

Секция «История»

Исследование княжеских актов XIV-XVI вв. с помощью технологии XML: вопросы методологии.

Замоздра Галина Николаевна

Соискатель

*Челябинский государственный университет, Исторический факультет, Челябинск,
Россия*

E-mail: doroga-life@yandex.ru

Проект по исследованию духовных и договорных грамот московских князей XIV – XVI вв. с помощью информационных технологий возник в связи с личным интересом к источнику, а также обуславливается актуальностью вопроса. Остается неохваченной вся совокупность княжеских актов как с конкретно-исторической точки зрения, так и со стороны дипломатики.

Основными целями проекта являются: анализ структуры, контента грамот, а также самого исследовательского процесса.

Создание машиночитаемой копии не должно нарушать внутренний строй средневекового источника. Помимо самого контента в разметку следует включить также всю совокупность метаданных о документе (место хранения, более ранние публикации, данные палеографии и т. д.). В результате получается полнотекстовая база данных, которую можно охарактеризовать как «слабоструктурированная», «источнико-ориентированная» [Иванов Варфоломеев 2005, с. 254-255.]. Техническая составляющая исследования – конструирование “скелета” базы данных – вопрос, который обсуждается.

Для реализации проекта выбор был остановлен на технологии XML (Extensible Markup Language – расширяемый язык разметки). В качестве свода общих синтаксических правил XML позволяет хранить структурированные данные в текстовом формате. Важны иерархическая структура размечаемого документа (“дерево”, состоящее из системы вложенных друг в друга тегов), возможность создания пользовательских элементов и преобразование XML-документа в доступный пользователю вид (отображение в web-браузере). XML-документ прост в создании. Для редактирования и преобразований существует свободное ПО.

В работе с текстами грамот используется издание под редакцией Л.В. Черепнина 1950 года — наиболее полное, актуальное и доступное на сегодняшний день — в ходе подготовки публикации Л.В. Черепнин впервые собрал воедино все выявленные завещания и договоры князей. [ДДГ 1950, с. 4-5.]

Устойчивая внутренняя организация княжеских актов, размеченных в XML, позволяет абстрактно представить их в виде формализованной структуры с четкой иерархией тегов (дерево, граф, множество). Сравнить индивидуальные модели (формуляры) и выявить группы конкретных формуляров отдельно для духовных и для договорных грамот — означает проследить эволюцию внутренней структуры документов; можно сравнивать один документ со всеми другими и каждый документ с каждым другим документом.

Специфика княжеских грамот — наличие формуляра — в тексте по смыслу можно выделить статьи, состоящие из предложений, которые делятся на обороты; обороты могут

включать в себя элементы, а у элементов могут быть характеристики; среди элементов и характеристик различают реалии, формулы и описания.[Каштанов 1970, с. 26-47.] В данном случае реалиями будут в основном топографические термины (населенные пункты), а также имена собственные (в основном княжеские, боярские, имена митрополитов) – то есть, отдельные понятия. Формулы – это как правило словосочетания – отдельные обороты. Исходя из этого, в разметку текста грамот изначально закладывается маркирование смысловых единиц текста – отдельных понятий, имен собственных, фактов, оборотов в виде формул; тем, выраженных в отдельных частях текста.

Запросы к имеющейся базе должны решать задачи разных уровней организации информации. Система ТАСТ соответствует большинству задач (от простого запроса на поиск отдельного слова до наиболее сложных запросов, связанных с подсчетом совместной встречаемости нескольких слов или категорий)[Историческая информатика 1996, с. 133-141], однако не поддерживает дореволюционную кириллицу. Поэтому планируется создание специализированного программного обеспечения, включающего в себя лемматизатор и, возможно, web-интерфейс. Количественные и качественные данные, полученные в результате запросов, в дальнейшем могут быть загружены в одну из реляционных СУБД для последующей их обработки.

Литература

1. Духовные и договорные грамоты великих и удельных князей XIV-XVI вв. М.; Л., 1950.
2. Иванов А.С., Варфоломеев А.Г. Технология XML как инструмент компьютерного источниковедения (на примере формулярного анализа документов приказного делопроизводства) // Круг идей: Алгоритмы и технологии исторической информатики. М.; Барнаул. 2005.
3. Историческая информатика / Под ред. Л.И. Бородкина, И.М. Гарсковой: М., 1996.
4. Каштанов С.М. Очерки русской дипломатики М., 1970.