рукописи;

- сравнение единиц может быть осуществлено в пределах *одной рукописи*, если в ней встречаются повторяющиеся единицы (фрагменты);

- сравнение единиц может быть осуществлено *в пределах разных текстов* (например, в том случае, если произведение включает цитаты из других текстов).

2.5. Соотношение рукопись / текст / фрагменты / лингвистические единицы – словари:

- текст и его фрагменты имеют связь с соответствующими текстологическими словарями (справочниками), которые содержат единицы с различными мета- и аналитическими характеристиками;

- лингвистические единицы текста связаны с лингвистическими словарями.

2.6. Типы соотношений текстовых единиц – фрагментов и лингвистических единиц:

- текстовая единица – текстовая единица,

- связь текстовых единиц – связь текстовых единиц,

- значение текстовой единицы – значение текстовой единицы,

- текстовая единица – словарная единица,

- значение текстовой единицы – значение словарной единицы.

2.7. Виды соотносимых лингвистических единиц в рукописях:

- словоформа – словоформа,

- словоформа – Ø единицы (грамматические нули и случаи ситуативно неполных структур),

- словоформа – клон (для случаев контекстуально неполных структур),

- словоформа – сочетание словоформ, которым может быть:

- словоформа с компонентом (предложно-падежная форма, словоформа с артиклем, словоформа с ся, словоформа с не и под.),
- словоформа аналитическая,
- словосочетание,
- синтаксический оборот / конструкция,
- синтаксическая группа,
- предикативная часть,
- фраза,
- несколько фраз,
- синтаксически не оформленное сочетание словоформ и другие комбинации перечисленных единиц.

Примечание:

Нахождение соотносимых единиц может быть осуществлено автоматически при предварительном разделении текста на фрагменты и достаточном упрощении графического вида словоформ.

2.8. Сравнение по расположению соотносимых единиц в рукописях:

- место относительно начала текста (для фрагментов),
- место относительно начала фрагмента (для словоформ),
- место относительно начала группы синтаксически связанных словоформ, например, фразы (для словоформ).

Примечание:

Сравнение по расположению должно осуществляться автоматически.

2.9. Виды соотносимых единиц в словарях:

- лексема – лексема (если тип соотносимых единиц в рукописях: словоформа – словоформа),
- лексема – лексема (если соотносятся словоформа – словоформа с компонентом, словоформа – аналитическая словоформа),
- лексема – словосочетание (главная словоформа в начальной форме),

- лексема – синтаксический оборот / конструкция (главная словоформа в начальной форме),

- лексема – синтаксическая группа (главная словоформа в начальной форме),

- лексема – предикативная часть,

- лексема – фраза и другие комбинации перечисленных единиц.

2.10. Визуализация единиц

Трансформация единиц:

- оригинальный вид,

- преобразованный вид,

- современный эквивалент,

- словарный вид.

Расположение единиц:

- построчное

- пофрагментное (с показом границ строк, столбцов, листов).

2.11. Визуализация различий единиц

Визуализация различий должна осуществляться после указания типа различий (и/или их комбинаций):

- текстологические различия (различия в количестве, составе фрагментов и в их значениях),

- графические различия (различия в составе и количестве букв),

- орфографические различия (варианты основ и окончаний, в том числе по сравнению с нормализованными словарями),

- морфологические различия (словоформы одной лексемы с разным грамматическим значением и/или с разным составом компонентов),

- синтаксические различия (различия, связанные с Ø, клоном, расположением, количеством связанных единиц разного типа),

- лексические различия (словоформы различных лексем).

Примечание:

Сравнение по характеристикам единиц и по их словарным формам должно осуществляться автоматически при наличии необходимых характеристик словоформ и их связей со словарными единицами.

2.12. Визуализация различий отношений:

- различия в расположении соотносимых единиц,
- различия в наличии / отсутствии связей у соотносимых единиц,
- различия в расстоянии соотносимых единиц в группах,
- различия в направлении связей в группах соотносимых единиц,
- различия в синтаксических значениях единиц,
- различия в типе и виде связей в группах соотносимых единиц.

2.13. Визуализация грамматических аналогий

Устанавливается идентичность направления, типа и вида связи в группах соотносимых единиц при различиях в их расположении, форме, значении.

Примечание:

Визуализация отношений осуществляется при наличии грамматических связей между словоформами соотносимых рукописей и при наличии у словоформ синтаксических значений.

2.14. Свойства и значения единиц:

- любая из единиц может иметь свойства и значения, которые должны быть при необходимости показаны пользователю;
- любая из единиц может иметь идентифицирующие единицу значения одного или нескольких свойств (например, заголовков, идентификатор фрагмента и под.), которые должны быть при необходимости показаны пользователю.

2.15. Навигация:

- должны быть обеспечены переходы между соотносимыми единицами (переход по ссылке, по нумерованной ссылке и др.);
- должны быть обеспечены переходы от единиц перечней и указателей к контекстам и наоборот;
- средством навигации должна быть связь между соотносимыми единицами;
- средством навигации может быть связь текстовых единиц с единицами словарей (справочников);

- средством навигации должен быть адрес единицы, включающей аббревиатуру рукописи (текста), идентификатор фрагмента, номер листа, номер страницы, номер столбца, номер строки, на которых начинается единица, или диапазон адресов начала и конца единицы.

2.16. Комментирующая и иная информация:

- любая из единиц, ее связи и значения могут иметь комментирующую информацию, которая должна быть в случае необходимости показана пользователю. Виды комментариев: текстологический, исторический, литературоведческий, исторический, лингвистический, палеографический и др.;

- любая единица может иметь технические комментарии издателя, редактора и др.

3. Рассмотрим реализацию подготовки критического издания в специализированном редакторе OldEd, предназначенном для работы с древними текстами и являющемся одним из модулей информационно-аналитической системы «Манускрипт» (URL: <http://manuscripts.ru/>).

Заметим, что основной задачей редактора является определение отношений единиц и установление различий (разночтений) между исследуемыми документами. Визуализация различий осуществляется другим модулем.

Пользователь может создать неограниченное число реконструированных текстов, определяя назначение каждого из них – промежуточный список или протограф – на свое усмотрение (рис. 1).

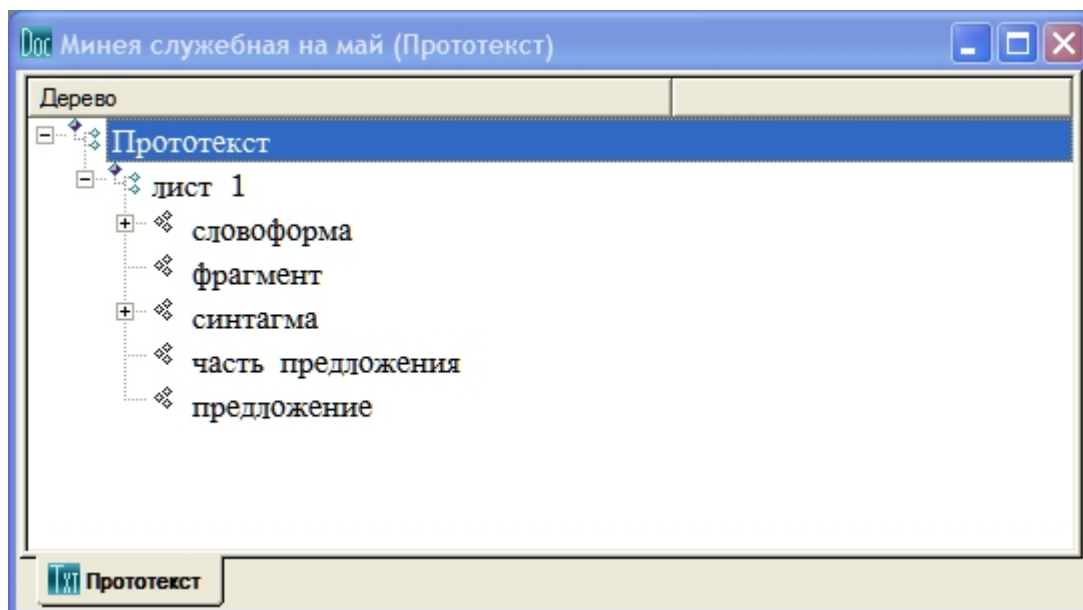


Рис. 1. Иерархия связанных с прототекстом единиц разного типа

В одном и том же реконструированном тексте (прототексте) пользователь может указать различия для разных типов единиц – словоформ, фрагментов, синтагм, предложений и их частей. Количество сравниваемых текстов также не ограничено.

Для удобства навигации прототексты разбиваются на листы.

Прототекст заполняется множеством единиц необходимых типов; указание различий текстов выполняется указанием связей между текстовыми единицами и единицами реконструированного текста (рис. 2).

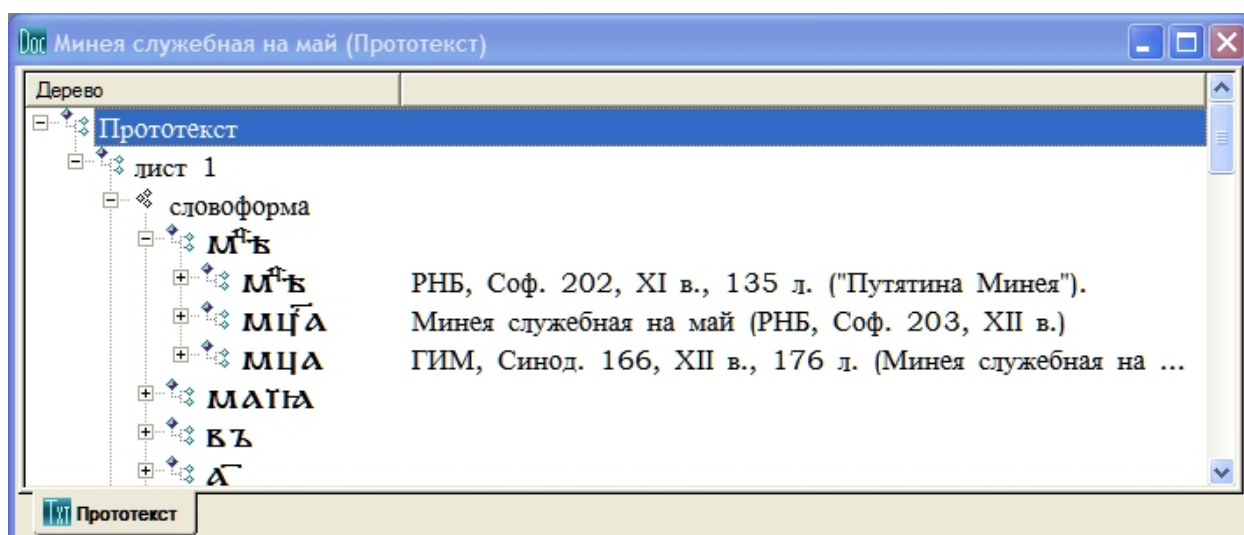


Рис. 2. Единицы прототекста и связанные с ними текстовые прецеденты списков

Создание единиц прототекста возможно в автоматическом режиме с использованием значений текстовой единицы одного из сравниваемых текстов.

Возможен просмотр и изменение свойств как для единиц реконструированного текста, так и для единиц сравниваемых текстов. Предусмотрено указание необходимых комментариев для любых единиц и их связей.

При работе с текстовыми единицами редактор позволяет увидеть их контекст, подчиненные единицы и прочие связи.

Схема подготовки критического издания с использованием редактора выглядит таким образом:

- создание необходимого прототекста и заполнение его необходимым числом листов в автоматическом режиме;
- открытие двух или нескольких сравниваемых текстов;
- создание единицы необходимого типа в реконструированном тексте – или указанием необходимых свойств вручную, или использованием автоматического создания на основании единицы одного из текстов;
- указание отношений (разночтений) между единицами сравниваемых текстов относительно единицы прототекста.

Таким образом, сформулированные требования и функциональная доработка специализированного редактора OldEd являются первоначально достаточной теоретической и технологической базой для создания электронного критического издания на платформе информационно-аналитической системы «Манускрипт».

Благодарности

Работа поддержана Российским гуманитарным научным фондом (проект № 07-04-00369а).

Список литературы

- Баранов и др. 2003 – *Баранов, В. А.* Специализированный текстовый редактор «Манускрипт» Системы обработки древних рукописей / В. А. Баранов, А. А. Вотинцев, Р. М. Гнутиков, А. Н. Миронов, В. А. Романенко // Информационный бюллетень Ассоциации "История и компьютер". – № 31. Сент. 2003. – М., 2003. – С. 159-165.
- Baranov 2004 – *Baranov, Victor.* Old Slavic Manuscript Heritage: Electronic Publications and Full-Text Databases / Victor Baranov, Andrey Votintsev, Roman Gnutikov, Aleksey Mironov, Sergey Oshchepkov, Vitaliy Romanenko // EVA 2004 London (Electronic Imaging, the Visual Arts Conference & Beyond) : Conference Proceedings / University College London. Institute of Archaeology ; principal Editor James Hemsley. – London, 2004. – P. 11.1-11.8.
- Баранов и др. 2006 – *Баранов, В. А.* Редактор OldEd как специализированный инструмент для редактирования документов в базе данных «Манускрипт» / В. А. Баранов, Р. М. Гнутиков // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам: материалы междунар. науч. конф., Ижевск, 13–17 июля 2006 г. / отв. ред. В. А. Баранов. — Ижевск : Изд-во ИжГТУ, 2006. — С 43-46.
- Baranov 2006 – *Baranov, Victor.* Information-Analytical System “Manuscript”: technologies and tools of creation of electronic collections of ancient and medieval documents [Электронный ресурс] / Victor Baranov // Dagstuhl Seminar Proceedings 06491: Digital Historical Corpora - Architecture, Annotation, and Retrieval / L. Burnard, M. Dobрева, N. Fuhr, A. Lüdeling; Dagstuhl Seminar 06491, 03.12. – 08.12.2006; Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany. – Режим доступа: <http://drops.dagstuhl.de/portals/index.php?semnr=06491>, свободный. – Загл. с титул. страницы.
- Baranov 2007 – *Baranov, Victor.* The ideology and technology of creating online full-text digital collections of ancient and medieval slavonic manuscripts / Victor A. Baranov // International Conference on Applied Natural Sciences, Trnava, November 7-9, 2007. – P. 199-207.

Гнутиков и др. 2005 – *Гнутиков, Р. М.* Редактор OldEd : Руководство пользователя (версия 1.1.0.2) [Электронный ресурс] / Удмуртский госуниверситет. Ижевский технический университет; Р. М. Гнутиков, В. А. Баранов, А. А. Вотинцев, А. Н. Миронов. – Электрон. издание. – Ижевск, 2005. – 21 с. – Режим доступа: http://manuscripts.ru/mns/main_sc?p1=4&p_lid=1, свободный. – Загл. с титул. страницы.

Лихачев 1983 – *Лихачев, Д. С.* Текстология / Д. С. Лихачев. – Л. : Наука, 1983. – 639 с.

A digital critical edition of a medieval text: defining the problem, setting basic requirements, and developing tools

Victor A. Baranov, Roman M. Gnutikov

Izhevsk State Technical University, Udmurtia State University, Izhevsk, Russia

This paper formulates the theoretical and technological requirements for a digital critical edition of a text for the Manuscript informational-analytical system. The Manuscript system includes a specialized editor that allows relationships to be constructed between content objects and allows a textual database to be constructed. The paper will discuss the use of the editor to create a prototext based on copies of the Slavonic Service Menaians for May.