

Принципы электронных публикаций комплексов исторических документов со средствами палеографического, текстологического и дипломатического анализа\*

А. Г. Варфоломеев, А. С. Иванов

Петрозаводский государственный университет, Россия

Даугавпилсский университет, Латвия

Изучение исторически сложившихся комплексов документальных источников имеет давние традиции в источниковедении [Шмидт 1997, Янин 1977], однако возможности такого изучения, как правило, ограничены несовершенством и громоздкостью традиционных «бумажных» публикаций. Очевидно, что обширный комплекс документов с достаточным научным и справочным аппаратом можно опубликовать только в электронной форме. Однако два серьезных фактора препятствуют этому: косность мышления, связанная с абсолютизацией общепризнанных традиционных подходов в археографии, и разрыв между экспоненциально растущими технологическими возможностями и сравнительно медленно меняющимися знаниями потенциальных «постановщиков задач» в области электронных публикаций. Часто электронные публикации просто копируют «бумажные», достигая при этом некоторых естественных преимуществ – улучшения качества передачи изображений документов (цифровая фотография лучше прежних литографий и фотокопий) и снятия ограничений на объем публикаций. Но ориентироваться в бумажном тексте проще, чем в его электронной копии, соответственно, указанные выше преимущества электронных публикаций не могут компенсировать некоторое неудобство их использования. Поэтому очень актуальной является задача разработки новых форм и инструментов электронных публикаций комплексов исторических документов. На наш взгляд, такие публикации должны быть ориентированы на решение не только археографических, но и аналитических задач, и предоставлять разнообразные возможности источниковедческой критики документов. Для достижения этой цели мы предлагаем специалистам-историкам,

---

\* Статья написана в рамках проекта, поддержанного грантом Российского гуманитарного научного фонда (проект № 08-01-12136в).

выступающим в роли постановщиков задач электронных публикаций, три принципа, или, если угодно, парадигмы: «публикация как семантическая сеть», «публикация как ящик с инструментами (toolbox)» и «публикация как сетевой сервис».

Все документы комплекса и, соответственно, все тексты в публикации связаны между собой множеством отношений. Некоторые отношения присутствуют в документах изначально, например, прямые ссылки на другие документы; другие отношения устанавливаются исследователями в процессе публикации, например, связи текстов документов с научными комментариями, указателями или изображениями. Такие связи легко представить себе в виде стрелок разного типа, соединяющих некие объекты, то есть в виде «семантической сети». Конечно, любая научная публикация источника включает его в некую семантическую сеть. Мы просто предлагаем делать ее более детальной. Так, в качестве объектов сети должны фигурировать не только тексты и изображения, но и их произвольные фрагменты. Фактически речь идет о замене текста как последовательности кодов символов на сложную систему объектов: с одной стороны, символов, словоформ, строк, листов, с другой – разнообразных логических фрагментов и ссылок на другие объекты. Чтобы не потерять при этом сам текст как целое, разумно использовать технологию XML [Иванов и др. 2005], создавая объекты как элементы разметки исходного текста. Что касается изображений, то их фрагменты также удобно задавать с помощью XML-разметки, записывая координаты в текстовый файл формата SVG.

Палеографический анализ источника, как составная часть научного аппарата традиционной публикации, предусматривает детальное описание физических особенностей документа и символов текста. Присутствие в электронной публикации фрагментов изображений документов вплоть до отдельных символов, а также связи между символами в тексте и их изображениями позволяет снабдить публикацию богатыми средствами палеографического анализа. Действительно, нет никаких препятствий для того, чтобы показывать по запросу пользователя все варианты написания того или иного символа в виде отдельных изображений в контексте всего

документа. Изображения можно выводить на экран вместе с линейками, демонстрирующими размеры символов. Более изощренные средства могут сравнивать изображения символов между собой или переводить их в векторный формат, например, для создания шрифта, максимально приближенного к особенностям почерка данного документа.

Дипломатический анализ комплекса документов предполагает выделение в текстах устойчиво повторяющихся или, напротив, уникальных логических фрагментов с тем, чтобы затем построить модели структуры текста и группы текстов – индивидуальный, конкретный и абстрактный формуляры [Каштанов 1988: 170]. Электронная публикация может демонстрировать уже проведенное деление текстов на фрагменты, предоставляя средства для сравнения структур ряда текстов (с помощью визуализации или путем подсчета по некоторому правилу «расстояния» между структурами). Но также вполне разумным выглядит вариант предоставления пользователю средства самостоятельного построения таких моделей текстов, чтобы, например, проверить выводы, сделанные другими исследователями.

Грамотно проведенный дипломатический анализ позволяет сравнивать между собой не только структуры документов, но и их содержание (как текстов в целом, так и их отдельных фрагментов) с тем, чтобы установить более глубокие взаимосвязи между документами, решить задачи атрибуции, датировки, построения стеммы списков и т.п. Как и в случае палеографического или дипломатического анализа, электронная публикация может включать в себя демонстрацию результатов уже проведенного текстологического анализа, сопровождающуюся ссылками на соответствующие фрагменты текстов или изображений, а также сравнительными таблицами и иллюстрациями (например, стеммами). Следует отметить, что наличие в составе публикации инструментов проведения такого анализа – средств визуального сравнения текстов и их фрагментов, фиксации сходств и различий, формирования гипотез о происхождении и связи текстов, построения сравнительных таблиц и стемм – значительно расширяет возможности использования электронной публикации.

Принятие принципа формирования программной «оболочки» электронной публикации в виде среды для проведения исследований с помощью встроенных инструментов приводит к необходимости уточнить введенное выше понятие публикации как «семантической сети» – это должна быть *открытая* сеть, то есть сеть, в которую можно добавлять объекты и связи между объектами.

Инструменты и результаты исследований, выступающие составной частью публикации, могут быть отделены от нее. Действительно, одни и те же инструменты можно использовать в разных публикациях, если будет сохраняться единство формата представления информации. Результаты исследований могут служить отправной точкой для других исследований того или иного комплекса документов. Для достижения таких возможностей публикации не должны быть изолированы друг от друга, их необходимо чем-то объединить. Поэтому мы предлагаем создавать электронные публикации как сервисы сети Интернет, предоставляющие свои тексты, изображения, средства анализа и результаты для совместного пользования. Можно сказать, что информация, содержащаяся и накапливающаяся в таких сетевых сервисах-публикациях, выступает в роли базы знаний сетевого сообщества исследователей комплексов исторических документов [Варфоломеев и др. 2007].

Принципы, предлагаемые в нашей статье, реализуются нами в проекте по подготовке электронной публикации комплекса документов «Moscovitica-Ruthenica» [Иванов и др. 2006].

#### Список литературы

- Варфоломеев и др. 2007 – *Варфоломеев, А. Г.* Приобретение и представление знаний в сетевом сообществе исследователей текстов / А. Г. Варфоломеев, И. В. Кравцов// Вторая конференция "Системный анализ и информационные технологии" САИТ-2007, Обнинск, Россия, 10-14 сентября 2007 г. : Тр. конф. : в 2 т. Т. 1. – М., 2007. – С.104-106.
- Иванов и др. 2005 – *Иванов, А. С.* Технология XML как инструмент компьютерного источниковедения (на примере формулярного анализа документов приказного делопроизводства) / А. С. Иванов,

А. Г. Варфоломеев // Круг идей: Алгоритмы и технологии исторической информатики : Тр. IX конференции Ассоциации «История и компьютер» / Ред. Л. И. Бородкин, В. Н. Владимиров. – М.; Барнаул : Изд-во Алтайского университета, 2005. – С. 241-281.

Иванов и др. 2006 – *Иванов, А. С.* Проект издания комплекса документов «Moscovitica-Ruthenica» / А. С. Иванов, А. Г. Варфоломеев // Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам : материалы междунар. научн. конф., Ижевск, 13-17 июля 2006 г. / Отв. ред. В. А. Баранов. – Ижевск: Изд-во ИжГТУ, 2006. – С. 65-70.

Каштанов 1988 – *Каштанов, С. М.* Русская дипломатика / С. М. Каштанов. – М., 1988.

Шмидт 1997 – *Шмидт, С. О.* Современные проблемы источниковедения / С. О. Шмидт // Шмидт С. О. Путь историка : Избранные труды по источниковедению и историографии. – М., 1997.

Янин 1977 – *Янин, В. Л.* Очерки комплексного источниковедения: Средневековый Новгород / В. Л. Янин. – М., 1977.

Principles of electronic publication of aggregate historical documents using  
paleographical, textological, and diplomatic analysis

Alexey G. Varfolomeev, Alexander S. Ivanov

Petrozavodsk State University, Russia

Daugavpils University, Latvia

This paper deals with the principles for representing historical records in digital publications (collections) that can provide historians with proper tools for external and internal criticism of historical sources. The authors substantiate three principles (“paradigms”), which should be taken into consideration by archaeographers and editors of digital collections of historical records: 1) digital publication is an open semantic network, which links arbitrary fragments of texts, images, and research abstracts and commentaries to each other; 2) digital publication is a specific toolbox that encompasses different tools of palaeographic, diplomatic, and textual analysis of the published historical records; 3) digital publication is a network service (Web

information system) that incorporates historical sources, research tools, and research results into a knowledge-based system, which is specially created for a network community.