

Варфоломеев А.Г. (Петрозаводск), Иванов А.С. (Даугавпилс)

**ТЕХНОЛОГИЯ XML КАК ИНСТРУМЕНТ КОМПЬЮТЕРНОГО
ИСТОЧНИКОВЕДЕНИЯ**
**(на примере изучения комплекса средневековых документов
по истории Динабурга)**

1. Цель доклада – раскрытие возможностей технологии XML как средства эффективного и достоверного анализа информации средневековых источников.

2. Для апробации подхода выбирается конкретный комплекс источников – документы Разрядного приказа (РГАДА, фонд 210), отражающие историю города Динабурга (ныне Даугавпилс) под властью царя Алексея Михайловича в 1656–1667 гг. Основной массив документов опубликован ¹.

3. Выбор данного комплекса обусловлен следующими причинами:

Сплошной обработке и изучению подвергается исторически сложившийся комплекс взаимосвязанных документов, что позволяет проводить их всестороннее изучение, не ограничиваясь информацией, запечатленной в отдельных, вырванных из контекста документах. Благодаря этому верификация извлекаемой информации получает высокую степень надежности.

В комплексе источников представлены практически все разновидности приказной документации: указы и указы воеводам, памяти, отписки воевод, росписи служивых людей, годовые сметы, докладные записки в Разряде, челобитные, расспросные речи. Это позволяет конкретную методку, апробированную на данном комплексе, без особых трудностей перенести на изучение других источников по истории России XVI–XVII вв.

4. Для изучения данного комплекса источников перспективным представляется последовательное применение методов актового источниковедения к материалам приказного делопроизводства, которые, хотя и являются документами неактового характера, относятся к объектам дипломатики ². При этом в центре внимания должно быть изучение именно внутренней формы документов (формулярный анализ), методы которого разработаны наиболее тщательно. Формулярный анализ делопроизводственных документов позволяет отделить индивидуальные описания, характеристики конкретного источника от формул, устойчивых штампов, этикетно-стилистических оборотов, обязательных для данной разновидности документов. В отступлениях индивидуального формуляра от абстрактного содержится ценная историческая информация. К сожалению, абстрактные формуляры для конкретных разновидностей делопроизводственных документов обозначены в исследованиях лишь в самых общих чертах. Поэтому практическое выявление индивидуальных особенностей внутренней формы документов возможно лишь путем сопоставления их индивидуальных формуляров. На наш взгляд, эффективное и достоверное решение этой задачи возможно лишь с помощью компьютерных технологий.

5. Идея отделения формуляра (логической структуры) документа от его содержания и индивидуальных особенностей лежит в основе технологии XML. Под XML (eXtensible Markup Language) понимают свод правил для создания новых языков разметки текста, подобных HTML. На сегодняшний день в мире существуют десятки языков, основанных на XML и претендующих на роль стандартов в описании тех или иных видов текстовой информации. Разметка выделяет некоторые смысловые единицы текста, которые могут вкладываться друг в друга и описывать структуру источника с любым

¹ Акты Московского государства. СПб., 1894, 1904. Т. 2,3; Иванов А., Кузнецов А. Динабург в документах Российского государственного архива древних актов (1656–1666). Даугавпилс, 2002. Ч. I-II.

² Каптанов С.М. Русская дипломатика. М., 1998.

уровнем детальности, вплоть до отдельных слов или символов. Смысловые единицы могут быть снабжены атрибутами, относящими их к тем или иным классам объектов или задающими для них стандартные значения (это касается, например, дат или географических названий). Разметка может быть использована также для связи фрагментов текста с их переводами и комментариями. Технология XML дает в руки исследователей сколь угодно гибкое средство структурирования и анализа текста, оставляющее источник неизменным, не разбитым на ячейки реляционных таблиц. Можно сказать, что это технология источник-ориентированных баз данных, доведенная до совершенства.

6. В нашем проекте мы собираемся использовать две схемы разметки – TEI и HEML. TEI (Text Encoding Initiative) – один из самых известных стандартов разметки, широко используемый в проектах по созданию электронных коллекций для гуманитарных наук³. Особое внимание в этом стандарте уделяется средствам анализа (интерпретации) текста, что делает его незаменимым для источниковедческих задач. Другой стандарт – HEML (Historical Events Markup and Linking) – специально создан для описания исторических событий, упоминаемых в текстах. Одной из главных целей создатели HEML видят визуализацию временной и пространственной информации о событиях с помощью динамически создаваемых карт и временных шкал⁴. В докладе на конкретных примерах будут рассмотрены особенности использования этих стандартов для изучения средневековой приказной документации.

³ Официальный сайт TEI – <http://www.tei-c.org>.

⁴ Сайт проекта HEML – <http://www.heml.org>.