

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

А. Г. Варфоломеев, А. С. Иванов

Компьютерное источниковедение

**Семантическое связывание информации
в репрезентации и критике
исторических источников**

Петрозаводск
Издательство ПетрГУ
2013

УДК 004.9
ББК 32.973.1
В187

Рецензенты:

доктор исторических наук *С. И. Корниенко*,
доктор исторических наук *А. М. Пашков*

Варфоломеев, А. Г., Иванов, А. С.

В187 Компьютерное источниковедение : семантическое связывание информации в репрезентации и критике исторических источников / А. Г. Варфоломеев, А. С. Иванов. – Петрозаводск : Изд-во ПетрГУ, 2013. – 204 с.

ISBN 978-5-8021-1573-2

Монография посвящена теоретическим проблемам, методам и технологиям нового научного направления – компьютерного источниковедения. Особое внимание уделено использованию технологий разметки текста и семантических публикаций исторических источников.

Для научных работников, специалистов по информатизации архивов, преподавателей, аспирантов, студентов вузов, всех, кто интересуется применением информационных технологий в гуманитарных науках.

УДК 004.9
ББК 32.973.1

Издается в рамках реализации комплекса мероприятий
Программы стратегического развития ПетрГУ на 2012–2016 годы

ISBN 978-5-8021-1573-2

© Варфоломеев А. Г.,
Иванов А. С., 2013
© Петрозаводский
государственный университет, 2013

Содержание

| | |
|---|-----|
| Введение. Компьютерное источниковедение как направление общего исторического источниковедения | 5 |
| Глава 1. Повышение информативной отдачи источников как центральная задача исторического источниковедения | 13 |
| Глава 2. Комплексный подход в компьютерном источниковедении и конкретные задачи изучения комплексов исторических источников | 22 |
| Комплексный подход к историческим источникам: теория и практика..... | 22 |
| Комплекс «Moscovitica–Ruthenica» и субкомплексы смоленско-рижских актов XIII – первой половины XIV века и полоцких грамот XV века | 30 |
| Комплекс источников по истории Динабурга середины XVII века ... | 40 |
| Задачи изучения комплексов исторических источников в компьютерном источниковедении | 46 |
| Глава 3. Разметка текстов исторических источников – средство представления, обработки и семантического связывания источниковой информации | 55 |
| XML: особенности, преимущества, схемы разметки TEI и CEI | 57 |
| Представление метаинформации и информации о палеографических особенностях источников | 65 |
| Информационные модели структуры и содержания отдельных источников и их комплексов | 77 |
| XML как технология агрегирования, связывания, анализа и репрезентации источниковой информации | 103 |
| Глава 4. Технологии семантического Web и возможности непосредственного связывания информации исторических источников | 114 |
| Виртуальные среды для научных исследований и «communities of practice» | 114 |

| | |
|--|-----|
| Семантические публикации исторических источников: концепция и ее реализация (SMW и ACE) | 126 |
| Семантические сети для реконструкции исторической реальности (HSN)..... | 154 |
| Заключение | 162 |
| Библиография | 164 |
| Список иллюстраций..... | 193 |
| Summary. Computer-based Records Studies: Semantic Linkage of Information in Representation and Criticism of Historical Records | 196 |
| Об авторах..... | 202 |

Введение

Компьютерное источниковедение как направление общего исторического источниковедения

В структуре исторического знания центральное место принадлежит историческому источниковедению и связанным с ним специальным и вспомогательным историческим дисциплинам. В эпоху внедрения информационных технологий в изучение истории закономерно возникает вопрос о том, приобретают ли в результате компьютеризации исторической науки какое-либо новое качество эти традиционные области исторического исследования, прежде всего – классическое историческое источниковедение.

В российской историографии термин «компьютерное источниковедение» появился уже в середине 90-х годов XX века¹. При этом вполне справедливо подчеркивалась его технологическая и аналитическая составляющая: «...“компьютерным источниковедением” мы ... называем как содержательные, так и внешние атрибуты источниковедческой работы, включающей создание и обработку машиночитаемых источников; совокупность связанных с этим процедур, которые выполняются на компьютере. Иными словами, “компьютерное источниковедение” – это источниковедческая работа (и ее результаты), ориентированная на определенный набор компьютерных технологий»². В отдельных случаях наблюдается отождествление компьютерного источниковедения и исторической информатики³, а сфера применения компьютерного источниковедения достаточно четко ограничивается критикой массовых источников. Здесь следует отметить, что компьютерное источниковедение можно с полным основанием считать составной частью исторической информатики⁴, однако отождествление компьютерного источникове-

¹ Тяжелникова В. С. Компьютерное источниковедение: К постановке проблемы. В кн.: *Круг идей: Развитие исторической информатики*. Москва, 1995. С. 270–284.

² Владимиров В. Н., Цыб С. В. Источниковедение в век компьютера (вместо предисловия). В кн.: *Источник. Метод. Компьютер: Сборник научных трудов*. Барнаул, 1996. С. 7.

³ Соколов А. К., Тяжелникова В. С. Массовые источники и компьютеризация исторических исследований. В кн.: *Источниковедение новейшей истории России: Теория, методология, практика*. Москва, 2004. С. 453.

⁴ Об исторической информатике см. подробнее: Гарскова И. М. Основные направления развития исторической информатики в конце XX – начале XXI века. *Вестник Московского государственного университета*. 2010. № 6. С. 75–103; Гарскова И. М. Источниковедческие проблемы исторической информатики. *Российская история*. 2010. № 3. С. 151–161.

дения и исторической информатики – более широкой отрасли знания – едва ли оправданно.

В наши дни компьютерное источниковедение все еще находится в процессе становления, а в среде специалистов существуют разногласия в определении его предмета. Вряд ли можно согласиться с попытками ограничить область применения компьютерного источниковедения и свести его к источниковедческой эвристике, т.е. к технологиям создания баз данных, репрезентации исторических источников или их информации в сети Интернет, подготовки электронных публикаций. Разумеется, представление исторических источников в источнико-ориентированных базах данных или электронных публикациях обеспечивает их доступность для исследователей, существенно облегчает и ускоряет работу с ними, а также способствует популяризации, а иногда и вульгаризации, знаний о прошлом. Однако, как уже неоднократно отмечалось исследователями, в наибольшей степени потенциал компьютерных технологий раскрывается именно при их использовании для проведения аналитической и синтетической критики исторических источников, о чем свидетельствуют публикации (некоторые из которых, к сожалению, имеют декларативный характер) по применению компьютерных методов в конкретных источниковедческих исследованиях⁵. Следует подчеркнуть, что компьютерные методы могут с успехом применяться к изучению не только массовых, но и уникальных исторических источников.

Итак, сфера применения методов компьютерного источниковедения определяется достаточно четко. Это, во-первых, источниковедческая эвристика – в настоящее время наиболее развитая область компьютерного источниковедения; во-вторых, методико-аналитиче-

⁵ См., например: Ivanovs, A., Varfolomeyev, A. Editing and Exploratory Analysis of Medieval Documents by Means of XML Technologies. In: *Humanities, Computers and Cultural Heritage*. Amsterdam, 2005. Pp. 155–160; Иванов А. С. Работа с XML-документом как воспроизведение основных этапов источниковедческой критики: новые технологии и возможность коррекции традиционных подходов. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. Специальный выпуск. № 34. Москва; Тамбов, 2006. С. 66–67; Варфоломеев А. Г., Иванов А. С. Технология XML: Современная реализация источнико-ориентированного подхода в работе с комплексами исторических документов. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. Специальный выпуск. № 34. Москва; Тамбов, 2006. С. 61–62; Varfolomeyev, A., Soms, H., Ivanovs, A. Knowledge-Based Information Systems in Research of Regional History. In: *Digital Humanities 2008*. Oulu, 2008. Pp. 210–211; Варфоломеев А. Г., Иванов А. С. Принципы электронных публикаций комплексов исторических документов со средствами палеографического, текстологического и дипломатического анализа. В кн.: *Современные информационные технологии и письменное наследие: от древних текстов к электронным библиотекам. El'Manuscript-08: Материалы Международной научной конференции*. Казань, 2008. С. 60–63 и др.

ское источниковедение⁶, в центре внимания которого – методика и техника изучения (критики) исторических источников.

Не вызывает сомнений, что специфика компьютерного источниковедения обусловлена методами и техникой изучения исторических источников на основе компьютерных технологий. Прикладной характер, технологическая (инструментальная) направленность компьютерного источниковедения в определенной степени лишают его теоретической и методологической самостоятельности: можно утверждать, что в настоящее время теоретическая и методологическая основа компьютерного источниковедения – традиционное теоретическое источниковедение, разрабатывающее основные принципы источниковедческого познания – теорию исторического источника, а также теорию познания источника⁷. Однако не исключено, что в перспективе компьютерное источниковедение сможет внести свой весомый вклад в дальнейшую разработку учения о методах исторического источниковедения и теорию исторического источника в контексте более широкого учения об информации, тем более что первые шаги в этом направлении уже сделаны⁸.

Неразрывная связь в теоретико-методологическом плане компьютерного источниковедения с традиционным историческим источниковедением ставит под вопрос обоснованность выделения компьютерного источниковедения из общего исторического источниковедения. К тому же для компьютерного источниковедения вряд ли может быть определен свой предмет и объект исследования, отличный от предмета и объекта традиционного источниковедения⁹, что вполне естественно, так как историческое источниковедение – комплексная вспомогательная (специальная) историческая дисциплина, совокупность методов которой постоянно расширяется за счет междисциплинарных связей различных областей знания. Таким образом, набор

⁶ Медушевская О. М. *Современное зарубежное источниковедение*. Москва, 1983. С. 6.

⁷ Шмидт С. О. Современные проблемы источниковедения. В кн.: Шмидт С. О. *Путь историка: Избранные труды по источниковедению и историографии*. Москва, 1997. С. 27–28.

⁸ Ковальченко И. Д. Исторический источник в свете учения об информации. (К постановке проблемы). *История СССР*. 1982. № 3. С. 129–148; Ковальченко И. Д. *Методы исторического исследования*. Москва, 1987. С. 106–126; Иванов А. С., Варфоломеев А. Г. Идеи И. Д. Ковальченко об информационной ценности и неисчерпаемости источника в современном компьютерном источниковедении. В кн.: *Идеи академика И. Д. Ковальченко в XXI веке: Материалы IV научных чтений памяти академика И. Д. Ковальченко*. Москва, 2009. С. 154–164.

⁹ См. подробнее: Шмидт С. О. *Современные проблемы источниковедения...; Источниковедение: Теория. История. Метод. Источники российской истории*. Москва, 1998. С. 9; Беленький И. Л. Источниковедение историческое. В кн.: *Источниковедение: Проблемные лекции. Учебно-методический модуль*. Москва, 2005. С. 100–112; Григорьева И. В. *Источниковедение новой и новейшей истории стран Европы и Америки: Учебник для студентов*. Москва, 1984. С. 7.

специфических методов исследования (многие из которых на новом технологическом уровне просто воспроизводят традиционные методы источниковедческого анализа и синтеза¹⁰) в компьютерном источниковедении еще не создает новую отрасль знания. Поэтому можно говорить лишь о *направлении* в рамках исторического источниковедения, а не о новой самостоятельной специальной (вспомогательной) исторической дисциплине.

Нельзя при этом не отметить, что выделение компьютерного источниковедения как направления в рамках общего исторического источниковедения возможно лишь в относительно кратковременной перспективе – пока компьютерные технологии являются чем-то «новым» и «необычным» для значительной части историков и источниковедов¹¹.

Относительная самостоятельность компьютерного источниковедения в рамках общего исторического источниковедения обусловлена задачами, которые новейшие компьютерные технологии позволяют решать на качественно более высоком технологическом уровне: изучение структуры источниковой информации, извлечение из источника косвенной (скрытой) информации с целью максимально полного выявления его информационного потенциала¹², целенаправленный поиск информации (реализация произвольных запросов), всесторонняя проверка полноты, надежности и достоверности свидетельств исторических источников, агрегирование и семантическое связывание информации естественных, исторически сложившихся комплексов источников (*linkage of sources*)¹³, восполнение пробелов в источниковой информации и в конечном итоге генерация новых знаний о прошлом¹⁴ в процессе реконструкции исторических фактов и многоаспектного воссоздания исторической реально-

¹⁰ Иванов А. С. Работа с XML-документом как воспроизведение основных этапов источниковедческой критики...

¹¹ Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения (на примере формулярного анализа документов приказного делопроизводства). В кн.: *Круг идей: Алгоритмы и технологии исторической информатики: Труды IX конференции Ассоциации «История и компьютер»*. Москва; Барнаул, 2005. С. 241–242.

¹² Иванов А. С., Варфоломеев А. Г. Идеи И. Д. Ковальченко об информационной ценности и неисчерпаемости источника...

¹³ Иванов А. С. Источниковедческие и археографические аспекты реконструкции исторических комплексов источников (отдел «Moscovitica–Ruthenica» в бывшем архиве Рижского магистрата). В кн.: *Сословия, институты и государственная власть в России. (Средние века и раннее Новое время): Сборник статей памяти академика Л. В. Черепнина*. Москва, 2010. С. 97–105.

¹⁴ Варфоломеев А. Г., Кравцов И. В. Приобретение и представление знаний в сетевом сообществе исследователей текстов. В кн.: *Вторая конференция «Системный анализ и информационные технологии» САИТ-2007, Обнинск, 10–14 сентября 2007 г.: Тр. конф.: в 2 т. Т. 1*. Москва, 2007. С. 104–106.

сти. В перспективе эффективное решение данных задач возможно при условии широкой репрезентации исторических источников (точнее – их информации) в семантических сетях¹⁵, использования технологий сетевых сообществ для организации распределенной обработки больших комплексов источников¹⁶ и обеспечения воспроизводимости исследований источников с помощью сохранения методик и результатов исследований в рамках баз знаний сетевых сообществ. К тому же семантические публикации исторических источников ценны и сами по себе, так как они вводят в интенсивный научный оборот новые группы источников. При этом следует отметить, что при представлении источников в семантических сетях не теряется и их «индивидуальность», что позволяет, с одной стороны, исследовать агрегированную информацию совокупностей источников, а с другой – углубленно изучать отдельные исторические источники.

Компьютерное источниковедение может быть полезным и при решении многих частных задач прикладного источниковедения, перечислить которые не представляется возможным. О потенциальных возможностях компьютерного источниковедения свидетельствуют проекты по созданию источник-ориентированных информационных систем для репрезентации и распознавания текстов¹⁷ и источниковедческого анализа исторических источников¹⁸, текстологические иссле-

¹⁵ См., например: Ahonen, E., Hyvönen, E. Publishing Historical Texts on the Semantic Web – A Case Study. In: *Proceedings of the Third IEEE International Conference on Semantic Computing (ICSC2009)*. Berkeley, 2009. Pp. 167–173; Варфоломеев А. Г., Иванов А. С. Семантические публикации информации исторических источников на основе технологии Wiki. In: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XX starptautisko zinātnisko lasījumu materiāli. Vēsture XIV*. Daugavpils, 2011. 339.–347. lpp.; Варфоломеев А. Г., Иванов А. С. Семантические публикации комплексов исторических источников. В кн.: *Информационные технологии и письменное наследие. El'Manuscript-10: Материалы международной научной конференции*. Уфа; Ижевск, 2010. С. 42–46; Иванов А. С., Варфоломеев А. Г. Публикация и анализ рукописных исторических документов с помощью технологии XML. *Humanitāro Zinātņu Vēstnesis*. 2009, Nr.16. 69.–84. lpp.; Варфоломеев А. Г., Иванов А. С. Модели для адекватной репрезентации неполной и/или противоречивой исторической информации в исторической семантической сети: К постановке проблемы. In: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XIX starptautisko zinātnisko lasījumu materiāli. Vēsture XIII*. Daugavpils, 2010. 210.–216. lpp.

¹⁶ Varfolomeyev, A., Soms, H., Ivanovs, A. Knowledge-Based Information Systems...

¹⁷ Гагарина Д. А., Кирьянов И. К., Корниенко С. И. Историко-ориентированные информационные системы: опыт реализации «Пермских» проектов. *Вестник Пермского университета. История*. 2011. Вып. 2 (16). С. 35–39.

¹⁸ См., например: Кирьянов И. К., Корниенко С. И. Стенографические отчеты Государственной Думы начала XX века: от традиционного к компьютерному источниковедению. *Вестник Челябинского государственного университета*. 2009. № 6 (144). (История, вып. 30). С. 144–148; Гагарина Д. А., Корниенко С. И. Рукописные и старопечатные кириллические книги в Интернете. *Вестник Пермского университета. История*. 2009. Вып. 3 (10). С. 112–118; Пигалева С. В., Корниенко С. И., Гагарина Д. А., Горбачева Н. Г. Газета «Пермские губернские ведомости» (1838–

дования письменных исторических источников¹⁹, электронные критические издания памятников письменности²⁰, специальные работы по атрибуции литературных произведений и других памятников письменности²¹, труды по вспомогательным историческим дисциплинам, в которых разрабатываются, например, новые методы кластеризации филигранных²², формулярного анализа актового и делопроизводственного материала²³ и многие другие вопросы.

В настоящей работе делается попытка раскрыть потенциальные возможности новейших компьютерных технологий в репрезентации, обработке, изучении и интерпретации информации слабоструктурированных уникальных средневековых источников по истории России. Следует отметить, что оценка потенциальных возможностей компьютерного источниковедения возможна лишь в контексте развития методологии общего исторического источниковедения, поэтому в монографии затрагиваются методологические вопросы повышения информационной отдачи письменных исторических источников как одной

1844 гг.): традиционный источниковедческий анализ и информационные технологии. *Вестник Пермского университета. История*. 2008. Вып. 7 (23). С. 72–76.

¹⁹ Бородкин Л. И., Милов Л. В. О некоторых аспектах автоматизации текстологического исследования (Закон Судный людем). В кн.: *Математические методы в историко-экономических и историко-культурных исследованиях*. Москва, 1977. С. 230–279; Бородкин Л. И. Математические методы классификации древних текстов. В кн.: *Методы количественного анализа текстов нарративных источников: Сб. статей*. Москва, 1983. С. 8–30; Бабалык М. Г., Варфоломеев А. Г. О применении XML-технологий в текстологическом анализе списков древнерусского апокрифа «Беседа трех святителей». В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 35. *Материалы XI конференции АИК*. Москва; Барнаул, 2008. С. 51–52; Варфоломеев А. Г., Бабалык М. Г., Пигин А. В. О проекте виртуальной среды для исследования списков «Беседы трех святителей». В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XII Всероссийской научной конференции “RCDL’2010”*. Казань, 2010. С. 551–556. Как правило, алгоритмы исследования текстов заимствуются из кладистики – биологической научной дисциплины, занимающейся построением эволюционных (филогенетических) деревьев. См., например: O’Hara, R., Robinson, P. Computer-Assisted Methods of Stemmatic Analysis. In: Blake, N., Robinson, P., eds. *The Canterbury Tales Project Occasional Papers*. Oxford, 1993. Vol. 1. Pp. 53–74. <http://rjohara.net/cv/1993-ctp/> (05.12.2012); Roos, T., Heikkilä, T. E. Evaluating Methods for Computer-assisted Stemmatology Using Artificial Benchmark Data Sets. *Literary and Linguistic Computing*. 2009. Vol. 24, no. 4. Pp. 417–433.

²⁰ См., например: Баранов В. А., Дубовцев С. В. Электронное критическое издание средневекового славянского текста: модель данных и визуализация лингвистических единиц. *Интеллектуальные системы в производстве*. 2010. № 1. С. 280–287.

²¹ *От Нестора до Фонвизина: новые методы определения авторства*. Москва, 1994.

²² Елизаров А. И., Есипова В. А., Калайда В. Т., Климкин В. М. Кластеризация филигранных по численным характеристикам изображений. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 31. Москва, 2003. С. 223–234.

²³ Ivanovs, A., Varfolomeyev, A. Editing and Exploratory Analysis of Medieval Documents...; Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения...

из центральных проблем современного источниковедения, в том числе и компьютерного источниковедения (глава 1). Одним из основных условий эффективного применения инструментария компьютерного источниковедения является комплексный подход к письменным историческим источникам, который позволяет конкретизировать общие источниковедческие задачи применительно к отдельным исторически сложившимся комплексам источников. Соответственно, в монографии в теоретическом плане затрагиваются некоторые общие вопросы комплексного источниковедения и возможности его реализации в компьютерном источниковедении; при этом дается характеристика двух основных комплексов исторических источников, послуживших эмпирической основой для апробации методов и технологий компьютерного источниковедения (глава 2). Авторы данной работы осознают, что всесторонняя и конкретная характеристика всех (или хотя бы наиболее распространенных) технологий и методов компьютерного источниковедения в рамках одной монографии вряд ли возможна²⁴. Поэтому в последующих главах они обращаются к базовым подходам и новейшим методам репрезентации и изучения источниковой информации, возникновение и развитие которых связано с появлением и прогрессом современных технологий семантических сетей (Semantic Web). По мнению авторов монографии, характерная особенность данных подходов и методов – это их универсализм, возможность применения как для анализа отдельных источников любых типов, видов и разновидностей, так и для системного изучения комплексов источников, что полностью отвечает требованиям историков и источниковедов к технологиям и методам источниковедческой критики. В качестве такого универсального метода репрезентации, обработки, интерпретации и, главное, связывания информации исторических источников может рассматриваться так называемая разметка текстов (а также метаинформации) исторических источников (глава 3). Разметка источников является одним из предварительных условий для их представления в семантических сетях, в том числе и прежде всего в виде семантических публикаций, которые сами по себе являются многообещающим инструментом их всестороннего исследования (глава 4).

²⁴ Подобные задачи могут ставиться лишь в учебных пособиях и в обобщающих руководствах по применению компьютерных методов в гуманитарных науках. См., например: Schreibman, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities*. Malden (MA); Oxford; Carlton, 2004; *Историческая информатика: Учебное пособие* [Москва], 1996; *Информационные технологии для историков: Учебное пособие к практикуму по курсу «Информатика и математика»*. Москва, 2006.

Апробация подходов и методов компьютерного источниковедения в монографии на основе комплексов средневековых документальных источников может иметь методологическое и прикладное значение также и для источниковедения истории нового и новейшего времени: если подобные технологии и подходы, подразумевающие определенную унификацию и стандартизацию информации источников, применимы в самых консервативных вспомогательных (специальных) исторических дисциплинах, то тем больше возможностей и перспектив открывается при их внедрении в анализ, синтез и интерпретацию информации исторических источников нового и новейшего времени.

Глава 1

Повышение информативной отдачи источников как центральная задача исторического источниковедения

Внедрение количественных (математических) методов и компьютерных технологий в историческую науку²⁵, в том числе в историческое источниковедение, изначально носило и по-прежнему носит противоречивый характер. С одной стороны, можно с полным основанием утверждать, что исследования, опирающиеся на компьютерные технологии и методы, уже образуют сравнительно самостоятельное, довольно развитое и популярное направление в гуманитарных науках (e-humanities). Об этом свидетельствуют и разветвленная инфраструктура, включающая в себя ассоциации, научные центры и общества²⁶, которые способствуют внедрению компьютерных технологий и математических методов в гуманитарные исследования, и регулярно проводимые международные научные конференции,²⁷ и рост числа специальных публикаций, и, разумеет-

²⁵ Об истории внедрения математических методов и компьютерных технологий в исторические исследования см. подробнее в обобщающих концептуальных очерках: Hockey, S. The History of Humanities Computing. In: Schreibman, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities...* Pp. 3–19; Boonstra, O., Breure, L., Doorn, P. *Past, Present and Future of Historical Information Science*. Amsterdam, 2004; Бородкин Л. И. Историческая информатика: Этапы развития. *Новая и новейшая история*. 1997. № 1. С. 3–22; Гарскова И. М. Основные направления развития исторической информатики...

²⁶ В последнее время в рамках e-humanities наиболее существенная роль в координации исследований и в организации процесса обмена идей принадлежит следующим организациям: The Alliance of Digital Humanities Organizations (ADHO), см. сайт данной организации: <http://adho.org/> (06.12.2012); The European Association for Digital Humanities, <http://www.allc.org/> (06.12.2012); The Association for Computers and Humanities (ACH), <http://www.ach.org/> (06.12.2012); Digital Medievalist, <http://www.digitalmedievalist.org/> (06.12.2012); The Australasian Association for Digital Humanities, <http://aa-dh.org/> (06.12.2012) и ряду других. Сравнительно долго, начиная с середины 80-х годов XX века, существовала международная организация, целью которой было способствовать внедрению информационных технологий в исторические исследования (Association for History and Computing – АНС). К сожалению, в последние годы действует только региональное отделение этой организации в России (с 1992 года) и в ряде стран СНГ – Ассоциация «История и компьютер» (АИК). См.: <http://aik-sng.ru/> (06.12.2012).

²⁷ Наиболее представительными можно считать международные конференции «Digital Humanities», которые проводятся ежегодно (поочередно в Европе и США) и объединяют представителей практически всех областей гуманитарного знания. Однако нельзя не отметить, что подавляющее большинство участников данных конференций – лингвисты и филологи. Последняя конференция была организована 16–22 июля 2012 года в Гамбурге. См.: <http://www.dh2012.uni-hamburg.de/> (06.12.2012). Можно также упомянуть специализированные конференции, например, конференцию «Digital Diplomats», посвященную методам дипломатического исследования актовых источников и вопросам репрезентации средневеко-

ся, стремительное увеличение количества ресурсов в сети Интернет, предназначенных для специалистов в области гуманитарных наук.

С другой стороны, с 80-х – 90-х годов XX века не прекращаются дискуссии о пределах и сферах применения точных, математических методов и компьютерных технологий в гуманитарных, в том числе и исторических, исследованиях, о сочетаемости этих методов и технологий с традиционными приемами и методами исторического исследования²⁸, о перспективах их применения в исторической науке и о возможности с их помощью полноценной реконструкции прошлого²⁹. В некоторых публикациях даже выражается сомнение в перспективах внедрения компьютерных технологий и математических методов в изучение истории, при этом отмечается, что характерные для 90-х годов XX века намерения тесно связать историческую науку и информатику так и не воплотились в жизнь³⁰.

В сущности, в центре внимания участников дискуссий – ключевые методологические вопросы о возможности и необходимости соответствия исторического знания критериям научности, о точности и верифицируемости данных исторической науки. Нельзя не отметить, что подобные дискуссии в исторической науке происходят регулярно. Одна из первых усилила кризис и во многом обусловила закат позитивистской историографии в конце XIX – начале XX века³¹, последняя же началась в 70-х – 80-х годах XX века³² и фактически продолжается по сей день. Современные попытки пересмотреть роль математических методов и компьютерных технологий в изучении истории не являются случайными; представляется, что сама природа ис-

вых актов в сети Интернет. См., например, информацию о последней подобной конференции (Digital Diplomats 2011: Tools for the Digital Diplomatist), состоявшейся в сентябре – октябре 2011 года в Неаполе: <http://www.cei.lmu.de/digdipl11/> (06.12.2012). Для развития исторической информатики особое значение имеют международные конференции, которые регулярно, раз в два года, проводятся Ассоциацией «История и компьютер».

²⁸ См., например: Thaller, M. Controversies around the Digital Humanities: An Agenda. *Historical Social Research*. 2012. Vol. 37, no. 3 (A Special Issue: Controversies around the Digital Humanities). Pp. 7–23.

²⁹ Бородкин Л. И. Квантитативная история в системе координат модернизма и постмодернизма. *Новая и новейшая история*. 1998. № 5. С. 3–16; Бородкин Л. И. Клиометрика: pro et contra (виртуальный диалог). В кн.: *Экономическая история. Обзор*. Вып. 7. Москва, 2001. С. 114–132; Thomas, W. G. Computing and the Historical Imagination. In: Schreibman, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities...* Pp. 56–68; Hughes, L. Conclusion: Virtual Representations of the Past – New Research Methods, Tools and Communities of Practice. In: Greengrass, M., Hughes, L., eds. *The Virtual Representation of the Past*. [S.l.], 2008, reprinted in 2010. Pp. 191–201.

³⁰ Boonstra, O., Breure, L., Doorn, P. *Past, Present and Future...* Pp. 88–89.

³¹ См.: Кроче Б. *Теория и история историографии*. Москва, 1998. С.185–186.

³² Stone, L. *The Past and the Present*. Boston, 1981; Stone, L. The Revival of Narrative: Reflections on a New Old History. *Past and Present*. 1979. Vol. 85, no. 1. Pp. 3–24.

торического знания и характер конечного «продукта» научно-исследовательской деятельности историка – описательной реконструкции прошлого³³ – существенно ограничивают сферу применения аналитических, чисто исследовательских процедур³⁴ и неизбежно склоняют историков к нарративному жанру, апробированному уже давно, в период зарождения историописания. Этим объясняется и консерватизм исторической науки, который гораздо сильнее стремления отдельных историков находится в русле современных методологических исканий, использовать новейшие методы и подходы в своей исследовательской практике. При этом не обязательно происходит полный отказ от нового исследовательского инструментария, так как одно из проявлений консерватизма – утилитарный подход к технологиям и методам исследования, приводящий к тому, что новейшие методы могут применяться лишь постольку, поскольку они способны дать значимый результат для исторической науки. То есть новая технология или метод исследования, с точки зрения историка, не самоцель, а лишь средство. Если технология и выглядит устаревшей, но при этом по-прежнему способствует решению задач историка, то это не является причиной для замены ее новейшей. Такой «минимализм» и самоограничение исторической науки, с одной стороны, вполне закономерны, а с другой – таят в себе опасности, главная из которых – «не заметить» какие-либо новые методы исследования, предлагаемые компьютерным источниковедением, исторической информатикой и смежными научными дисциплинами, особенно в условиях «лавинообразного» умножения компьютерных технологий. В результате будут безвозвратно упущены и возможности, которые открываются благодаря этим методам и технологиям.

К тому же современные компьютерные технологии могут внести определенный вклад в решение основной методологической проблемы исторической науки и тем самым способствовать достижению идеала (или, точнее, приближению к идеалу) исторического познания – по возможности всесторонней и достоверной реконструкции прошлого на основе информации исторических источников. Нельзя не заметить, что периодически повторяющиеся кризисы в исторической науке во многом связаны с разочарованием историков в возможности

³³ Ср. подходы к организации и репрезентации итогов исторического исследования, обоснованные в разное время в трудах теоретиков исторической науки: Дройзен И. Г. *Историка*. СПб., 2004 (1858). С. 494-500; Риккерт Г. *Границы естественнонаучного образования понятий. Логическое введение в исторические науки*. СПб., 1997 (1902); Уайт Х. *Метаистория*. Екатеринбург, 2002 (1973). С. 493-496 и др.

³⁴ Риккерт Г. *Границы естественнонаучного образования понятий...* С. 247-256.

получения точной, всесторонней и объективной ретроспективной информации. Единственный носитель информации о прошлом – исторический источник, соответственно, в центре внимания историков всегда были и по-прежнему остаются вопросы полноты, объема и качества – достоверности, степени объективности (или, напротив, субъективности) – свидетельств исторических источников. Поэтому поиск путей и средств повышения информативной отдачи исторических источников в течение уже длительного времени остается одной из насущных задач не только исторического источниковедения, но и исторической науки в целом.

Однако содержание исследовательских операций, входящих в традиционный источниковедческий анализ и синтез, остается, в сущности, неизменным с конца XIX – начала XX века³⁵. Первые же конструктивные попытки решить проблему повышения информативной отдачи исторических источников были сделаны лишь в рамках методологии неопозитивизма во второй половине XX века представителями таких направлений в мировой историографии, как квантитативная история и «новая история» (также «новая социальная история»)³⁶. Актуализация вопросов, связанных с источниковой информацией, не была случайной: данные школы в историографии пытались на основе междисциплинарных связей с социологией, математикой и информатикой сделать изучение прошлого максимально точным, достоверным и объективным. Для достижения этой цели в духе социологии уточнялся объект исторического изучения, при этом приоритет отдавался системному исследованию истории социальных отношений, социальных структур и социальных групп, истории повседневности, частной жизни рядовых людей, истории общественных преобразований и социальных конфликтов³⁷, истории идей³⁸. Применительно к данным областям исследования обосновывались и новые научные методы, опирающиеся на возможности математических и компьютерных технологий³⁹, благодаря которым у истории должен был появиться шанс стать в определенном смысле «точной» наукой⁴⁰. При этом создается

³⁵ См., например, работу французских историков, давно уже ставшую классической: Ланглуа Ш.-В., Сеньобос Ш. *Введение в изучение истории*. СПб., 1899.

³⁶ В концентрированном виде методология «новой истории» представлена в следующих работах: Himmelfarb, G. *The New History and the Old*. Cambridge (MA), 1987; Rabb, T. K., Rotberg, R. I., eds. *The New History, the 1980-s and Beyond: Studies in Interdisciplinary History*. Princeton (NJ), 1982.

³⁷ Burke, P. *Sociology and History*. London; Boston, 1980. Pp. 13, 31.

³⁸ См., например: Ратмэн Д. Б. «Новая социальная история» в США. *Новая и новейшая история*. 1990. № 2. С. 72.

³⁹ Там же. С. 77.

⁴⁰ Aydelotte, W. O. Quantification in History. In: Rowney, D. K., Graham, J. Q., eds. *Quantitative History: Selected Readings in the Quantitative Analysis of Historical Data*. Homewood (Ill.), 1969.

впечатление, что специфика неопозитивистских школ в историографии, в том числе и «новой истории», была связана не столько с выбором объектов изучения, сколько с применением современных методов, среди которых особое значение придавалось созданию и оперированию компьютерными (реляционными) базами данных.

К сожалению, представители направления «новой (социальной) истории» подходили к решению проблемы информативной отдачи исторических источников чисто эмпирически, в рамках конкретных «ситуационных» исследований (case studies). Возможно, именно поэтому их наработки, связанные с применением количественных методов в критике источников, не привели к какому-либо значимому прорыву в методологии источниковедения, а поэтому не стали надежной основой для дальнейшего системного и целенаправленного внедрения компьютерных технологий в источниковедческий анализ и синтез.

В качестве прорыва в методологии источниковедения, благодаря которому были заложены теоретические и методические основы для дальнейшего продуктивного применения математических методов и компьютерных технологий в изучении, репрезентации и верификации информации исторических источников, можно с полным основанием считать труды академика И. Д. Ковальченко, в творческом наследии которого важное место занимают поиски путей и средств повышения информативной отдачи исторических источников. Данное направление творческого поиска было четко обозначено в 1982 году в программной статье «Исторический источник в свете учения об информации»: «Главная из прикладных задач источниковедения состоит во все более явственно усиливающейся потребности повышения информативной отдачи источников. Такая потребность существовала всегда. Она обусловлена тем, что всегда имеет место определенное несоответствие между той информацией, которая необходима историку для изучения тех или иных явлений или процессов, и тем, что непосредственно отражено в источниках»⁴¹.

Хотя в классическом источниковедении более распространен термин «свидетельства исторических источников», введение в *активный* научный оборот понятия «информация исторических источников» (что является бесспорной заслугой академика И. Д. Ковальченко), вполне обоснованно, поскольку применение положений учения об информации в источниковедении не только подтверждает

Р. 3. Отметим, что подобный (возможно, чрезмерный) оптимизм был характерен скорее для 60–70-х годов XX века, чем для более позднего времени.

⁴¹ Ковальченко И. Д. Исторический источник в свете учения об информации... С. 131. См. также: Ковальченко И. Д. *Методы исторического исследования*... С. 125–126.

и углубляет существующие представления об историческом источнике⁴², но и способствует внедрению в традиционные методы внешней и внутренней критики источников современного исследовательского инструментария, опирающегося в том числе на математический аппарат и компьютерные технологии.

Четкое определение основной проблемы любого источниковедческого исследования как *информационной* было принципиально новым положением в теории и методологии источниковедения – здесь наблюдается выход за рамки дискурса классической историографии XIX – первой половины XX века и перевод вопросов изучения исторических источников в сферу исследований информации. Данный подход к изучению исторических источников как бы снимает границу между исторической наукой и «точными» науками, поставленную еще в конце XIX века. В результате в историческом исследовании могут применяться общие теоретико-методологические и, соответственно, практические подходы, используемые во всех областях знания, имеющих дело с информацией – с ее возникновением, фиксации, сохранением, утратой, изменением, передачей, восприятием и т.д. При этом можно утверждать, что подход к изучению информации исторических источников, обоснованный И. Д. Ковальченко и апробированный им на обширном эмпирическом материале, полностью соответствовал (и по-прежнему соответствует) новейшим мировым тенденциям в развитии теории и методологии исторического исследования. Академик И. Д. Ковальченко одним из первых историков в СССР плодотворно применял количественные методы к изучению свидетельств исторических источников, в том числе и к выявлению, репрезентации и анализу их структурной информации⁴³. Тем самым он стал одним из основоположников компьютерного источниковедения в российской (советской) историографии. При этом И. Д. Ковальченко рассматривал новейшие технологии прежде всего в качестве инструментов для извлечения и аналитической обработки ин-

⁴² Ковальченко И. Д. Исторический источник в свете учения об информации... С. 137–145.

⁴³ См., например: Ковальченко И. Д. Об опыте математико-статистической обработки выборочных данных о крестьянском хозяйстве России XIX в. *Вестник Московского университета*. Серия 8. История. 1966. № 1. С. 76–89; Ковальченко И. Д. О применении математико-статистических методов в исторических исследованиях. В кн.: *Источниковедение: Теоретические и методологические проблемы*. Москва, 1969. С. 115–133; Ковальченко И. Д., Милов Л. В. *Всероссийский аграрный рынок. XVIII – начало XX в.: Опыт количественного анализа*. Москва, 1974; Ковальченко И. Д. О моделировании исторических явлений и процессов. *Вопросы истории*. 1978. № 8. С. 72–93; Ковальченко И. Д., Селунская Н. Б., Литваков Б. М. *Социально-экономический строй помещичьего хозяйства Европейской России в эпоху капитализма: Источники и методы изучения*. Москва, 1982; Ковальченко И. Д. Применение количественных методов и ЭВМ в исторических исследованиях. *Вопросы истории*. 1984. № 9. С. 61–73 и др.

формации исторических источников, что полностью отвечает современной парадигме компьютерного источниковедения.

Из всех направлений исследования информации исторических источников, обозначенных в трудах академика И. Д. Ковальченко, наиболее перспективным можно считать подход к изучению структуры источниковой информации, и особенно – к извлечению скрытой (структурной, связанной) информации⁴⁴, который был, бесспорно, новаторским⁴⁵: он способствовал и по-прежнему способствуют расширению и углублению творческой, теоретической мысли историков и источниковедов⁴⁶.

К сожалению, в прикладных исследованиях теоретико-методологическое наследие И. Д. Ковальченко востребовано лишь частично. Так, в центре внимания авторов находятся вопросы внешней и внутренней критики источников, а методы извлечения, проверки и использования их структурной информации в ходе источниковедческого анализа и синтеза разрабатываются явно недостаточно. Возможно, это связано с определенной консервативностью мышления историков, большинство из которых – сторонники традиционных подходов и давно апробированных методов исследования⁴⁷. К тому же для историка, как и для любого другого специалиста, характерно стремление избирать наиболее прямой путь для достижения желаемого результата; а при этом нельзя отрицать, что использование прямых свидетельств источников (в терминологии, предложенной И. Д. Ковальченко, – выраженной информации) требует меньших затрат времени, чем тщательное системное изучение срезов информации памятников прошлого. Это означает, что в историческом исследовании доминирует «экстенсивный подход»: с целью получения новых фактов-знаний историки, как правило, просто расширяют круг использованных источников.

⁴⁴ Ковальченко И. Д. Исторический источник в свете учения об информации... С. 133.

⁴⁵ В определенном смысле, этот подход по-прежнему является новаторским, так как и в новейшем российском источниковедении недостаточная четкость в определении структуры источниковой информации приводит к тому, что рекомендации по извлечению так называемой косвенной информации приобретают неопределенный, абстрактный характер. См., например: *Источниковедение: Теория. История. Метод...* С. 140–141.

⁴⁶ Об этом свидетельствует, например, использование и развитие идей И. Д. Ковальченко в современных публикациях по историческому источниковедению: Соколов А. К., Тяжелникова В. С. Массовые источники...

⁴⁷ Не исключено, что данный тезис авторов настоящей монографии является всего лишь расхожим штампом, поскольку эмпирические исследования особенностей работы историков, специфики их «творческой лаборатории», кажется, не проводились. Публикации по данной теме опираются, как правило, на логические конструкции, основанные на методологии структурализма и постмодернизма. См., например: Копосов Н. Е. *Как думают историки*. Москва, 2001.

Представляется, что данное искусственное ограничение творческого поиска историков и источниковедов коренится также и в сомнительном теоретическом положении источниковедения об «оптимальном объеме нужной для получения достоверных знаний информации», согласно которому «избыточная» источниковая информация не способствует приращению фактического знания.

Данный тезис был выдвинут еще в советском источниковедении: «Историк, ... собирая материал, ... сталкивается с таким моментом, когда новые источники ничего существенного к уже накопленным фактам не прибавляют... Итак, практика последующего развития исторического процесса, с одной стороны, и практика научного познания, с другой, выступают критерием истинности исторических знаний и, следовательно, критерием правильности определения оптимального объема нужной для получения достоверных знаний информации»; при этом «установление момента, за которым дальнейшее собирание фактов ведет к избыточности информации, не должно происходить стихийно»⁴⁸. Фактически данной рекомендации осознанно или неосознанно следует большинство историков. Разумеется, этот тезис нельзя оценивать однозначно негативно, так как в конкретном исследовании, посвященном ограниченному кругу вопросов и аспектов исторического прошлого, закономерно приходится ограничивать объем информации⁴⁹. Однако стратегическая цель исторической науки – всесторонняя реконструкция прошлого – не позволяет согласиться с тем, что избыточная источниковая информация не способствует приращению фактического знания. Отсутствие в картине прошлого хотя бы одного фрагмента будет означать, что знание не обладает достаточной полнотой, поэтому любой аспект новой информации ценен для исторической науки⁵⁰. В контексте же источниковедения об «избыточной» информации источника вряд ли вообще можно говорить, учитывая направленность данной специальной исторической дисциплины⁵¹.

Из сомнительного тезиса об «избыточной» источниковой информации логично вытекает положение о том, что любой источник должен использоваться прагматично, для изучения четко очерченного круга вопросов в соответствии со своей прямой социальной функцией. Тем самым информационная ценность источника, зачастую, опреде-

⁴⁸ Варшавчик М. А. Вопросы логики исторического исследования и исторический источник. *Вопросы истории*. 1968. № 10. С. 87.

⁴⁹ Ср.: Ковальченко И. Д. Методы исторического исследования... С. 110.

⁵⁰ Отметим, что лишь гипотетически можно допустить возможность полного дублирования информации в различных источниках.

⁵¹ См. подробнее: Шмидт С. О. Современные проблемы источниковедения...

ляется его непосредственным целевым назначением⁵². К сожалению, здесь забывается бесспорное положение источниковедения о неисчерпаемости источника и его полифункциональности⁵³, которое в исследовательской практике означает, что в любом историческом памятнике запечатлелись самые разнообразные аспекты действительности. При этом полифункциональность исторических источников подтверждает тезис теоретического источниковедения об информационной неисчерпаемости источников: как отметил И. Д. Ковальченко, «многообразие и безграничность взаимосвязей, присущих явлениям объективного мира, обуславливают то, что в исторических источниках, несмотря на избирательное отражение ими действительности, содержится, в сущности, безграничный объем скрытой информации, характеризующей взаимосвязи явлений действительности»⁵⁴.

Итак, основная задача источниковеда – максимально полная актуализация информационного потенциала источника, включая и более глубокие пласты его информационной структуры. При этом эффективность выявления потенциальной (прежде всего – структурной) информации может быть существенно повышена при помощи современных методов и технологий работы с источниками, включающих представление источника в семантических сетях, компьютерные технологии извлечения информации и математические методы ее обработки, т.е. то, что и составляет инструментарий компьютерного источниковедения.

Возможностей повышения информативной отдачи источников много, о чем свидетельствуют в том числе новейшие тенденции в развитии компьютерного источниковедения, которые охарактеризованы далее в настоящей монографии. Однако все они реализуются только в творческой деятельности историка, поэтому не востребованная исследователем – историком или источниковедом – информация источника всегда будет оставаться потенциальной.

Хотелось бы надеяться, что использование инструментария компьютерного источниковедения в решении основной задачи общего исторического источниковедения станет стимулом и методологической основой для его развития, а также убедительным аргументом для историков и источниковедов, все еще сомневающихся в перспективах и потенциальных возможностях компьютерных технологий в исторических и источниковедческих исследованиях.

⁵² См. подробнее: Илизаров Б. С. О понятии ценности документального исторического источника. *История СССР*. 1986. № 5. С. 109.

⁵³ Тартаковский А. Г. Социальные функции источников как методологическая проблема источниковедения. *История СССР*. 1983. № 3. С. 125–126.

⁵⁴ Ковальченко И. Д. Исторический источник в свете учения об информации... С. 140.

Глава 2

Комплексный подход в компьютерном источниковедении и конкретные задачи изучения комплексов исторических источников

Комплексный подход к историческим источникам: теория и практика

Методологические положения источниковедения о путях и способах максимально полного раскрытия информативного потенциала исторических источников лишь частично могут быть реализованы при изучении отдельных, пусть даже уникальных источников, особенно если они изучаются изолированно, в отрыве от других, связанных с ними исторических источников. Как правило, применение методов аналитической и синтетической критики подразумевает использование ряда взаимосвязанных источников. Однако в традиционном источниковедении перспективы одновременного использования обширных совокупностей источников при проведении аналитических и синтетических операций существенно ограничены возможностями исследователя воспринимать, обобщать и анализировать большие объемы информации. Исследовательский процесс затрудняет также рассредоточенность исторических источников по различным архивохранилищам, а наличие всех необходимых материалов в одном архиве или даже в одном обширном фонде лишь незначительно облегчает работу. При этом традиционные («бумажные») публикации, как правило, предоставляют исследователю только небольшую часть необходимых ему источников, эффективное использование которых, в свою очередь, затруднено громоздкостью академических публикаций и несовершенством научно-справочного аппарата, который не позволяет осуществить многоаспектный поиск необходимой источниковой информации.

В отличие от традиционного источниковедения, компьютерное источниковедение, подразумевающее репрезентацию текстов исторических источников, метаинформации о них и научно-справочного аппарата в электронных источник-ориентированных базах данных⁵⁵,

⁵⁵ Об источник-ориентированных базах данных см. подробнее: Бородкин Л. И. Историческая информатика: Этапы развития. *Новая и новейшая история*. 1997. № 1. С. 3–22; Юмашева Ю. Ю. Утопия или утро завтрашнего дня? (Новые средства работы с электронными источниками). В кн.: *Новые информационные ресурсы и технологии в исторических исследова-*

а в последнее время – и в семантических сетях, дает реальную возможность проводить полноценное комплексное изучение практически неограниченных совокупностей памятников письменности. Это, бесспорно, повышает как качество источниковедческой критики, так и надежность исторической реконструкции в целом. При этом любые синтетические и аналитические операции могут быть гибко модифицированы в зависимости от задач конкретного исследования; ограничение здесь накладывает только способность историка или источниковеда ставить исследовательские задачи. В сущности, комплексный подход является естественным в компьютерном источниковедении⁵⁶, а при отсутствии более или менее обширной выборки исторических источников применение инструментария компьютерного источниковедения во многом теряет смысл.

Однако для корректного использования методов и технологий компьютерного источниковедения в изучении информации больших и малых совокупностей исторических источников необходимо опираться на точное представление о сущности, особенностях и границах естественных комплексов источников. К сожалению, в общем историческом источниковедении данные вопросы разработаны лишь частично⁵⁷.

На первый взгляд, в современном источниковедении комплексному подходу в изучении исторических источников и интерпретации их свидетельств действительно уделяется определенное внимание,

ниях и образовании. Москва, 2000. С. 261–262; Иванов А. С. Источниково-ориентированные базы данных в изучении локальной истории. В кн.: *XI Zinātnisko lasījumu materiāli: Vēstures sekcija*. 5. krājums. Daugavpils, 2001. 37.–42. lpp.; Иванов А. С. База данных как динамическая модель истории региона. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 30. Москва, 2002. С. 170–172.

⁵⁶ Данный тезис подтверждают и наработки авторов монографии по применению методов компьютерного источниковедения в изучении относительно обширных совокупностей источников. См., например: Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения...; Ivanovs, A., Varfolomeyev, A. *Editing and Exploratory Analysis of Medieval Documents...* и др.

⁵⁷ О комплексном подходе в источниковедении см. подробнее: Иванов А. С. Источниковедческие и археографические аспекты реконструкции исторических комплексов источников...; Ivanovs, A. *Complex Approach to the Research of the 1949 Deportation (the Case of Latvia)*. *Lituanistica*. 2011. T. 57, Nr. 1 (83). Pp. 106–111; Иванов А. С. Комплекс источников по истории Великого Новгорода в Латвийском государственном историческом архиве (XII–XVI вв.). В кн.: *Новгородика-2010. Вечевой Новгород*. Часть 2. Великий Новгород, 2011. С. 89–103; Ivanovs, A. *Dokumenti par Rīgas, Gotlandes un Hanzas attiecībām ar Lielo Novgorodu Latvijas Valsts vēstures arhīvā (12.–16. gs.)*. *Latvijas Arhīvi*. 2010, Nr. 2. 7.–34. lpp.; Варфоломеев А. Г., Иванов А. С. Семантические публикации комплексов исторических источников...; Ivanovs, A. *Dokumentu kompleksa arzināšana, rekonstrukcija un izpēte: arheogrāfiskie un avotpētnieciskie aspekti*. *Humanitāro Zinātņu Vēstnesis*. 2006, Nr. 10. 76.–88. lpp.; Ivanovs, A. *Vēstures avotu kompleksa rekonstrukcijas problēma: kolekcija Moscowitica–Ruthenica*. Grām.: *Vēsture: Avoti un cilvēki*. *Humanitārās fakultātes XIV starptautisko zinātnisko lasījumu materiāli*. *Vēsture VIII*. Daugavpils, 2004. 49.–57. lpp.

особенно при осуществлении перекрестной проверки надежности и достоверности информации источников, а также при использовании источниковой информации для полноценной исторической реконструкции⁵⁸. В то же время в источниковедческой практике реализация комплексного подхода иногда приобретает упрощенные формы, в результате чего сам подход частично лишается реального содержания и превращается в декларацию. Как правило, в исследованиях комплексный подход проявляется как одновременное, если не параллельное, использование ряда исторических источников, принадлежащих к разным видам, реже – типам, при этом могут применяться современные и довольно-таки оригинальные междисциплинарные методы извлечения и верификации их информации⁵⁹. В сущности, такой подход в аналитической и синтетической критике источников не является новым, так как еще в XIX веке источниковеды-теоретики писали о проверке информации одних источников свидетельствами других источников, о констатации и восполнении пробелов в историческом материале, об историческом повествовании, опирающемся на максимально полную и всестороннюю источниковую базу⁶⁰.

Принципиально новым является положение современного источниковедения об исторических, объективных связях между источниками, которые образуют комплекс или систему⁶¹. Следует отметить, что связи между источниками, образующими исторические комплексы, уже стали объектом изучения в специальных источниковедческих трудах⁶². Осмысление опыта источниковедческого изучения данного вопроса дает возможность выделить основные признаки «естественного» комплекса источников: общность происхождения источников; спонтанность формирования комплекса, иерархичность его структуры, наличие устойчивых связей между источниками. Указанные выше признаки естественного комплекса исторических источников позво-

⁵⁸ Янин В. Л. *Очерки комплексного источниковедения: Средневековый Новгород*. Москва, 1977. С. 5–21; Пронштейн А. П., Данилевский И. Н. *Вопросы теории и методики исторического исследования*. Москва, 1986. С. 62; Шмидт С. О. *Современные проблемы источниковедения...* С. 50–53.

⁵⁹ В качестве примера можно упомянуть некоторые доклады, представленные на международных конференциях «Комплексный подход в изучении Древней Руси». См.: *Древняя Русь. Вопросы медиевистики*. 2003, № 4 (14). С. 5–86; 2005, № 3 (21). С. 5–120; 2007, № 3 (29). С. 5–130; 2008, № 3 (33). С. 5–75; 2009, № 3 (37). С. 5–131; 2011, № 3 (45). С. 5–142.

⁶⁰ См.: Ланглуа Ш.-В., Сеньобос Ш. *Введение в изучение истории...* С. 159–168; Дройзен И. Г. *Историка...* С. 223–229.

⁶¹ Янин В. Л. *Очерки комплексного источниковедения...* С. 20–21; Шмидт С. О. *Современные проблемы источниковедения...* С. 52.

⁶² См., например: Черепнин Л. В. *Русские феодальные архивы XIV–XV веков*. Часть первая. Москва; Ленинград, 1948; Часть вторая. Москва, 1951; Янин В. Л. *Очерки комплексного источниковедения...*

ляют изучать эти источники системно и всесторонне, тем самым полностью реализовать методико-аналитический потенциал исторического источниковедения⁶³.

Именно поэтому реконструкцию исторических, естественных комплексов источников следует считать одной из приоритетных задач источниковедения, в том числе – и компьютерного источниковедения. Точное установление комплекса необходимо для выработки адекватной стратегии исследования и выбора методов изучения как комплекса в целом, так и отдельных источников, входящих в него. Успех применения методов компьютерного источниковедения во многом зависит от полноты комплекса источников, равно как и от его обширности, так как предполагается или восполнение отсутствующей в одном источнике информации за счет информации других источников, или же вероятностное, гипотетическое моделирование фрагментов реальности – как исторических фактов, так и самих источников, которые до нас не дошли, – на основе логической, ситуационной, функциональной интерпретации информации сохранившихся источников. К тому же адекватная реконструкция комплекса является необходимым условием для подготовки высококачественной публикации источников как традиционной («бумажной»), так и электронной.

Реконструкция комплекса источников, в сущности, всегда в большей или меньшей степени гипотетична, поскольку она требует воспроизведения среды возникновения источников и контекста их дальнейшего исторического бытования⁶⁴, при этом в центре внимания историков должны быть связи между источниками, складывавшиеся в процессе выполнения ими своих непосредственных, первичных функций⁶⁵.

Конкретные способы реконструкции комплексов исторических источников уже давно разрабатываются в одной из вспомогательных исторических дисциплин, а именно – в археографии, поскольку любую публикацию исторических источников, которая претендует на полноту и всесторонность, можно, в определенном смысле, рассматривать как попытку реконструкции комплекса источников. Тем самым в ходе многолетней археографической деятельности были

⁶³ См. подробнее: Иванов А. К вопросу о повышении информативной отдачи документов приказного делопроизводства в исторических исследованиях. В кн.: *Проблемы источниковедения и историографии*. Калининград, 2001. С. 22–27; Иванов А. С., Кузнецов А. М. Комплексный подход к изучению документов приказного делопроизводства. *Древняя Русь: Вопросы медиевистики*. 2003. № 4 (14). С. 30–31.

⁶⁴ Черепнин Л. В. *Русские феодальные архивы XIV–XV веков*. Часть первая. С. 5–6.

⁶⁵ Об изучении источников в контексте их функционирования в реальной жизни прошлого см. подробнее: Тартаковский А. Г. Социальные функции источников...

фактически верифицированы различные подходы к установлению исторических комплексов источников⁶⁶, которые обобщены в трудах по теории и практике археографии⁶⁷.

В археографической практике наиболее распространенным является исторически-тематический принцип в определении совокупности публикуемых документов⁶⁸. В данном случае составители публикаций весьма произвольно формируют комплексы источников в соответствии с жестко заданной программой издания и априорной исторической интерпретацией, которую информация источников определяет лишь в незначительной степени. Такие публикации источников напоминают хрестоматии. Нельзя не отметить, что на основании таких же подходов формируется и источниковая база многих исторических трудов, что свидетельствует об определенном отрыве практики исторического исследования от теоретического источниковедения.

В качестве реконструкций комплексов исторических источников можно условно рассматривать также систематические публикации определенных видов или разновидностей источников – актов, отписок, указных грамот, договорных грамот, писем и т.п. Однако здесь доминирует формальный критерий, который однозначно ограничивает выборку публикуемых источников принадлежностью их к определенному виду или разновидности, тем самым «отсекаются» все прочие источники других видов и разновидностей.

Третий подход к установлению комплексов источников – региональный. В отличие от двух вышеупомянутых подходов, региональный подход в наибольшей степени соответствует идее комплексного

⁶⁶ К сожалению, вопросы развития теории и практики отбора исторических источников к публикации лишь фрагментарно освещены в специальной литературе. См., например: Козлов В. П. *Российская археография конца XVIII – первой четверти XIX века*. Москва, 1999. С. 306–315; Чирков С. В. *Археография в творчестве русских ученых конца XIX – начала XX века*. Москва, 2005. С. 94–101, 149–160; Королев Г. И. *Медиевистическая археография за рубежом: Труды XVI–XVIII веков*. Москва, 2001; Королев Г. И. *Медиевистическая археография за рубежом: Труды XIX – начала XX в.* Москва, 2003.

⁶⁷ См.: *Отбор исторических источников для публикации в различных типах и видах изданий документов. (Методические рекомендации)*. Москва, 1986; *Правила издания исторических документов в СССР*. Москва, 1990. С. 21–35; Козлов В. П. *Основы теоретической и прикладной археографии*. Москва, 2008. С. 57–91. Данные издания содержат ценные рекомендации, которые необходимо учитывать в практике реконструкции естественных комплексов исторических источников.

⁶⁸ О подходах к воссозданию (в отдельных случаях – произвольному конструированию) комплексов источников в археографии см. подробнее: Иванов А. С. *Источниковедческие и археографические аспекты реконструкции исторических комплексов источников...*; Ivanovs, A. *Dokumentu kompleksa arzināšana, rekonstrukcija un izpēte...*; Ivanovs, A. *Vēstures avotu kompleksa rekonstrukcijas problēma...* и др.

подхода в источниковедении⁶⁹. Перспективы реализации комплексного подхода в репрезентации и изучении информации по региональной (локальной) истории обусловлены тем, что любой исторический регион представляет собой самодостаточный и в то же время компактный объект изучения для историка⁷⁰. Соответственно, комплекс источников, относящийся к истории региона, является органичным и сравнительно небольшим, а поэтому может быть более или менее всесторонне представлен как в традиционных (печатных) изданиях источников, так и в электронных публикациях, которые включены в источник-ориентированные базы данных⁷¹. Именно поэтому в публикациях исторических источников, относящихся к какому-либо конкретному региону, естественный исторический комплекс документов воспроизводится точнее и полнее, чем, например, в тематических изданиях.

Однако и в данном случае системы источников, представленные в публикациях, нельзя рассматривать в качестве полноценных реконструкций реально существовавших комплексов документов. Нельзя отрицать, что сама принадлежность совокупности источников к определенной территории, региону, государственному образованию в рамках определенного периода уже подразумевает наличие некоторых устойчивых связей внутри этой совокупности. В то же время, региональная выборка источников игнорирует более тесные, более значимые связи между источниками; для канцелярских документов это, например, связи, сложившиеся в делопроизводственной практике – в ходе оборота документов. Поэтому реконструкция комплекса на основе чисто регионального подхода может стать формальной и даже искусственной.

Характеристика основных подходов к установлению комплексов источников, апробированных в археографической практике, позволяет утверждать, что существует значительный разрыв между теорией и практикой как в источниковедении, так и в археографии: в работе с источниками теория требует следовать принципу ком-

⁶⁹ См. подробнее: Иванов А. С. Комплексное источниковедение региональной истории. В кн.: *Научные труды Рижского института мировой экономики*. Вып. 2. Рига, 1998. С. 71–75.

⁷⁰ Ivanovs, A. Reģionālisma problēma Latgales vēstures kompleksajā izpētē. (Problēmas nostādne). Grām.: *IX Zinātnisko lasījumu materiāli. Vēstures sekcija*. 3. krājums. Daugavpils, 2000. 9.–10. lpp.

⁷¹ Моделирование такой базы данных принципиально не отличается от формирования традиционного сборника документов. Разработка структуры базы данных – это чисто логическая операция, которая опирается на принятую в историографии модель истории региона, т.е. на реконструкцию идеального образа края. См.: Иванов А. С. Методологические проблемы создания базы данных для региональных исследований: (Пример Латгальского края). В кн.: *Новые информационные ресурсы и технологии в исторических исследованиях и образовании*. Москва, 2000. С. 71–72.

плексности, всесторонности и полноты, на практике же лишь небольшая часть источников публикуется, изучается и используется в исторических исследованиях⁷². Как правило, отбор источников и в археографической, и в источниковедческой практике носит случайный характер или производится по формальным критериям.

Представляется, что основная причина разрыва теории и практики – подчинение источниковедческой практики и археографической деятельности господствующим в исторической науке представлениям и интерпретациям: специалисты зачастую вынуждены заниматься поисками примеров для иллюстрации положений своей школы в историографии⁷³. Последовательная реализация иллюстративного, равно как и формально-видового или регионального подхода в источниковедении и археографии приводит к тому, что исторические источники искусственно «вырываются» из контекста своей эпохи и своего исторического существования, игнорируется их место в системе других источников, что не позволяет полноценно раскрыть их специфические функции.

В качестве альтернативы иллюстративному подходу можно предложить метод «сплошного» изучения и, соответственно, публикации естественных (реальных) исторических комплексов источников, что позволяет в значительной мере исключить элемент субъективности при проведении аналитической и синтетической критики источников, а также при их отборе к публикации и формировании источниковой базы исследований. При этом в отдельных случаях нет необходимости проводить какую-либо специальную работу по реконструкции исторических комплексов источников, так как, например, пофондовые издания документов⁷⁴ могут опираться на существующие в наше время реальные исторические комплексы документальных исторических источников – архивные фонды⁷⁵.

⁷² Здесь нельзя не отметить, что введение в активный научный оборот именно средневековых источников сильно зависит от качества и полноты их публикаций. Подтверждением тому могут служить многочисленные исторические труды, источниковая база которых обычно не выходит за рамки корпуса опубликованных источников. Именно это и определяет особую ответственность источниковеда и археографа, необходимость в своей работе опираться на надежные реконструкции реальных комплексов источников.

⁷³ Добрушкин Е. М. Археографическая теория и источниковедческая практика (к вопросу о взаимосвязи). В кн.: *Россия в IX–XX веках*. Москва, 1999. С. 131–132.

⁷⁴ См. подробнее: *Отбор исторических источников...* С. 55–59.

⁷⁵ Закономерности формирования архивных фондов, благодаря чему в архивохранилищах концентрируются естественные комплексы исторических источников, подробно рассмотрены в специальной архивоведческой литературе. См., например: Старостин Е. В. *Зарубежное архивоведение: Проблемы истории, теории и методологии*. Москва, 1997. С. 49–62; Duchein, M. Theoretical Principles and Practical Problems of Respect des Fonds in Archival Science.

Однако в других случаях реконструкция исторических комплексов требует определенных усилий, при этом результаты работы зависят не только от настойчивости и квалификации источниковеда, но и от ряда объективных условий: рассредоточенности документов комплекса по различным фондам и архивохранилищам, степени его сохранности и полноты, наличия старых описей, реестров и других источников, на основе изучения которых можно воссоздать комплекс, и т.д. Нужно также учитывать и специфику самого комплекса, обусловленную его видовым составом, зависимостью от социальной, политической и культурной среды, в которой он изначально сформировался.

Несмотря на то, что разнообразие исторических комплексов существенно затрудняет выработку единых рекомендаций по их реконструкции, все же можно назвать самый общий подход, на основе которого возможно воссоздание утраченных или неполных комплексов памятников письменности. Это подход, который еще в 40-е годы XX века был успешно и последовательно реализован академиком Л. В. Черепниным в работе по воссозданию цельных архивных фондов из русских феодальных архивов⁷⁶. Данный подход можно назвать функциональным, поскольку существовавшая ранее система источников восстанавливается на основе тщательного изучения их функционирования в конкретных условиях своего времени. Естественно, что наибольшее внимание при этом уделяется месту каждого отдельного памятника в системе делопроизводства и документооборота, возникновению и дальнейшей эволюции архивных собраний, в которых отложились документы, истории текстов источников в связи с конкретно-историческими условиями их создания и использования.

Как уже отмечалось во введении к данной монографии, апробация потенциальных возможностей и перспектив применения инструментария компьютерного источниковедения авторами производилась на конкретном материале, который представляет два естественных комплекса исторических источников. Характеристика этих комплексов также позволяет в дальнейшем сформулировать общие задачи изучения исторических источников, наиболее эффективное решение которых может предложить именно компьютерное источниковедение.

Archivaria. Summer 1983, Vol. 16. Pp. 64–82; Abukhanfusa, K., ed. *The Principle of Provenance*. Stockholm, 1994 и многие др.

⁷⁶ Черепнин Л. В. *Русские феодальные архивы XIV–XV веков...* К сожалению, идеи Л. В. Черепнина об органичных, системных связях источников, когда-то входивших в определенные комплексы, долгое время не были в полной мере востребованы в российском источниковедении.

Комплекс «Moscovitica–Ruthenica» и субкомплексы смоленско-рижских актов XIII – первой половины XIV века и полоцких грамот XV века

Комплекс средневековых документальных источников, с 30-х годов XVII века известный под историческим названием «Moscovitica–Ruthenica»⁷⁷, формировался с XIII века по XVII век в архиве рижского магистрата. До начала 60-х годов XX века документы данного комплекса хранились в Рижском городском государственном архиве, а ныне находятся в Латвийском государственном историческом архиве – структурном подразделении новообразованного Латвийского национального архива⁷⁸.

Собрание документальных памятников «Moscovitica–Ruthenica» представляет собой комплекс исторических источников об отношениях Риги, Ливонии, Ганзы и отдельных немецких городов с древнерусскими землями, княжествами и городами, Московским государством, а также с Великим княжеством Литовским и Речью Посполитой с конца XII века и до начала XVIII века. Состав документов «Moscovitica–Ruthenica» обусловлен функциями высшего органа управления Риги – магистрата: важной областью его деятельности было поддержание экономических и политических связей с русскими землями, посредничество в торговле Руси с Ганзой. Как результат документирования этой функции магистрата и возник комплекс «Moscovitica–Ruthenica». В результате в комплексе «Moscovitica–Ruthenica» представлена большая часть видов и разновидностей актового и делопроизводственного материала, характерного как для Руси, так и для регионов, входивших в сферу немецкого политического и культурного влияния. Объем документального массива значителен:

⁷⁷ Название комплекса впервые было зафиксировано в хронологических реестрах «Ruthenica» и «Moscovitica» Внешнего архива рижского магистрата: «Ruthenica», capsula R. IV¹⁰, 1228–1615. Латвийский национальный архив – Латвийский государственный исторический архив (Latvijas Nacionālā arhīva Latvijas Valsts vēstures arhīvs; далее – ЛНА–ЛГИА). Ф. 673. Оп. 1. Д. 1482; «Moscovitica», capsula M III¹⁰, 1558–1663. ЛНА–ЛГИА. Ф. 673. Оп. 1. Д. 1483.

⁷⁸ Об истории формирования комплекса «Moscovitica–Ruthenica», составе материалов, публикациях документов см. подробнее: Иванов А. С. «Moscovitica–Ruthenica» в Латвийском государственном историческом архиве: история формирования комплекса, состав и введение в научный оборот. *Древняя Русь: Вопросы медиевистики*. 2004. № 3 (17). С. 47–54; № 4 (18). С. 94–106; Ivanovs, A. Kompleksa «Moscovitica–Ruthenica» ieviešana zinātnes aprītē: arheogrāfisks apskats.” *Latvijas Arhīvi*. 2004. Nr. 2. 47.–85. lpp.; Ivanovs, A. Vēstures avotu kompleksa rekonstrukcijas problēma...; Ivanovs, A., Varfolomejev, A. Editing and Exploratory Analysis of Medieval Documents by Means of XML Technologies...; Ivanovs, A. Zespół dokumentów «Moscovitica–Ruthenica» w Łotewskim Państwowym Archiwum Historycznym. *Zapiski Historyczne*. 2006. Tom LXXI. Zeszyt 2–3. S. 125–143.

787 документов в фонде Внешнего архива рижского магистрата⁷⁹, около 80 документов в фонде Внутреннего архива рижского магистрата⁸⁰, а также отдельные документы в фонде коллегии ландратов Лифляндии⁸¹. Документы комплекса написаны на древнерусском, немецком, латинском, польском и шведском языках.

Выбор комплекса «Moscovitica–Ruthenica» для апробации возможностей технологий компьютерного источниковедения обусловлен, во-первых, его исторической ценностью, обширностью и уникальностью, а также разнообразием видового состава источников. При этом учитывалось, что комплекс «Moscovitica–Ruthenica» как единое целое лишь в редких случаях становился объектом специального, системного изучения⁸². По-прежнему актуальной остается задача введения источников комплекса в активный научный оборот: больше половины данной совокупности документов не были опубликованы и не использовались в исторических исследованиях⁸³. Только отдельные документы комплекса, преимущественно – древнейшие, издавались неоднократно, благодаря чему активно используются в научных исследованиях. Именно поэтому авторы настоящей монографии разработали концепцию системного применения компьютерных технологий для введения в научный оборот комплекса «Moscovitica–Ruthenica»⁸⁴.

⁷⁹ ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18–20.

⁸⁰ ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 14–18, 41, 72; Capsula B. № 42; Capsula C. № 1–11, 23, 27, 34, 43; ЛНА-ЛГИА. Ф. 8. Оп. 4. № 6–58.

⁸¹ ЛНА-ЛГИА. Ф. 214. Оп. 6. Д. 114, 115, 116.

⁸² За исключением упомянутых выше работ авторов данной монографии, комплекс «Moscovitica–Ruthenica» характеризовался лишь в общих чертах в отдельных статьях. См.: Енш Г. «Polonica» в Рижском городском государственном архиве. В кн.: *Ученые записки Института славяноведения АН СССР*. Т. II. Москва; Ленинград, 1950. С. 311–325; Дзярновіч А. Гісторыя архіўнага комплексу Moscovitica–Ruthenica Рыжскага магистрата. В кн.: *Віцебска-Рыжскія акты XIII–XVII ст.* Вып. I. Мінск, 2005. С. 10–19.

⁸³ Некоторые документы до сих пор не датированы и не описаны. См., например: ЛНА-ЛГИА. Ф. 8. Оп. 4. № 6–58.

⁸⁴ Иванов А. С., Варфоломеев А. Г. Использование технологии XML для введения в научный оборот комплекса документов Moscovitica–Ruthenica. В кн.: *RCDL'2004. Электронные библиотеки: перспективные методы и технологии, электронные коллекции*. Пуццино, 2004. С. 285–289; Варфоломеев, А. Г., Иванов, А. С. Технология XML: Современная реализация источник-ориентированного подхода...; Иванов А. С. Проблемы введения в научный оборот обширных комплексов источников: коллекция «Moscovitica–Ruthenica» в Латвийском государственном историческом архиве. В кн.: *XXI век: Актуальные проблемы исторической науки*. Минск, 2004. С. 100–101; Иванов А. С., Варфоломеев А. Г. Проект издания комплекса документов «Moscovitica-Ruthenica». В кн.: *Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам*. Ижевск, 2006. С. 65–70. Несмотря на то, что концептуально вопросы введения обширных комплексов исторических источников в научный оборот разработаны достаточно детально, проект по сплошной репрезентации документов «Moscovitica–Ruthenica» все еще не реализован.

Во-вторых, комплекс «Moscowitica–Ruthenica» сам по себе является идеальным объектом реконструкции, поскольку исторические источники, входящие в него, сохранились весьма полно. При этом хронологические, тематические и видовые границы комплекса могут быть установлены достаточно определенно и четко. Надежность реконструкции определяется наличием многообразных источников, содержащих информацию о складывании и эволюции комплекса, о его первоначальном составе и дальнейшей судьбе его документов. Пожалуй, важнейшим источником для реконструкции являются старые описи Внешнего архива рижского магистрата, а также описи Внутреннего («секретного») архива рижского магистрата⁸⁵. Описи позволяют восстановить первоначальную структуру комплекса «Moscowitica–Ruthenica», изучить особенности систематизации документов в средневековом архиве. В качестве дополнительных источников для реконструкции служат сами документы, входящие в комплекс. Изучение их текстов, делопроизводственных и архивных помет, палеографических особенностей дает возможность прояснить характер системных и делопроизводственных связей внутри комплекса. Таким образом, именно функциональный подход становится наиболее плодотворным и многообещающим при практическом осуществлении реконструкции.

В-третьих, наличие системных, закономерных и четких связей между документами комплекса «Moscowitica–Ruthenica» позволяет их всех включить в единую источникориентированную базу данных, равно как и в семантическую сеть, и подвергнуть многоаспектному источниковедческому изучению. Причем результаты такого изучения (в перспективе осуществления обширного исследования) должны обладать высокой степенью достоверности, поскольку выводы будут строиться на основе изучения не отдельных, вырванных из контекста эпохи и своего бытования документов, а спонтанно сложившейся их совокупности⁸⁶.

Для апробации конкретных технологий и методов компьютерного источниковедения в монографии использовались также две отдельные выборки тесно взаимосвязанных документальных источников, которые образуют субкомплексы в рамках комплекса «Moscowitica–Ruthenica».

⁸⁵ См., например: ЛНА-ЛГИА. Ф. 8. Оп. 1. Д. 59 («Register der Privilegien der Stadt Riga auf Befehl des Rahts im J. 1507 angefertigt»).

⁸⁶ Иванов А. С., Варфоломеев А. Г. О перспективах и пределах применения методов компьютерного источниковедения в медиевистике (на примере комплексного изучения коллекции «Moscowitica–Ruthenica»). *Древняя Русь: Вопросы медиевистики*. 2005. № 3 (21). С. 34–35.

Первый субкомплекс представлен смоленско-рижскими актами XIII – первой половины XIV века⁸⁷. Он включает в себя так называемый «договор неизвестного смоленского князя» или список *K* (в сущности, это условия к договору Смоленска с Ригой и Готландом)⁸⁸, списки *A*⁸⁹, *B*⁹⁰, *C*⁹¹, *D*⁹² и *E*⁹³ смоленского договора с Ригой, Готландом и немецкими городами 1229 года, грамоты смоленских князей Федора Ростиславича⁹⁴, Александра Глебовича⁹⁵ и Ивана Александровича⁹⁶, которые исследователи обычно рассматривают в качестве документов, подтверждающих договор 1229 года, а также грамоту смоленского князя Федора Ростиславича по судному делу о немецком «колоколе» (т.е. гире) 1284 года⁹⁷ и грамоту архиепископа рижского смоленскому князю Федору Ростиславичу с обвинением жителей города Витебска в необоснованной жалобе на рижан⁹⁸ (см. рис. 2.1, 2.2, 2.3, 2.4).

⁸⁷ Эти исторические источники издавались неоднократно. Недавно вышла в свет их новая дипломатическая публикация, в которой были пересмотрена датировка и атрибуция ряда грамот. См.: Иванов А., Кузнецов А. *Смоленско-рижские акты: XIII в. – первая половина XIV в.: Документы комплекса Moscovitica–Ruthenica об отношениях Смоленска и Риги*. [Rīga], 2009.

⁸⁸ [1219–1228/9 гг.] [Смоленск]. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 16. Л. 1.

⁸⁹ [1229 г.] [Готланд]. Подлинник, черновик. Пергамен. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 14. Л. 1.

⁹⁰ [1229 г. или немного позже]. [Смоленск]. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 15. Л. 1.

⁹¹ [Первая половина XIV в.] [Рига?]. Список. Пергамен. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 1. Л. 1.

⁹² [1229 г.] [Смоленск]. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 17. Л. 1.

⁹³ [1229 г.] [Смоленск]. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 2. Л. 3.

⁹⁴ [18 мая] 1284 г. Смоленск. Грамота о предоставлении чистого пути немецким и русским купцам, написанная от имени смоленского князя Федора Ростиславича архиепископу рижскому Иоганну I, магистру Ливонского ордена Виллекину и ратманам города Риги. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 72. Л. 1.

⁹⁵ [1297–1313 гг.] Смоленск. Грамота смоленского князя Александра Глебовича рижским ратманам и рижанам с подтверждением прежних дружественных отношений между Смоленском и Ригой. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 12. Л. 3.

⁹⁶ [Не ранее 1 ноября 1338 г. – не позднее 1341 г.] Смоленск. Договорная грамота смоленского князя Ивана Александровича с магистром Ливонского ордена, архиепископом рижским, рижскими ратманами и рижанами о подтверждении прежних договоров между Смоленском и Ригой. Подлинник. Бумага. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 18. Л. 1.

⁹⁷ Смоленск. Подлинник. Пергамен. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 6. Л. 3.

⁹⁸ [1284–1297 гг., очевидно, ближе к 1285–1287 гг.] Рига. Современный список (концепт?). Пергамен. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3.

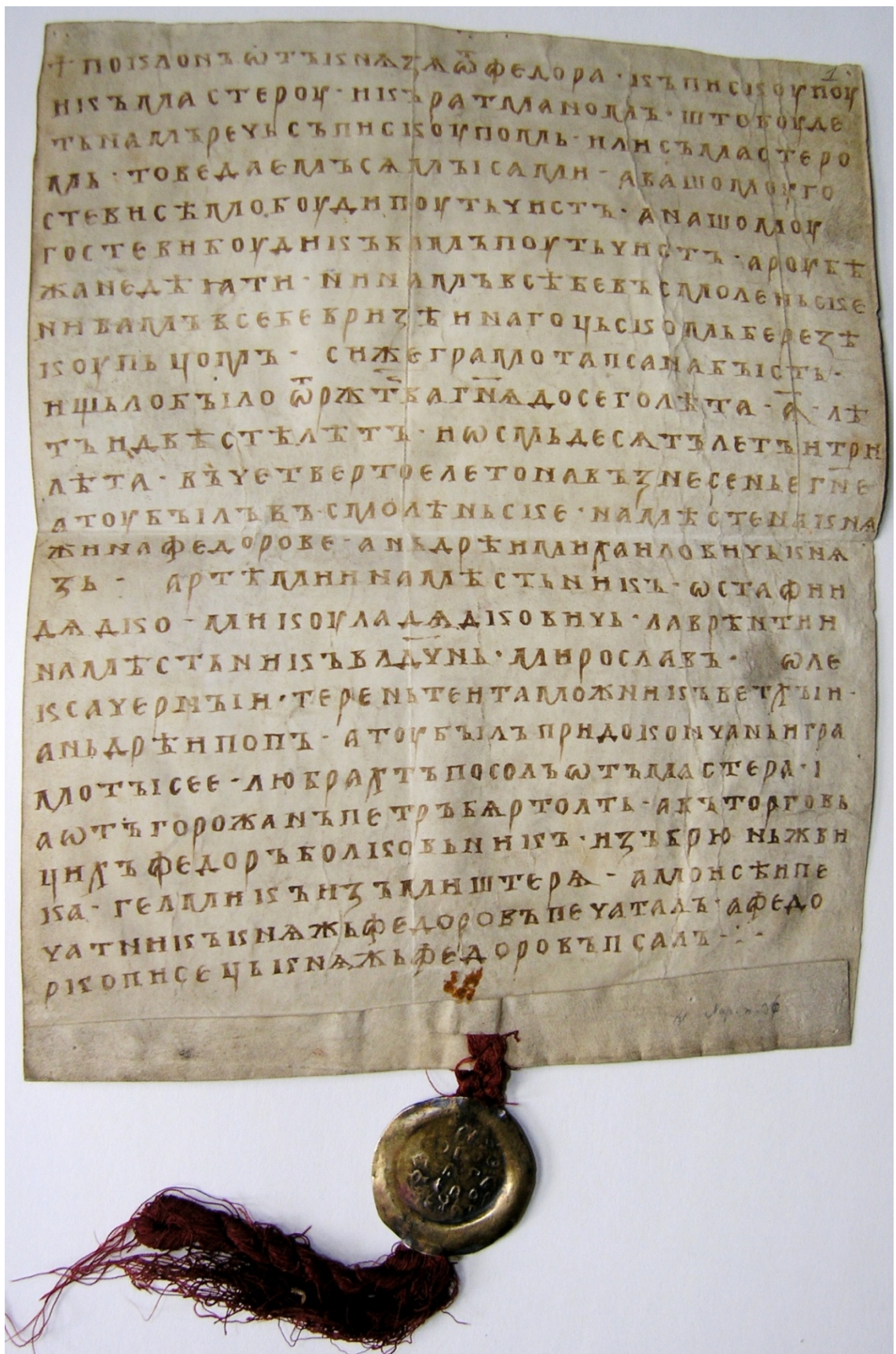


Рис. 2.2. 1284 г. [мая 18]. Грамота о предоставлении чистого пути немецким и русским купцам, написанная от имени смоленского князя Федора Ростиславича. ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 72. Л. 1 (акт № 4)

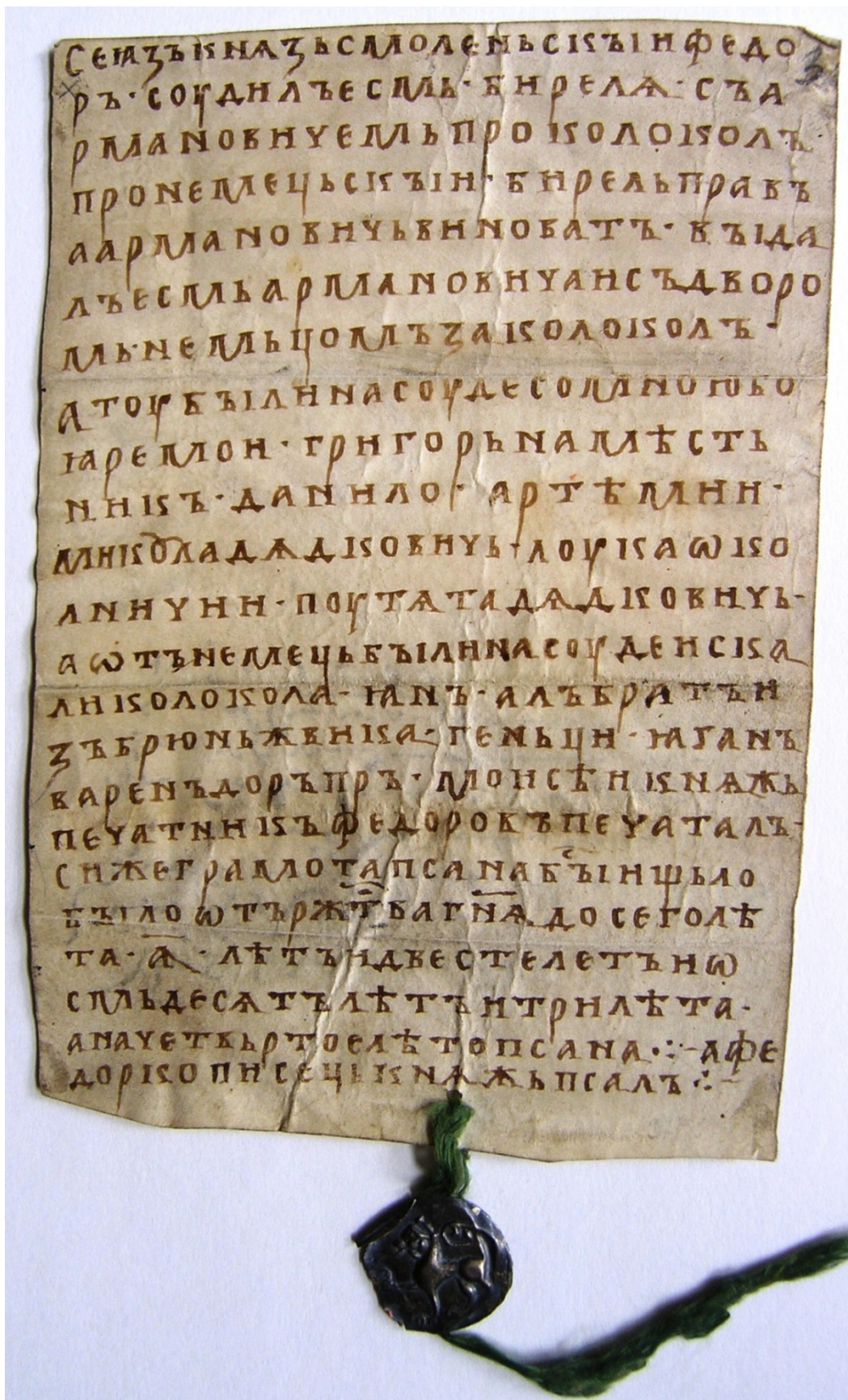


Рис. 2.3. 1284 г. Грамота смоленского князя Федора Ростиславича по судному делу о немецком «колоколе».
ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 6. Л. 3 (акт № 5)

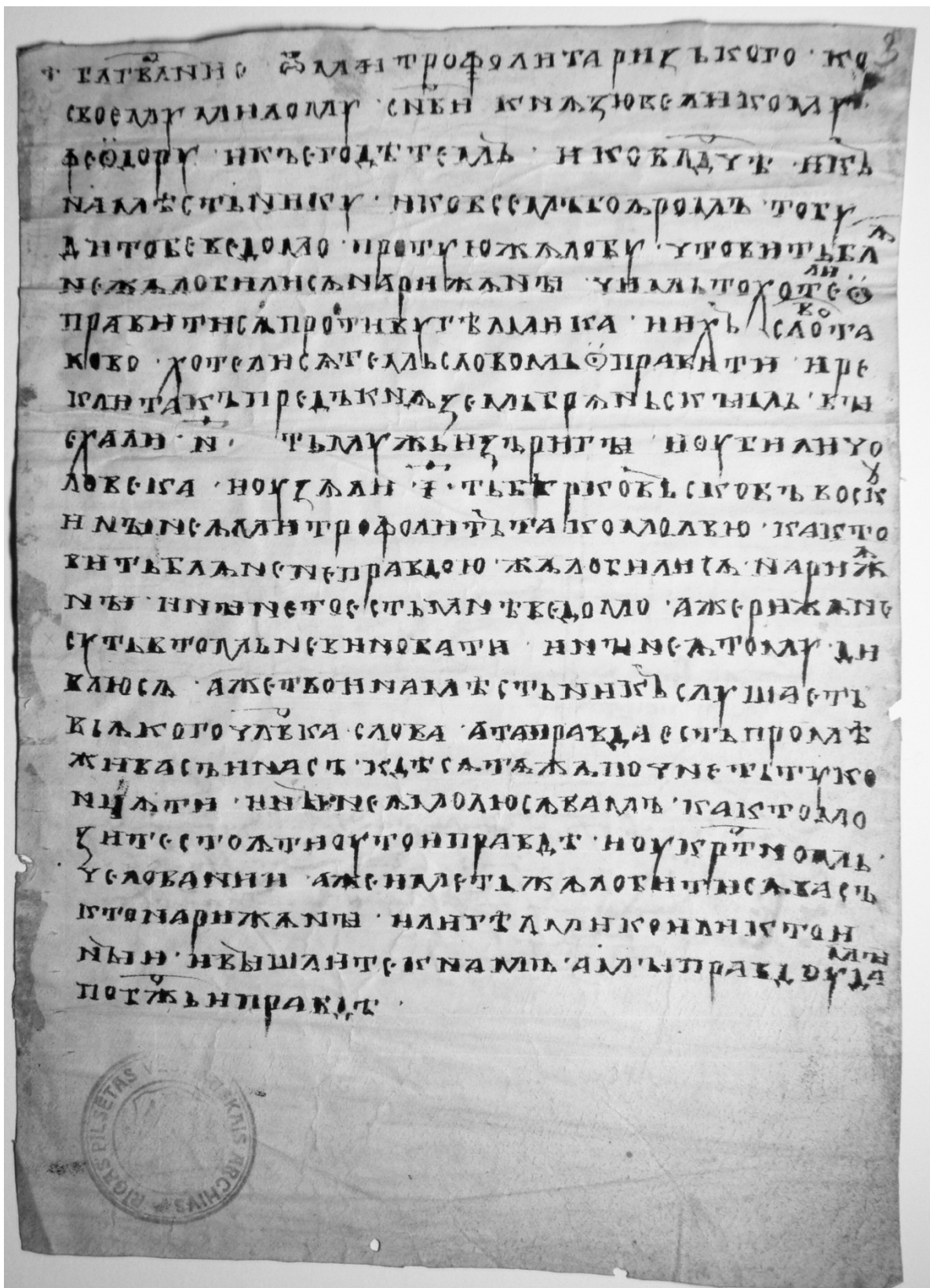


Рис. 2.4. [1284–1297 гг.] Грамота архиепископа рижского Иоганна II смоленскому князю Федору Ростиславичу.
ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3 (акт № 6)

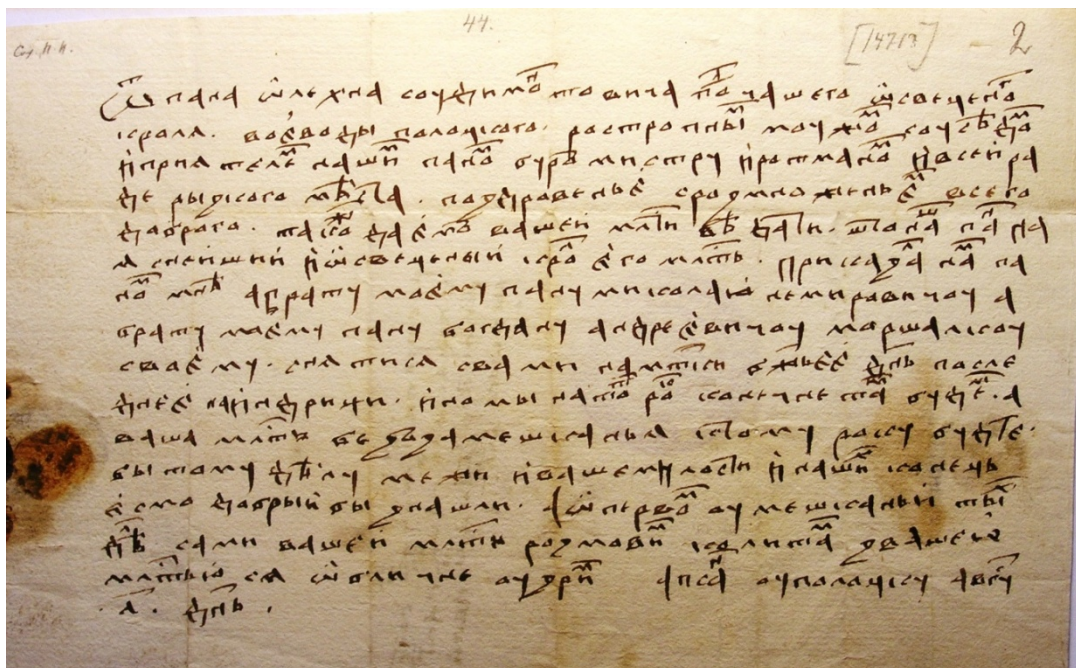


Рис. 2.5. [1471 г.] августа 30. Грамота полоцкого воеводы Олехны Судимонтовича рижскому магистрату.
ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 7. Л. 2

Смоленско-рижские акты действительно образуют органичную систему взаимосвязанных документов, возникшую в ходе подготовки, заключения и последующих действий по подтверждению и возобновлению договора 1229 года. При этом между списками договора (A, B, C, D и E) прослеживаются весьма определенные текстуальные связи, что позволяет установить генеалогию этих списков и гипотетически реконструировать историю их возникновения именно в процессе выработки заключительного текста договора 1229 года. Другие грамоты также связаны с этим договором и, соответственно, со списками, в которых запечатлен его текст. Так, «договор неизвестного смоленского князя» послужил основой для выработки положений договора 1229 года, а в последующих подтвердительных грамотах смоленских князей содержатся как косвенные, так и прямые отсылки к статьям базового документа, на основе которого вплоть до середины XIV века строились отношения Риги и других немецких городов со Смоленском. Связи между документами субкомплекса обусловлены не только процессом документирования отношений между Смоленском и Ригой и системой документооборота своего времени, но и отражением в документах «пересекающихся» аспектов исторической реальности, наличием параллельных или альтернативных свидетельств об определенных исторических фактах, явлениях и лицах. Многообразие и многоплановость связей между грамотами делают данный субкомплекс исторических источников незаменимым для проверки возможностей

технологий семантических сетей по связыванию информации источников, а также для апробации других компьютерных технологий и методов источниковедческой критики⁹⁹.

Для изучения потенциальных возможностей технологии XML в источниковедческом анализе актового материала использован еще один субкомплекс источников, входящий в комплекс «Moscovitica–Ruthenica»¹⁰⁰. Данный субкомплекс образуют лишь несколько полоцких грамот 60–70-х годов XV века¹⁰¹: грамота полоцкого городского совета рижскому магистрату о передаче задержанных в Полоцке рижских товаров слуге рижского магистрата Ольбрахту¹⁰², удостоверительная грамота полоцкого городского совета о выдаче Ольбрахту рижских товаров, задержанных в Полоцке¹⁰³, грамота полоцкого воеводы Олехны Судимонтовича рижскому магистрату с просьбой разрешить торговый конфликт¹⁰⁴, грамота полоцкого воеводы Олехны Судимонтовича рижскому магистрату о причине переноса съезда литовских и рижских послов¹⁰⁵ и грамота полоцкого воеводы Олехны Судимонтовича рижскому магистрату о готовности литовских послов начать переговоры с ливонскими послами¹⁰⁶ (рис. 2.5). Грамоты близки по времени создания и к тому же достаточно репрезентативно отражают особенности (в том числе – формуляр) данной разновидности документов, посланных из Полоцка в Ригу¹⁰⁷.

⁹⁹ См. главу 4-ю настоящей монографии. См. также: Ivanovs, A., Varfolomeyev, A. Semantic Publications of Charter Corpora (The Case of a Diplomatic Edition of the Complex of Old Russian Charters «Moscovitica–Ruthenica»). In: *International Conference «Digital Diplomats: Tools for the Digital Diplomatist»*. [Naples], 2011. Pp. 36–40; Варфоломеев А. Г., Иванов А. С. Модели структуры и содержания исторических источников. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 37. Петрозаводск, 2011. С. 25–31; Varfolomeyev, A., Ivanovs, A. Wiki Technologies for Semantic Publication of Old Russian Charters. In: *Digital Humanities 2012*. Hamburg, 2012. Pp. 405–407; Иванов А. С., Варфоломеев А. Г. Репрезентация содержания исторических источников в семантических сетях: Attempto Controlled English (ACE). В кн.: *Информационные технологии и письменное наследие. El'Manuscript–2012*. Петрозаводск; Ижевск, 2012. С. 112–120.

¹⁰⁰ См., например: Варфоломеев А., Иванов А. XML-технологии в анализе информации полнотекстовых баз данных. В кн.: *Vēsture: Avoti un cilvēki. Starptautiskās konferences «XV Zinātniskie lasījumi» materiāli. Vēsture IX*. Daugavpils, 2006. 164.–174. lpp.; Ivanovs, A., Varfolomeyev, A. Editing and Exploratory Analysis of Medieval Documents by Means of XML Technologies...

¹⁰¹ Данные грамоты уже публиковались, однако воспроизведение текстов не было идеальным. См.: Хорошкевич А. Л., сост. *Полоцкие грамоты XIII–начала XVI вв.* I вып. Москва, 1977. № 119, с. 221–222; II вып. Москва, 1978. № 145, с. 35–36; № 146, с. 36–37; № 147, с. 37–39; № 148, с. 39–40.

¹⁰² [1471 г. после марта 30]. Полоцк. Подлинник. Бумага. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 2. Л. 1.

¹⁰³ [1471 г. после марта 30]. Полоцк. Подлинник. Бумага. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 8. Л. 2.

¹⁰⁴ [1464–1465 гг.] ноября 17. Полоцк. Подлинник. Бумага. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 10. Л. 1.

¹⁰⁵ [1471 г.] июля 3. Полоцк. Подлинник. Бумага. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 5. Л. 2.

¹⁰⁶ [1471 г.] августа 30. Полоцк. Подлинник. Бумага. ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 7. Л. 2.

¹⁰⁷ Разумеется, выборка подобных документов могла быть и более обширной, так как в комплексе «Moscovitica–Ruthenica» полоцкие грамоты представлены особенно широко. Однако,

Комплекс источников по истории Динабурга середины XVII века

Второй обширный комплекс источников, материалы из которого используются преимущественно в третьей главе монографии, – это совокупность документов приказного делопроизводства, в которых всесторонне отразились различные аспекты истории ливонского города Динабурга, который в 1656–1667 годах находился под властью царя Алексея Михайловича¹⁰⁸. Вскоре после взятия русскими войсками город, в полном соответствии с традициями того времени, был переименован и в честь особо почитаемых в России святых страстотерпцев Бориса и Глеба стал называться Борисоглебским городом, а в некоторых документах того времени – также Борисоглебовым городом¹⁰⁹.

Данный комплекс исторических источников отложился в фонде Разрядного приказа Российского государственного архива древних актов¹¹⁰, что не было случайным: город Динабург в течение всего периода русской власти находился на окраинных землях Русского государства, являлся важным форпостом в продвижении России к Балтийскому морю, часто подвергался набегам польских войск, в городе постоянно находился гарнизон русских служилых людей. Соответственно, воеводы города были вынуждены выполнять прежде всего чисто военные функции и лишь во вторую очередь – собственно административные. Поэтому закономерно, что не только назначение и смена воевод, но и порой само оперативное управление городом находилось в общем ведении Разрядного приказа, что полностью отвечало его

по мнению авторов монографии, использованных грамот вполне достаточно для апробации технологии XML при проведении формулярного анализа.

¹⁰⁸ См.: Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов (1656–1666)*. Часть 1. Daugavpils, 2002. С. 21–60; Иванов А. С., Кузнецов А. М. Динабург в документах времен царя Алексея Михайловича. В кн.: *Динабург, Двинск, Даугавпилс в истории, культуре, литературе*. Даугавпилс, 1999. С. 9–46; Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения (на примере формулярного анализа документов приказного делопроизводства)...; Варфоломеев А. Г., Иванов А. С. Технология XML как инструмент компьютерного источниковедения: (на примере изучения комплекса средневековых документов по истории Динабурга). В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 32, апрель 2004. Москва; Томск, 2004. С. 137–139.

¹⁰⁹ См. подробнее: Кузнецов А. М. Почему Даугавпилс назывался Борисоглебовым. *Русская речь*. 1994. № 1. С. 83–86; Иванов А. С., Кузнецов А. М. Динабург в начальный период Первой Северной войны (1655–1656). В кн.: *Славянские чтения*. II. Даугавпилс; Резекне, 2002. С. 222–236.

¹¹⁰ Российский государственный архив древних актов (далее – РГАДА). Ф. 210. Московский стол. Столб. 254, 270, 272, 279, 285, 289, 306, 322, 342(2), 344, 349, 352(1), 364, 365; Севский стол. Столб. 175; Белгородский стол. Столб. 394, 400, 429, 518, 542; Новгородский стол. Столб. 126, 136, 148, 162, 187, 188, 257, 276, 297, 299.

функциям¹¹¹. Следует отметить, что большая часть документов комплекса опубликована¹¹², в том числе и в издании, подготовленном одним из авторов данной монографии¹¹³.

В комплексе представлены практически все разновидности приказной документации, характеризующие процесс воеводского управления небольшим городом: указы и наказы воеводам, памяти, отписки воевод, росписи служилых людей, годовые сметы, докладные записи в Разряде, челобитные, расспросные речи. Некоторая нетипичность данного комплекса источников связана не с видовым составом документов, а с их содержанием, особенностью отражения исторических реалий: как уже отмечалось, Динабург был прифронтовым городом и его гарнизон терпел многие лишения и нужду, поскольку прилегающие к Динабургу волости, с которых «ратные люди живилис[ь], розорено без остатку»¹¹⁴. Центральные власти, воеводы Новгорода и Пскова, которым непосредственно подчинялся динабургский воевода, не могли решить проблему снабжения кормами и припасами, поэтому для русских служилых людей пребывание в Динабурге было настоящим осадным сиденьем. Именно эти черты жизни осажденного города наиболее интенсивно представлены в данном комплексе документов, обусловив специфику их прямой и косвенной информации.

Выбор документов из данного комплекса для апробации компьютерных технологий в историческом источниковедении обусловлен рядом причин. Во-первых, это высокая степень репрезентативности видового состава комплекса: входящие в него исторические источники позволяют в перспективе разработать всестороннюю методику изучения приказного делопроизводства XVII века, включая как специальные подходы к анализу отдельных разновидностей приказных документов, так и общие методы критики приказной документации. Во-вторых, документы по истории Динабурга 1656–1667 годов (как и актовый материал комплекса «*Moscowitica–Ruthenica*») представляют собой полноценный комплекс взаимосвя-

¹¹¹ Чернов А. В. *Вооруженные силы Русского государства в XV–XVII в.в.* Москва, 1954. С. 182; Ерошкин Н. П. *История государственных учреждений дореволюционной России.* Москва, 1968 С. 39–40.

¹¹² См.: *Акты Московского государства, изданные Императорской Академией Наук.* Том II: *Разрядный приказ. Московский стол. 1635–1659* / Под ред. Н. А. Попова. СПб., 1894. № 647, 851, 852, 855, 862–864, 866–868, 870, 874, 879, 881, 883, 912, 928, 932, 937, 966; *Акты Московского государства, изданные Императорской Академией Наук.* Том III: *Разрядный приказ. Московский стол. 1660–1664* / Под ред. Д. Я. Самоковасова. СПб., 1901. № 59, 382, 440, 453, 456, 565, 571, 580, 582, 647, 664, 673, 681, 690.

¹¹³ Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов (1656–1666).* Часть 1 и 2. Daugavpils, 2002.

¹¹⁴ РГАДА. Ф. 210. Новгородский стол. Столб. 148. Л. 156.

занных источников¹¹⁵ – связи эти, прежде всего, делопроизводственные: отписки воевод, как правило, являются ответами на указы, в свою очередь, указные грамоты составлены с учетом поступивших в Разряд отписок и т.д., а утраченные документы частично воспроизводятся (цитируются) в сохранившихся.

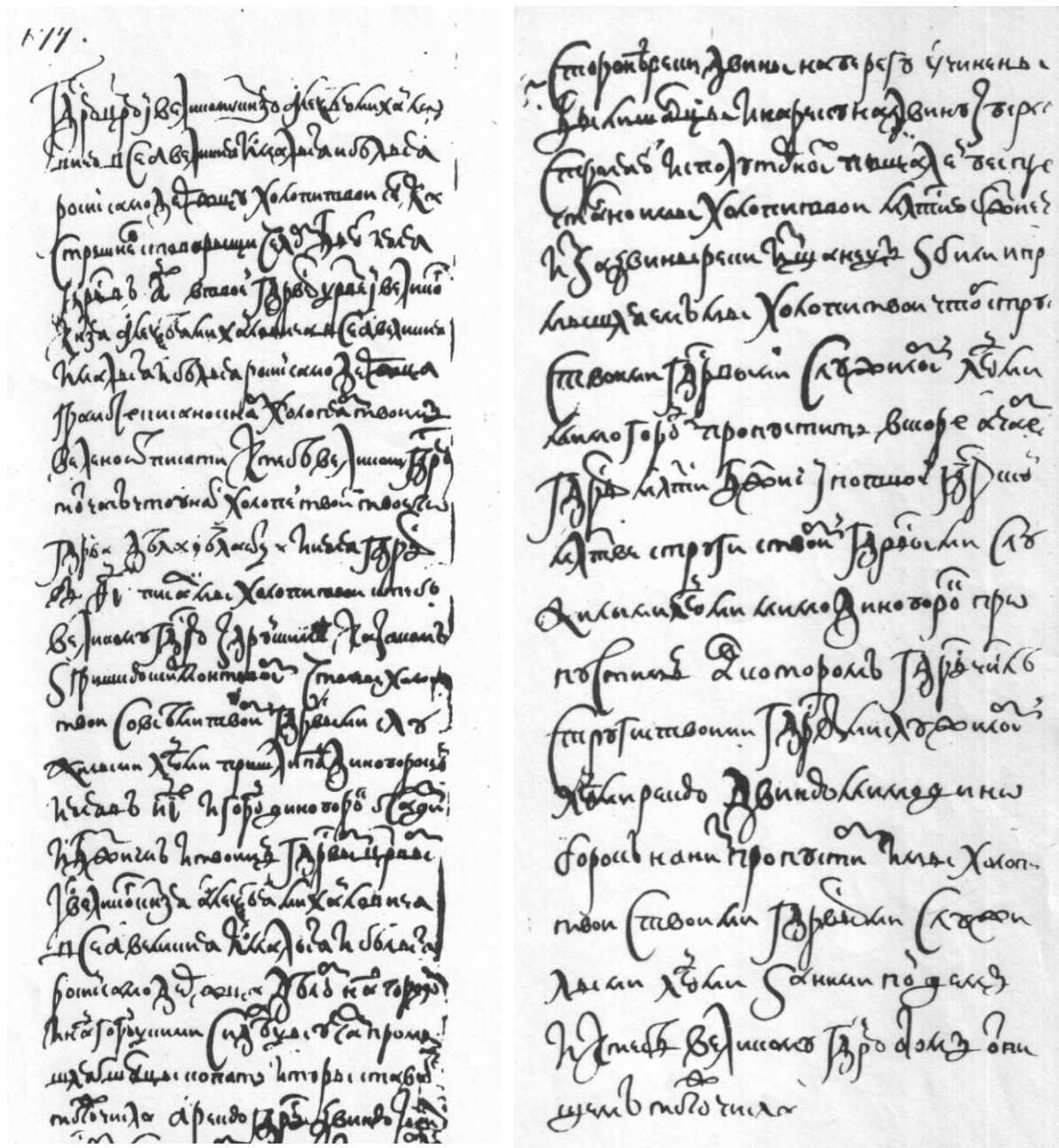


Рис. 2.6. [1656 г. не ранее июля 19]. Отписка боярина и воеводы С. Л. Стрешнева царю Алексею Михайловичу об осаде города Динабурга. РГАДА. Ф. 210. Московский стол. Столб. 270. Л. 674–675 (документ № 1)

¹¹⁵ Иванов А. К вопросу о повышении информативной отдачи документов приказного делопроизводства...; Иванов А. С., Кузнецов А. М. Комплексный подход к изучению документов приказного делопроизводства... и др.

Комплексный подход и, соответственно, применение инструментария компьютерного источниковедения способны не только повысить эффективность, но и углубить источниковедческое изучение документов приказного делопроизводства. В течение долгого времени специальное изучение этих источников идет обычно в русле источниковедения XIX – первой половины XX века. В центре внимания специалистов – авторов учебных пособий по источниковедению и по истории и организации делопроизводства¹¹⁶, а также специальных исследований¹¹⁷ – находятся преимущественно вопросы аналитической критики источников. В результате рекомендации по использованию делопроизводственных документов в исторических исследованиях не отличаются разнообразием, зачастую ограничиваясь прямой информацией источников¹¹⁸. Технологии же компьютерного источниковедения нацелены, прежде всего, на выявление косвенной, скрытой информации. При этом конкретная методика источниковедческого исследования, отработанная на данном комплексе, без особых трудностей может быть перенесена на изучение других источников по истории России XVI–XVII веков, а возможно – и более позднего времени.

Задачи изучения комплексов исторических источников в компьютерном источниковедении

Комплексный подход, заявленный и частично обоснованный в историческом источниковедении, ставит лишь самые общие задачи изучения органических, спонтанно сложившихся совокупностей исторических источников. Однако особенности охарактеризованных выше комплексов источников – исторические условия и специфика формирования комплексов, исторический период, видовой состав источников, особенности внешней и внутренней формы, а также содержания, характер взаимосвязей между памятниками письменности, особенности отражения исторической реальности в текстах, направления использования источников в исследованиях и т.д. – позволяют конкретизировать общие задачи их всестороннего и системного изучения в компьютерном источниковедении.

¹¹⁶ См.: Митяев К. Г. *История и организация делопроизводства в СССР*. Москва, 1959. С. 32–41; *История делопроизводства в СССР*. Москва, 1974. С. 10–18 и др.

¹¹⁷ См., например: Тихомиров М. Н. Приказное делопроизводство в XVII веке. В кн.: Тихомиров М. Н. *Российское государство XV–XVII веков*. Москва, 1973. С. 348–383.

¹¹⁸ *Источниковедение: Теория. История. Метод...* С. 408; Демидова Н. Ф. *Служилая бюрократия в России XVII в. и ее роль в формировании абсолютизма*. Москва, 1987. С. 9–14.

Первой задачей является реконструкция комплексов исторических источников. Следует отметить, что реконструкция этих комплексов в общих чертах была произведена еще во время подготовки традиционных (печатных) публикаций без применения методов компьютерного источниковедения, что вполне закономерно и оправданно на стадии апробации инструментария компьютерного источниковедения, который в перспективе может быть использован для решения аналогичных задач. Фактически, авторы монографии применяли компьютерные технологии не для непосредственной реконструкции комплексов источников, а для проверки результатов уже проведенной работы по воссозданию исторических совокупностей источников. При этом верификацию результатов применения комплексного подхода в источниковедении можно считать самостоятельной и весьма значимой задачей компьютерного источниковедения. В ходе проведения верификации было установлено, что работу по реконструкции обоих комплексов исторических источников нельзя считать завершенной.

Так, субкомплекс смоленско-рижских актов, соответственно, и весь комплекс «*Moscowitica–Ruthenica*» может быть расширен за счет включения в него других ливонских актов, в которых упоминаются те же исторические события и лица, что и в смоленско-рижских актах¹¹⁹. Не исключена также возможность выявления дополнительных документов в других фондах и коллекциях Латвийского государственного исторического архива¹²⁰. Необходимо также учитывать, что некоторые документы комплекса «*Moscowitica–Ruthenica*» в свое время отложились в архивохранилищах в Дерпте, Кенигсберге и других ганзейских городах¹²¹. О наличии системных связей между документами из различных архивохранилищ свидетельствуют сами источники комплекса «*Moscowitica–Ruthenica*» – переписка Рижского магистрата, отпуски и концепты документов. В связи с этим возникает вопрос о допустимых пределах расширения исторического комплекса источников в ходе его реконструкции. Документы, возникшие в результате

¹¹⁹ «Внешние» связи смоленско-рижских актов с документами, не принадлежащими комплексу «*Moscowitica–Ruthenica*», отмечены в модели семантической сети, охарактеризованной в четвертой главе монографии.

¹²⁰ См., подробнее: *Центральный государственный архив Латвийской ССР: Краткий справочник*. Часть I: (1220–1918). Рига, 1980. С.66–68; Bes, L., Frankot, E., Brand, H., eds. *Baltic Connections: Archival Guide to the Maritime Relations of the Countries around the Baltic Sea (Including the Netherlands), 1450–1800*. Vol. II: *Latvia, Lithuania, the Netherlands*. Leiden; Boston, 2007. Pp. 787–873.

¹²¹ См., например: Benninghoven, F., hrsg. *Liv-, est- un kurländische Urkundenregesten bis zum Jahre 1300*. Hamburg, 1959.

деятельности различных учреждений и отложившиеся в различных архивах, безусловно, образуют единую систему, границы которой заданы системой отношений Рижского магистрата со своими корреспондентами, другими органами управления в Ливонии, с немецкими (ганзейскими) городами, с восточными соседями. Однако подобное расширение комплекса может оказаться бесконечным. На первый взгляд, целесообразно ограничить реконструкцию рамками, заданными в делопроизводственной и архивной практике в период становления и эволюции комплекса. Однако данный тезис оправдан лишь в контексте традиционного исторического источниковедения. Компьютерное источниковедение предлагает иной подход к установлению пределов расширения комплексов источников: возможность репрезентации (и последующего изучения) количественно неограниченных совокупностей источников в сети Интернет позволяет сделать процесс реконструкции комплексов «открытым» – стремящимся к полноте и завершенности, но никогда не завершенным.

Особо важная задача, решение которой будет способствовать полноценному источниковедческому изучению как документальных комплексов, так и отдельных исторических источников, – это подготовка полноценных их публикаций как в печатной, так и в электронной форме. Причем подготовка электронных публикаций, с точки зрения компьютерного источниковедения, является приоритетной, так как только высококачественная репрезентация источников в электронной форме позволяет эффективно использовать компьютерные технологии и методы. Разумеется, должна быть выбрана наиболее адекватная технология для подготовки таких публикаций.

Зачастую современные электронные публикации просто копируют «бумажные», достигая при этом некоторых естественных преимуществ – улучшения качества передачи изображений источников и снятия ограничений на объем публикаций. Но ориентироваться в бумажном тексте проще, чем в его электронном варианте, 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 памятников письменности. Такие публикации должны не только вводить в научный оборот новые источники и обеспечивать их доступность в сети Интернет, но и предоставлять разнообразные возможности их источниковедческой критики.

Наиболее перспективным и многообещающим видом публикаций комплексов источников, бесспорно, является их семантическая публикация, которая целиком и полностью отвечает идеальной модели высококачественной публикации исторических источников. Так, в научной публикации тексты исторических источников и научный аппарат издания («конвой»¹²²) неизбежно связаны между собой множеством отношений: например, альтернативные или параллельные свидетельства ряда источников, относящихся к одним и тем же аспектам исторической реальности, прямые и косвенные ссылки в текстах источников на другие источники, связи текстов документов с научными комментариями, указателями, иллюстрациями и т.п. Такие связи легко представить в виде стрелок разного типа, соединяющих некие объекты, то есть в виде «семантической сети»¹²³. Это означает, что любая научная публикация источника, в сущности, включает его в некую семантическую сеть. Такая сеть должна стать по возможности более детальной. Так, в качестве объектов сети должны фигурировать не только тексты и изображения источников, 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. При этом нельзя «потерять» сам текст как целое¹²⁴.

В некоторых случаях электронная публикация источников должна быть не только критической или дипломатической, но и «палеографической», т.е. максимально точно отражать палеографические особенности документов, включая особенности начерка букв и других символов. Данное требование относится, прежде всего, к грамотам на древнерусском языке, составляющим субкомплекс смоленских грамот в рамках комплекса «Moscovitica–Ruthenica», поскольку палеографические особенности важны как для специалистов по истории древнерусского языка, так и для историков и источниковедов, ведь именно палеографические особенности данных документов (большая часть которых не имеет даты) являются едва ли не решающим аргументом при обосновании их датировки и атрибуции. Присутствие в электронной публикации фрагментов изображений документов вплоть до отдельных символов, а также связи между симво-

¹²² Довольно удачный термин, предложенный В. В. Козловым. См.: Козлов В. П. *Основы теоретической и прикладной археографии...* С. 125.

¹²³ См. подробнее в четвертой главе монографии.

¹²⁴ Подходы к решению этой конкретной задачи на основе технологии XML рассмотрены в третьей главе.

лами в тексте и их изображениями позволяют снабдить публикацию эффективными 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.

Электронная публикация может включать в себя самые разные инструменты для решения конкретных источниковедческих задач, например, средства для проведения формулярного¹²⁵ и текстологического анализа актовых материалов и делопроизводственных источников. Данная задача актуальна в изучении как комплекса «*Moscowitica–Ruthenica*», так и приказных документов по истории Динабурга. Как показано в третьей главе монографии, наличие в составе публикации инструментов проведения такого анализа значительно расширяет возможности использования электронной публикации.

Принцип формирования программной «оболочки» электронной публикации в виде среды для проведения исследований с помощью встроенных инструментов¹²⁶ означает, что электронная публикация должна быть открытой семантической сетью, то есть сетью, в которую можно добавлять объекты и связи между ними. При этом электронные публикации не должны быть изолированы друг от друга; они должны создаваться 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 источников¹²⁷.

¹²⁵ Формулярный анализ является одним из способов изучения процесса складывания и эволюции структуры (внутренней формы) конкретных разновидностей документов, прежде всего – актов, что является одной из важнейших задач актового источниковедения. См.: Каштанов С. М. *Русская дипломатика*. Москва, 1988. С. 190–191.

¹²⁶ 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 исторических источников.

¹²⁷ Варфоломеев А. Г., Кравцов И. В. Приобретение и представление знаний в сетевом сообществе исследователей текстов...

Еще одной задачей источниковедческого изучения комплексов исторических источников можно считать «динамическое» воссоздание (реконструкцию) исторической реальности, взаимосвязи между аспектами которой отразились в источниках¹²⁸. Возможности для этого предоставляет сам комплекс источников, если он образует, например, систему документов, в которой, как правило, ни один аспект исторической реальности (исторический факт) не появляется изолированно от других. Напротив, он органично включается в систему, которая может быть рассмотрена в синхронном и диахронном разрезе. Так, в комплексе документальных источников по истории Динабурга входящий документ (отписка, например), фиксирующий какой-либо факт, одновременно предоставляет информацию о связи этого факта с другими аспектами реальности, с ситуацией в целом, отразившейся в ряде других документов комплекса, при этом тот же самый входящий документ инициирует составление другого документа (исходящего или внутреннего), где фиксируется дальнейшее развитие указанного факта или имеется свидетельство о появлении производных фактов. Обычно цепочка документирования исторического процесса на этом не прерывается. Представление комплекса источников в семантической публикации позволяет ясно вычлениить, а иногда – и воссоздать, связи между многообразными аспектами информации, зафиксированной в различных источниках. В конечном итоге, на основе констатации этих связей возникает принципиальная возможность создания достоверной динамической модели факта или более крупного фрагмента реальности¹²⁹. При этом качество и достоверность «динамического» воссоздания реальности во многом зависит от объема информации, то есть, в конечном итоге, – от количества и информационной насыщенности источников, включенных в публикацию. В ней должен быть представлен максимально полный (если не исчерпывающий) набор выделенных (размеченных) признаков и характеристик, встречающихся в документах, для многомерного связывания всех аспектов информации источников. Основное же внимание следует уделять описаниям и реалиям, поскольку формулы, штампы, устойчивые обороты не несут какую-либо существенную информацию о фактах, а

¹²⁸ Ковальченко И. Д. Исторический источник в свете учения об информации... С. 145.

¹²⁹ Разумеется, подобное моделирование будет описательным, а не собственно математическим, так как в условиях острого недостатка объективных (в том числе количественных) характеристик исторических явлений моделируются сами исторические реалии и ситуации. Ср.: Бородкин Л. И. Методология моделирования в зарубежной клиометрике: новые тенденции. В кн.: *Проблемы исторического познания*. Москва, 1999. С. 139.

лишь свидетельствуют об особенностях формуляра¹³⁰ данной разновидности делопроизводственной документации.

Динамическое моделирование фрагментов исторической реальности позволяет решить еще одну проблему, которая в общем виде заявлена в современном источниковедении. Речь идет о гипотетической, вероятностной реконструкции «утраченных» исторических фактов, которые прямо ни в одном из источников не отразились¹³¹, а поэтому восстанавливаются на основе скрытой информации источников. Путь к решению этой задачи связан с созданием гибкой вероятностной, «расширенной» модели фрагмента реальности. В структуре этой модели заполненные ниши (прямая и косвенная информация источников об аспектах реальности) перемежаются с незаполненными, которые, в конечном итоге, и восполняются путем логических умозаключений. В результате должна возникнуть непротиворечивая версия фрагмента реальности.

При системном изучении достаточно большой совокупности документальных источников, образующих комплекс, восполнение пробелов в информации отдельных источников облегчается промежуточной процедурой – гипотетическим воссозданием содержания утраченных документов в рамках естественного комплекса исторических источников. Решение этой задачи возможно, поскольку во многих источниках – древнерусских актах и документах приказного делопроизводства – имеются прямые и косвенные ссылки на другие документы. Следует отметить, что содержание упоминаемых документов достаточно полно и точно излагается именно в документах приказного делопроизводства XVII века. Так, если исходящие документы (например, указы, указы, указы) сохранились в виде отпусков, то при сравнении их описательной части со ссылками на тот же документ в отписках, составленных в ответ на указы, появляется возможность с достаточной степенью уверенности реконструировать окончательный (отосланный) вариант указа или наказа. Еще одну возможность реконструкции содержания утраченных документов предоставляет ситуационный анализ действий должностных лиц, которые руководствовались полученными распорядительными документами.

На этой основе с большой степенью достоверности можно реконструировать и утраченные факты. Как и при решении предыду-

¹³⁰ Формуляр – это внутренняя форма документа, структура его текста. См.: Каштанов С. М. *Русская дипломатика...* С. 26.

¹³¹ Прежде всего пробелы в источниках необходимо констатировать. См., например: Бобинская Ц. Пробелы в источниках. Методологический анализ. *Вопросы истории*. 1965. № 6. С. 76–86.

щей задачи, при реконструкции как самих документов, так и исторических фактов необходимо гибкое многофакторное изучение больших массивов документации, что требует применения адекватных компьютерных технологий. Решение данной задачи особенно актуально при изучении комплекса документов по истории Динабурга 1656–1667 годов, так как несмотря на в общем-то хорошую сохранность документов, в комплексе источников есть и пробелы¹³².

В источниковедческом изучении комплексов документальных источников большой интерес может представлять выявление особенностей отражения различных сторон «живой» реальности сквозь призму стандартизированного делового языка. Здесь можно использовать опыт лингвистического источниковедения, изучающего специфику деловой письменности, которая нивелирует реальность и обладалает своей стилистикой и лексикой¹³³. Так, в качестве основного метода изучения языка приказных документов применяется дифференцированный лексический анализ текстов¹³⁴, эффективность которого существенно повышается при использовании статистических методов. В итоге, интерпретация полученных результатов будет обладать большей степенью репрезентативности, надежности и достоверности. Гипотетически можно предложить следующую модель интерпретации результатов лексического анализа: любые отклонения от стандартных, устойчивых форм приказного (делового) языка будут нести наиболее существенную и более объективную информацию о ситуации в городе, так как само использование нестандартизированной, разговорной лексики говорит о «разрыве» между видением ситуации бюрократической системой и реальным положением дел. Чем больше этот разрыв, тем ярче раскрывается реальная ситуация. Иногда разрыв настолько значителен, что применение точных количественных методов даже не является обязательным. Например, при ухудшении военного положения в районе Динабурга в 60-е годы XVII века, при обострении проблем со снабжением, с усилением дезертирства из русского гарнизона язык отписок становится живым,

¹³² Наиболее полно представлены документы начального периода русской власти в городе (1656–1658 гг.), выборочно – за последующие годы. См.: Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов (1656–1666)*. Часть 1. С. 24–34.

¹³³ См., например: Кортова Т. В. *Московский приказный язык XVII века как особый тип письменного языка*. Москва, 1998. С. 20–30.

¹³⁴ Волков С. С. *Лексика русских челобитных XVII века: Формуляр, традиционные этикетные и стилевые средства*. Ленинград, 1974. С. 9–11.

образным, позволяющим прочувствовать и зримо представить ситуацию в городе, увидеть ее глазами воевод¹³⁵.

Приведенный выше список задач изучения комплексов исторических источников, в решении которых можно и нужно использовать инструментарий компьютерного источниковедения, не является исчерпывающим, поскольку задачи возникают именно в творческой деятельности источниковеда и историка. Предназначение же компьютерного источниковедения в том, чтобы предложить инструментарий для решения этих и любых других задач. К сожалению, этот инструментарий разработан явно недостаточно, поэтому и практическая реализация комплексного подхода в компьютерном источниковедении не всегда принимает конкретные формы, зачастую остается лишь декларацией. В то же время, алгоритм применения инструментария компьютерного источниковедения к решению данных задач выглядит достаточно определенным: (1) перевод текстов источников и метаинформации о них в электронную форму; (2) разметка текстов и метаинформации (а в идеальном варианте – репрезентация информации на одном из естественных контролируемых языков); (3) связывание информации в рамках традиционной источник-ориентированной базы данных или (лучше) формирование семантической сети; (4) применение аналитического инструментария компьютерного источниковедения в аналитической и синтетической критике источников; (5) в идеале, генерация новых знаний на основе автоматического вывода, но обязательно при творческом участии исследователя.

¹³⁵ См., например, характеристику ситуации в городе в 1665г. в отписке воеводы Б. Неклюдова: «... А те, г[о]с[у]д[а]рь, твои, великого г[о]с[у]д[а]ря, ратные люди борисоглебские осадные сидел[ь]цы н[ы]не дошло время имъ до бол[ь]шие скудости. А опасуюс[ь], г[о]с[у]д[а]рь, я, холоп твои, прежнего самовол[ь]ства от тех твоих, великого г[о]с[у]д[а]ря, ратных людеи ... Опасно, г[о]с[у]д[а]рь, мне, холопу твоему, чтоб твоего, великого г[о]с[у]д[а]ря, города и казны неприятел[ь]ским литовских полков ратемъ не вручили, не перетерпя нужного времени и н[ы]нешних холодных дни. А мне б, холопу твоему, за непостереган[ь]е от тебя, великого г[о]с[у]д[а]ря, в опале не быт[ь] и смерть, г[о]с[у]д[а]рь, принят[ь] мне, холопу твоему, по твоему, великого г[о]с[у]д[а]ря, указу вместо небеснаго дару. А страшит неприятел[ь]ская неволя ...» (РГАДА. Ф. 210. Новгородский стол. Столб.148. Л. 156–157).

Глава 3

Разметка текстов исторических источников – средство представления, обработки и семантического связывания источниковой информации

При выборе способа цифрового представления письменного исторического источника зачастую возникает дилемма – изображение или текст¹³⁶. Изображение кажется наиболее естественным способом цифрового представления, так как позволяет максимально полно передать все особенности внешней формы источника, включая даже такие детали, которые не заметны для невооруженного взгляда исследователя¹³⁷. Можно привести примеры, когда комплексы исторических источников представлены именно в виде их изображений в растровом формате, снабженных метаописаниями¹³⁸. Однако в компьютерном источниковедении приоритет, безусловно, должен быть отдан репрезентации прежде всего текстов исторических источников¹³⁹, так как такое представление позволяет не только выполнять поисковые запросы и связывать информацию источников, но и применять к источникам сколь угодно сложные методы аналитической и синтетической критики. Для цифрового представления текстов источников, ориентированного на их всестороннее изучение, можно использовать технологии баз данных, основанных либо на реляционной (табличной) модели представления информации, либо на модели XML-дерева, так как другие форматы представления текста (например, HTML, PDF) ориентированы прежде всего на его визуальное восприятие.

Поскольку основной целью компьютерного источниковедения является выявление информационного потенциала – т.е., главным образом, смыслового содержания – исторических источников, в на-

¹³⁶ Разумеется, такой вопрос не возникает в случае репрезентации, например, изобразительных источников. См. подробнее: Bolvig, A., Lindley, Ph., eds. *History and Images: Towards a New Iconology*. Turnhout, 2003. Pp. 59–246 («Image Databases and History»). Хотя методы критики иконографических источников могут существенно обогатить инструментарий компьютерного источниковедения, эти вопросы все же выходят за рамки настоящей монографии.

¹³⁷ Hollaus, F., Gau, M., Miklas, H., Sablatnig, R. *Modern Technologies for Manuscript Research*. In: *Информационные технологии и письменное наследие. El'Manuscript–2012*. Петрозаводск; Ижевск, 2012. С. 57–64.

¹³⁸ Олонецкие губернские ведомости. <http://ogv.karelia.ru/> (17.11.2012); Codices Electronici Sangallenses. <http://www.cesg.unifr.ch> (17.11.2012).

¹³⁹ Желательно при этом текст источника объединить с его изображением, как это делается в ряде проектов. См., например: Édition électronique des chartes de l'abbaye de Saint-Denis. <http://saint-denis.enc.sorbonne.fr> (17.11.2012).

стоящей монографии выбор сделан в пользу технологии XML (Extensible Markup Language¹⁴⁰), которая позволяет выделять в источниках не только фрагменты текста и иерархию его структуры, но и смысловые блоки и элементы, а также включать их в единое «информационное пространство» семантической сети¹⁴¹. При этом можно утверждать, что метод разметки текстов источников¹⁴² на основе технологии XML в наше время является универсальным способом стадальной обработки, структурирования, агрегирования, связывания и интерпретации как собственно источниковой информации, так и метаинформации¹⁴³. Таким образом, разметка источников становится непременным условием для последующего осуществления аналитических и синтетических операций, в том числе и с использованием технологий семантического Web. Не следует забывать и о том, что технология XML весьма эффективна и для подготовки публикаций исторических источников, как печатных, так и электронных¹⁴⁴.

¹⁴⁰ Extensible Markup Language (XML). <http://www.w3.org/XML/> (09.12.2012).

¹⁴¹ Преимущества технологии XML, особенно в сравнении с технологией создания реляционных баз данных, рассмотрены в разделе «XML: особенности, преимущества, схемы разметки TEI и CEI» этой главы монографии.

¹⁴² Понятие «разметка текстов» возникло в полиграфии, так называли подготовку машинописных текстов к типографскому набору с помощью внесения в них различных маркеров для обозначения курсива, жирного шрифта и т.д. С появлением компьютерных технологий разметкой стали называть внесение в текст служебных символов и слов, выделяющих в нем структурные или смысловые фрагменты. Разметка стала применяться не только в компьютерной полиграфии, но и как средство структурирования текстовой информации в электронном документе. Процесс компьютеризации гуманитарных наук уже в 80-е годы XX века привел к использованию технологий разметки в процессе подготовки публикаций и в последующем исследовании памятников письменности.

¹⁴³ В последние годы разметка переходит из разряда технических на уровень методологических проблем e-humanities. См., например: Schmidt, D. The Role of Markup in the Digital Humanities. *Historical Social Research*. 2012. Vol. 37, no. 3. Pp. 125–146.

¹⁴⁴ Об особенности применения технологии XML в источниковедческом анализе и синтезе, а также для разметки исторических источников и создания электронных архивов памятников письменности подробно писали авторы монографии. См.: Иванов А. С., Варфоломеев А. Г. Использование технологии XML для введения в научный оборот комплекса документов *Moscowitica–Ruthenica...*; Ivanovs, A., Varfolomeyev, A. *Editing and Exploratory Analysis of Medieval Documents by Means of XML Technologies...*; Иванов, А. С. Работа с XML-документом как воспроизведение основных этапов источниковедческой критики...; Филатов В. О., Кравцов И. В., Варфоломеев А. Г. Информационная система для работы с полнотекстовыми базами данных исторических документов на основе технологии XML. В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Восьмой Всероссийской научной конференции (RCDL'2006)*. Ярославль, 2006. С. 337–344; Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения...; Филатов В. О., Кравцов И. В. Технологии создания информационной системы для работы с полнотекстовыми базами данных исторических документов. В кн.: *Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам*. Ижевск, 2006. С. 168–173; Варфоломеев А. Г., Кравцов И. В. Использование технологии XML для публикации методики и результатов

XML: особенности, преимущества, схемы разметки TEI и CEI

XML¹⁴⁵ можно определить как язык разметки текста, подобный HTML¹⁴⁶, но с произвольным набором тегов. Однако язык с произвольным набором слов (тегов) – это не язык, а бесконечное множество разных языков, один из которых – HTML. Поэтому точнее будет определить XML как метаязык, то есть язык для создания (описания) других языков, или различных схем разметки текста. Т.е. XML – это свод правил для создания новых языков разметки текста, представляющих текст в виде вложенных друг в друга логических фрагментов.

С 80-х годов XX века существует метаязык SGML (Standard Generalized Markup Language)¹⁴⁷, на основе которого разработан HTML. XML был создан как, в определенном смысле, альтернатива SGML. В отличие от SGML, новый метаязык XML изначально был ориентирован на применение в сети Интернет. Кроме того, он проще в использовании, чем SGML. И, наконец, XML является более последовательной и четкой, чем SGML, реализацией универсальной технологии, позволяющей записывать в виде текстовой строки любую систему данных, структура которой может быть представлена в виде дерева (см. рис. 3.1). Указанные выше преимущества и объясняют популярность XML.

На сегодняшний день существует много схем разметки, созданных в рамках XML и претендующих на роль стандартов в описании тех или иных видов информации: например, MathML¹⁴⁸ (язык описа-

исследования текстов исторических источников. В кн.: *Информационный бюллетень Ассоциации «История и компьютер». Специальный выпуск*, № 34. Москва; Тамбов, 2006. С. 63–64; Варфоломеев А. Г., Иванов А. С. XML-технологии в анализе информации полнотекстовых баз данных...; Варфоломеев, А. Г., Иванов, А. С. Технология XML: Современная реализация источникориентированного подхода...; Варфоломеев А. Г., Иванов А. С. Принципы электронных публикаций комплексов исторических документов со средствами палеографического, текстологического и дипломатического анализа...; Иванов А. С., Варфоломеев А. Г. Публикация и анализ рукописных исторических документов с помощью технологии XML... См также: Hansen, A. M. Text Encoding of Manuscripts: Danish Prayer Books from the 16th Century. *Le Médiéviste et l'ordinateur. Histoire médiévale, informatique et nouvelles technologies*. 2002, no. 41. http://lemo.irht.cnrs.fr/41/mo41_09.htm (05.12.2012); *The Menota Handbook: Guidelines for the Electronic Encoding of Medieval Nordic Primary Sources. Version 1.0*. Bergen: Medieval Nordic Text Archive, 2003. www.hit.uib.no/menota/guidelines (05.12.2012) и др.

¹⁴⁵ <http://www.w3.org/XML/> (17.11.2012).

¹⁴⁶ Шафер С. *HTML, XHTML и CSS. Библия пользователя*. 5-е изд. Москва, 2010; W3C HTML. <http://www.w3.org/html/> (17.11.2012).

¹⁴⁷ Herwijnen, E. van. *Practical SGML*. 2nd ed. Boston, 1994; Shirrell, J. *XML: A Deeper Understanding*. [Online Book]. Chapter 2: SGML. <http://www.xmlbook.info/> (17.11.2012).

¹⁴⁸ W3C Math Home. <http://www.w3.org/Math/> (17.11.2012).

ния математических формул), SMIL¹⁴⁹ (язык описания мультимедийных презентаций), SVG¹⁵⁰ (язык описания векторной графики), охарактеризованные далее TEI и CEI и др.

Текст на таких языках состоит из служебных слов – тегов, которые помещаются в угловые скобки: <tag>. Теги служат для выделения логических фрагментов XML-документа – элементов, для этого они объединяются в пары из открывающего и закрывающего тегов вида <tag>...</tag>. Внутри такой пары могут находиться другие элементы, а также сам текст, не заключенный в угловые скобки, – текстовое содержимое элемента. Если внутри элемента ничего нет, то открывающий тег совмещают с закрывающим, получая так называемый пустой тег: <tag/>. Имена тегов отличают друг от друга разные типы фрагментов текста источника, кроме того, для уточнения этих различий теги снабжаются атрибутами (служебными словами, обозначающими свойства элементов), принимающими определенные значения:

```
<tag attribute1="value1" attribute2="value2" ... >  
...  
</tag>
```

Таким образом, разметка исходного текста источника XML-тегами способна выделять разные по типам и свойствам структурные и смысловые единицы текста, которые могут вкладываться друг в друга и описывать структуру и, в формализованном виде, содержание исторического источника с любым уровнем детальности, вплоть до отдельных слов или даже символов. Однако на разметку накладываются определенные ограничения. Во-первых, элементы не должны пересекаться, то есть в XML запрещены конструкции типа <tag1>...<tag2>...</tag1>...</tag2>. Во-вторых, все элементы XML-документа должны быть вложены в один корневой элемент, который охватывает весь размеченный текст исторического источника. В результате, как уже отмечалось, XML-документ всегда можно представить как разветвляющуюся структуру с одним корнем (дерево).

¹⁴⁹ Bulterman, D. C. A., Rutledge, L. W. *SMIL 3.0. Flexible Multimedia for the Web, Mobile Devices and Daisy Talking Books*. 2nd ed. Berlin, 2009; Synchronized Multimedia Integration Language (SMIL 3.0). W3C Recommendation 01 December 2008. <http://www.w3.org/TR/2008/REC-SMIL3-20081201/> (09.12.2012).

¹⁵⁰ Scalable Vector Graphics (SVG). <http://www.w3.org/Graphics/SVG/> (09.12.2012); Eisenberg, J. D. *SVG Essentials*. [S.l.], 2002.

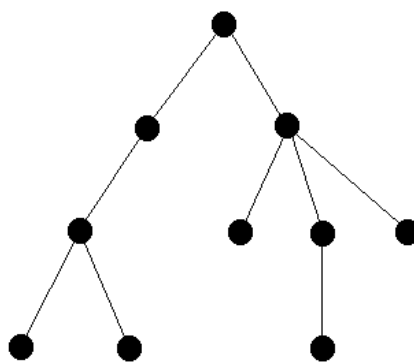


Рис. 3.1. Принципиальная структура XML-документа

Черными кружками на рис. 3.1 отмечены так называемые узлы дерева, соответствующие элементам XML-документа, а прямые линии показывают иерархическую вложенность одних элементов в другие. Каждый узел такого дерева может иметь «дочерние» узлы, текстовое «содержимое» и ряд атрибутов (свойств), принимающих те или иные значения. Узлы могут быть «пустыми», то есть не иметь «дочерних» узлов и текстового содержимого. Каждому узлу дерева в XML-документе соответствует элемент (пара из открывающего и закрывающего тега вместе с тем, что находится между ними, либо пустой тег).

Наряду с таблицей, дерево является одной из фундаментальных структур данных. Интуитивно понятно, что для представления одних данных (например, для описания множества однородных объектов с одинаковым набором признаков) подходят таблицы, а в других случаях (например, для отражения истории возникновения и взаимосвязей списков летописей) логичнее и удобнее использовать именно дерево. Таким образом, XML-документ является способом представления структурированных данных, а технология XML в целом – одной из парадигм в проектировании баз данных.

Алгоритм применения технологии XML заключается в том, что сначала разрабатывается (или подбирается) подходящая схема (схемы) разметки, способная передать все интересные для исследователя палеографические и языковые особенности источников, особенности внутренней формы (структуры, формуляра), смысловые аспекты и т.п. Затем тексты исторических источников переводятся в электронный вид, размечаются, и, наконец, создается программный инструментарий для поисковых и аналитических запросов к созданным XML-файлам, а также для преобразования их в форматы, удобные для электронных и бумажных публикаций.

XML-документ как база данных существенно отличается от реляционной базы данных. Реляционные базы ориентированы на хранение жестко структурированных данных, например, метаданных об историческом источнике, хотя гибкость этой технологии позволяет передать с ее помощью любую текстовую разметку. Для этого надо разбить текст на простейшие элементы (слова или символы), из которых затем собираются более сложные фрагменты. Однако хранение текста в виде таблицы слов или символов может быть неудобным – теряется целостность источника, а для работы с разметкой приходится создавать специализированную программную оболочку. Структура XML-документа намного менее регулярна – узлы могут содержать разное число «дочерних» узлов и атрибутов, разным узлам могут соответствовать одинаковые имена тегов, подчеркивая их однотипность, наконец, порядок элементов в XML-документе не произвольный, как у записей в таблице; он связан с тем текстом, на который накладывается разметка. В результате получается такая база данных, которую можно охарактеризовать как слабоструктурированную и полнотекстовую. Т.е. технология XML выступает как альтернатива реляционным базам данных и при этом является технологией, хорошо подходящей для создания источник-ориентированных баз данных¹⁵¹.

Для компьютерного источниковедения особое значение имеет то, что технология XML сохраняет целостность исторического источника как отдельного текстового файла, доступного для редактирования и обработки с помощью универсального программного обеспечения (например, простейших текстовых редакторов типа Notepad++¹⁵²). Технология XML особенно эффективна, если возникает необходимость интеграции комплекса исторических источников в единую базу данных с целью проведения операций источниковедческого анализа и синтеза. В данном случае база данных – это единый XML-файл, интегрирующий совокупность источников и представляющий информационную модель комплекса источников.

Важным отличием технологии XML от технологии создания реляционных баз данных является наличие стандартных схем разметки. Конечно, любой исследователь, создающий XML-документ, может разработать свою собственную схему разметки, но тогда он не сможет обмениваться данными с другими исследователями, работающими в той же области. Поэтому любое научное сообщество, использующее XML, старается выработать стандарт разметки, ори-

¹⁵¹ Гарскова И. М. *Базы и банки данных в исторических исследованиях*. Москва; Göttingen, [1994].

¹⁵² Notepad++ Home. <http://notepad-plus-plus.org/> (17.11.2012).

ентированный на специфику исходных данных своей области исследования. Синтаксические правила нового стандарта строго описываются с помощью специальных языков разметки – DTD (Document Type Definition)¹⁵³, XML Schema¹⁵⁴ или RELAX NG¹⁵⁵.

Одним из самых известных стандартов разметки, ориентированных на электронные коллекции текстов для гуманитарных наук, является созданный еще в 1994 году стандарт TEI (Text Encoding Initiative¹⁵⁶), активно используемый историками, литературоведами и лингвистами¹⁵⁷. Этот формат сначала основывался на технологии SGML, которая была адаптирована к разметке электронных версий литературных произведений с сохранением всех особенностей бумажного издания. Затем, начиная с 4-й версии (TEI P4¹⁵⁸), он стал XML-форматом. С 2007 года существует и продолжает совершенствоваться новая версия – TEI P5, учитывающая тенденции развития XML-технологий¹⁵⁹. В результате именно технология XML стала ведущей как в подготовке электронных изданий памятников письменности и для внесения в электронный документ метаинформации об источнике, так и в исследовательской интерпретации текста, изучении его структуры, агрегировании информации коллекций текстов.

Изначально этот стандарт был ориентирован на описание печатных изданий литературных произведений с «разбивкой» текстов на главы и абзацы, стихов – на строфы, с указанием концов страниц, выделением прямой речи героев и т.д. Но возможности, заложенные в стандарт, намного шире, что позволяет активно использовать его не только для создания полноценных электронных копий печатных изданий, но и для выделения системы логических фрагментов в текстах произвольной природы, а также для связи этих фрагментов между собой. В результате TEI можно считать прежде всего инструментом анализа и интерпретации текста, что делает его незаменимым для решения задач компьютерного источниковедения.

Традиционно TEI имеет модульную структуру – кроме набора базовых элементов и атрибутов, имеется также много дополнитель-

¹⁵³ DTD Tutorial. <http://www.w3schools.com/dtd/> (17.11.2012).

¹⁵⁴ XML Schema. <http://www.w3.org/XML/Schema> (17.11.2012).

¹⁵⁵ RELAX NG Home page. <http://relaxng.org> (17.11.2012).

¹⁵⁶ Text Encoding Initiative. <http://www.tei-c.org> (07.08.2012).

¹⁵⁷ Использование формата TEI для публикации и анализа литературных памятников имеет давние традиции особенно в литературоведении. См., например: Vanhoutte, E. An Introduction to the TEI and the TEI Consortium. *Literary and Linguistic Computing*. 2004. Vol. 19, Issue 1. Pp. 9–16.

¹⁵⁸ TEI P4. Guidelines for Electronic Text Encoding and Interchange. TEI Consortium, 2001. <http://www.tei-c.org/Vault/P4/doc/html/> (09.12.2012).

¹⁵⁹ Wittern, C., Ciula, A., Tuohy, C. The Making of TEI P5. *Literary and Linguistic Computing*. 2009. Vol. 24, no. 3. Pp. 281–296. doi: 10.1093/lc/fqp017.

ных комплексов элементов и атрибутов, ориентированных на конкретные задачи или виды и жанры текстов. На сайте Консорциума TEI¹⁶⁰ приводится список проектов, в той или иной мере использующих этот стандарт для различных литературоведческих и лингвистических целей, электронных публикаций печатных изданий, писем и рукописей. Есть среди них и проекты, посвященные коллекциям исторических документов, например, средневековым историческим источникам. Для историка представляют интерес следующие проекты (разумеется, список не является исчерпывающим):

– Menota (Medieval Nordic Text Archive) – проект, посвященный электронной публикации средневековых рукописей Норвегии, Швеции, Исландии и Дании¹⁶¹;

– MEP (Model Editions Partnership) – проект по разработке единой схемы разметки для электронных публикаций писем исторических деятелей (США)¹⁶²;

– Newton Manuscript Project – электронная публикация рукописей Исаака Ньютона¹⁶³;

– Repertorium of Old Bulgarian Literature and Letters¹⁶⁴;

– CELT Project, посвященный созданию электронной коллекции исторических и литературных ирландских текстов¹⁶⁵.

Характерной чертой этих проектов является то, что они предлагают специфические схемы разметки, основанные на TEI¹⁶⁶, но ориентированные на конкретные особенности текстов, входящих в коллекции, и на решение определенных задач, стоящих перед археографами (издателями) и исследователями.

Как показано далее, стандарт TEI действительно может применяться для многоуровневой и многоплановой разметки исторических источников, причем эта разметка будет ориентирована, прежде всего, на решение задач компьютерного источниковедения, в том числе – комплексного изучения документальных источников. Однако нельзя не отметить, что в данном случае стандарт TEI нуждается в расши-

¹⁶⁰ Projects Using the TEI. <http://www.tei-c.org/Activities/Projects/> (17.11.2012).

¹⁶¹ Medieval Nordic Text Archive (Menota) <http://www.menota.org> (17.11.2012).

¹⁶² The Model Editions Partnership. <http://wyatt.elasticbeanstalk.com/mep/home> (17.11.2012).

¹⁶³ The Newton Project. <http://www.newtonproject.sussex.ac.uk> (17.11.2012).

¹⁶⁴ Repertorium of Old Bulgarian Literature and Letters. <http://clover.slavic.pitt.edu/~repertorium/> (17.11.2012).

¹⁶⁵ CELT. Corpus of Electronic Texts. <http://www.ucc.ie/celt/> (17.11.2012).

¹⁶⁶ Poupeau, G. Réflexions sur l'utilisation de la TEI pour coder les sources diplomatiques à partir de l'exemple du Cartulaire blanc de l'abbaye de Saint-Denis. *Le Médiéviste et l'ordinateur: Histoire médiévale, informatique et nouvelles technologies*. 2004, no. 43. <http://lemo.irht.cnrs.fr/43/43-12.htm> (05.12.2012); Hansen, A. M. Text Encoding of Manuscripts: Danish Prayer Books from the 16th Century...

рении – дополнении его элементами (тегами), которые позволяют выделять как структурные части актовых источников и документов приказного делопроизводства, так и смысловые фрагменты, в которых отразились различные аспекты исторической реальности.

Ряд конкретных задач компьютерного источниковедения позволяет решить и созданная на основе стандарта TEI схема разметки средневековых актов CEI (*Charters Encoding Initiative*)¹⁶⁷, которая в последние годы более-менее интенсивно разрабатывается в Мюнхенском университете (Ludwig-Maximilians-Universität München); при этом в разработке и апробации схемы принимают участие специалисты из многих стран, в которых историческая дипломатика развивается давно и успешно, – из Германии, Франции, Италии, Испании, Нидерландов и др. Преимущество данного стандарта – наличие специальных элементов для очень детальной разметки формуляра средневековых актов. Причем наименования этих элементов (разумеется, на латинском языке) полностью соответствуют принятым в дипломатике обозначениям структурных частей средневекового западноевропейского акта¹⁶⁸. Другое преимущество схемы – возможность более точной разметки, соответственно, и представления метаинформации об актовых источниках, принятой в актовой археографии¹⁶⁹. Таким образом, схема CEI ориентирована как на формулярный анализ документов, так и на подготовку высококачественных печатных и электронных дипломатических изданий актов.

Нельзя не отметить, что в последующих разделах данной главы авторы монографии использовали разные схемы разметки исторических источников – TEI, CEI, а также собственные схемы (точнее – элементы). Такое разнообразие отражает восьмилетний процесс апробации технологии XML с непрекращающимся поиском схемы разметки, адекватно соответствующей задачам комплексного компьютерного источниковедения. Очевидно, что готового «идеального» стандарта с полным набором необходимых элементов нет и, судя по всему, не будет. Да и вряд ли такой стандарт, в принципе, возможен, так как нельзя предвидеть, какие конкретные задачи может поставить творческое воображение историка перед компьютерным источниковедением. Нужно учитывать и бесконечное разнообразие как источ-

¹⁶⁷ CEI – Charters Encoding Initiative. <http://www.cei.lmu.de/> (17.11.2012).

¹⁶⁸ Ortí, M. M. C., ed. *Vocabulaire International de la Diplomatie*. 2nd ed. València: Universitat de València, 1997. Дополненная и частично модернизированная версия словаря доступна также в Интернете: *Vocabulaire Internationale de Diplomatie (Terminologie der Diplomatie)*. <http://www.cei.lmu.de/VID/> (05.12.2012).

¹⁶⁹ См. подробнее: Каштанов С. М. *Актовая археография*. Москва, 1998. С. 19–103.

ников и их комплексов, так и источниковой информации. Наверно, придется смириться с тем, что выбор определенных схем разметки в компьютерном источниковедении всегда будет носить ситуационный характер. Но это не значит, что результаты работы неизбежно будут носить локальный характер – базовые подходы и принципы разметки вряд ли придется менять, а конкретный набор элементов (тегов) и, соответственно, схема разметки могут быть соотнесены с имеющимися стандартами. В свою очередь, гибкое использование различных схем разметки может стать эффективным средством актуализации как свидетельств источников, так и информации о них.

В источниковедческой практике давно сложился определенный алгоритм выделения (репрезентации) информации об исторических источниках¹⁷⁰, при этом каждый источник фактически рассматривается как многоуровневый объект. Прежде всего предоставляется «внетекстовая» информация об источнике – исторических условиях и обстоятельствах его возникновения, об авторе, месте хранения, о палеографических особенностях (внешней форме) источника. Далее выделяется собственно источниковая информация, при этом происходит как бы многоуровневая репрезентация его информации: сначала – информация в рамках крупных логических фрагментов, отражающих внутреннюю форму (структуру) источника, затем – свидетельства источников о конкретных аспектах исторической реальности. Специально выделяется также информация о стилистических, лингвистических и других особенностях текстов. Выделение данных информационных аспектов является основой для интерпретации информации как отдельного источника, так и их совокупностей¹⁷¹. В соответствии с алгоритмом репрезентации информации в традиционном источниковедении, в компьютерном источниковедении должна соблюдаться стадийность в представлении данных, а разметка должна быть многоуровневой и сочетать в себе самые различные параметры, которые могут представлять интерес для историка или источниковеда. В ходе разметки неизбежно происходит формализация информации источников, однако это не трансформирует оригинальный текст источника, поскольку формализация здесь проявляется как условное, стандартизированное и унифицированное, обозначение (маркировка) выделен-

¹⁷⁰ *Источниковедение: Теория. История. Метод...* С. 127–143.

¹⁷¹ Впрочем, само по себе выделение информационных аспектов уже является определенной интерпретацией источника, составной частью аналитической критики.

ных частей документа¹⁷², то есть, в сущности, формализация является интерпретацией источника, «надстройкой» над его текстом.

Представление метаинформации и информации о палеографических особенностях источников

С точки зрения источниковедения и археографии, метаинформация (или метаданные) – это информация о публикуемом памятнике письменности, которую предоставляет издатель. Минимально необходимый набор метаданных включает сведения о месте и времени появления памятника письменности, авторе, месте хранения оригинала (архивный шифр). В научных изданиях об источнике представляется максимально полная информация, которая включается в издательский заголовок, археографическую легенду и комментарии. Соответственно, к метаинформации относятся не только дата, место возникновения, сведения об индивидуальном или коллективном авторе источника, месте его хранения, но и издательский заголовок, порядковый номер документа в комплексе (публикации), вид и разновидность источника, язык, сведения о предыдущих публикациях, обоснование датировки и атрибуции (по необходимости)¹⁷³, сжатое или, наоборот, детальное описание палеографических особенностей рукописей¹⁷⁴ и др.

В компьютерном источниковедении разметка любого исторического источника начинается с репрезентации метаинформации о самом источнике, которая дается в соответствии с правилами археографии¹⁷⁵. Как правило, метаинформация соответствует заголовку и легенде¹⁷⁶, которые сопровождают текст источника при его публикации, поскольку и разметка сама по себе может рассматриваться как средство подготовки источника к печатной публикации или репрезентации в сети Интернет, что, в сущности, также является его публикацией.

Существующие схемы разметки TEI и CEI предоставляют практически исчерпывающий набор элементов, которые могут потребо-

¹⁷² Ср.: Орешников А. С. О применении формализации к описанию деловых текстов. *Источниковедение и история русского языка*. Москва, 1964. С. 150–156.

¹⁷³ См.: Каштанов С. М. *Актная археография...* С. 21–35, 74–76. Ср. с подходом представления информации в филологии: Лихачев Д. С. *Текстология: На материале русской литературы X–XVII веков*. Санкт-Петербург, 2001. С. 526–530.

¹⁷⁴ *Правила лингвистического издания памятников древнерусской письменности*. Москва, 1961. С. 29–34.

¹⁷⁵ *Правила издания исторических документов в СССР...* С. 88–121; Зимин А. А. *Методика издания древнерусских актов*. Москва, 1959.

¹⁷⁶ Валк С. Н. Археографическая «легенда». В кн.: *Археографический ежегодник за 1971 год*. Москва, 1972. С. 5–16.

ваться для репрезентации такой метаинформации. Поэтому репрезентация метаинформации является лишь вопросом техники.

Так, можно использовать стандартный набор тегов TEI, предназначенных для электронного кодирования титульного листа опубликованного текста, фиксации библиографической информации и описания источников. Однако данный стандарт не учитывает специфику описания архивных источников, соответственно, набор тегов должен быть расширен. Например, в заголовке источника археограф или исследователь должен определить разновидность документа, кратко обозначить содержание источника, предложить датировку. Поэтому в системе разметки необходимы дополнительные теги, которые могут быть частично импортированы из стандарта CEI, например, элемент для обозначения вида (разновидности) документа `<class>...</class>`; элемент для выделения краткого содержания источника в заголовке `<contents>...</contents>`; тег для выделения информации о месте хранения архивного оригинала `<repository>...</repository>`. Для сведений о предыдущих публикациях документа можно ввести тег `<published>...</published>`, который отсутствует и в схеме разметки TEI, и в стандарте CEI. При разметке той части текста легенды, которая относится к палеографическому описанию документа, набор тегов может потребовать существенного расширения. Так, необходимым является тег `<authenticity>...</authenticity>` для констатации подлинности документа. Система кодирования текстов TEI позволяет при разметке более детально представить метаинформацию об источнике: выделить персоналии и топонимы, упомянутые в заголовке документа, более детально разметить библиографическую информацию, представить в системе разметки и научный аппарат – примечания, комментарии.

В целом, репрезентация метаинформации об источнике на основе схемы разметки TEI выглядит следующим образом (в сокращении¹⁷⁷):

```
<doc id="doc1">
  <head>
```

¹⁷⁷ Здесь размечены заголовок и археографическая легенда одного из источников, входящего в комплекс документов приказного делопроизводства по истории Динабурга: отписка боярина и воеводы Семена Лукьяновича Стрешнева царю Алексею Михайловичу об осаде города Динабурга, которая датируется не ранее 19 июля 1656 года. См.: РГАДА. Ф. 210. Московский стол. Столб. 270. Л. 674–675. В комплексе источников по истории Динабурга идентификационный номер документа – 1 (в соответствии с нумерацией в публикации: Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов...* Часть 1. № 1, с. 67–68). Далее в примерах этот документ будет обозначен именно так.

```

<class> Отписка </class> <sender> боярина и
воеводы Семена Лукьяновича Стрешнева
</sender> <addressee> царю Алексею Михайло-
вичу </addressee> <contents> об осаде горо-
да Динабурга </contents>.
<docDate>
  не ранее 19 июля 1656 г.
</docDate>
</head>
<sourceNote>
  <repository>
    РГАДА. Ф.210, Московский стол. Столб.270.
    Л.674-675
  </repository>.
  <authenticity>
    Подлинник
  </authenticity>.
  <published>
    <bibl> Иванов А., Кузнецов А. Динабург в
    документах Российского государственного
    архива древних актов (1656-1666). Даугав-
    пилс, 2002. Ч.I. № 1.</bibl>
  </published>
</sourceNote>
  ...
</doc>

```

Альтернативная разметка метаинформации предлагается в схеме разметки CEI, впрочем, «альтернативной» ее можно назвать лишь условно, так как имеется ряд общих элементов, что означает, что обе схемы, в принципе, совместимы. Для средневекового актового материала схема CEI, безусловно, приоритетна, так как она полностью учитывает традицию эдиции данного вида исторических источников. В качестве примера разметки метаинформации на основе схемы CEI можно привести заголовок и легенду грамоты архиепископа рижского смоленскому князю Федору Ростиславичу с обвинением жителей города Витебска в необоснованной жалобе на рижан, датируемую 1284–1297 годами¹⁷⁸:

¹⁷⁸ ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3. Как и документам из комплекса источников по истории Динабурга, документам из субкомплекса смоленско-рижских актов присвоены номера по их последней публикации (Иванов А., Кузнецов А. *Смоленско-рижские акты...* № 6, с. 636). Номер данной грамоты – 6; под этим номером документ может быть идентифицирован в последующих примерах как в данной, так и в четвертой главе.

```

<charter n="6">
  <abstract>
    <class> Грамота </class> <issuer> архиепи-
    скопа Рижского Иоганна II (?) </issuer>
    <addressee> смоленскому князю Федору Рости-
    славичу </addressee> <contents> с обвинением
    жителей города Витебска в необоснованной жа-
    лобе на рижан </contents>
  </abstract>
  <issued>
    Рига, [1284-1297 гг., очевидно, ближе к
    1285-1287 гг.]
  </issued>
  <chDesc>
    <country> Латвия </country> <settlement>
    Рига </settlement> <repository> Латвий-
    ский национальный архив - Латвийский го-
    сударственный исторический архив
    (Latvijas Nacionālā arhīva Latvijas
    Valsts vēstures arhīvs). Ф. 673. Оп. 4.
    К. 18. № 8. Л. 3 </repository>
    <textLang> древнерусский </textLang>
    <authenticity> список (концепт?)
    </authenticity>
    <physDesc> <material> пергамен, 16,5 ×
    23,3 см </material> </physDesc>
    <bibl> Иванов А., Кузнецов А. Смоленско-
    рижские акты: XIII в. - первая половина
    XIV в.: Документы комплекса Moscovitica-
    Ruthenica об отношениях Смоленска и Риги.
    [Rīga], 2009. № 6; Russisch-livländische
    Urkunden. Gesammelt von K. E. Napiersky.
    St. Petersburg, 1868. N. 34; Liv-, Esth-
    und Curländisches Urkundenbuch nebst Re-
    gesten. Hrsg. von Dr. Fr. G. von Bunge.
    Bd. VI. Riga, 1873. N. 3051 </bibl>
  </chDesc>
  <tenor>
    ...
  </tenor>
</charter>

```

Приведенный пример несколько упрощенной разметки метаинформации на основе схемы разметки CEI показывает, что есть воз-

возможность достаточно полно выделить данные о палеографических особенностях оригинала, включая детальное описание, например, печатей, скрепляющих документ, материала, почерка, филиграней, архивных помет и т.д.¹⁷⁹ Однако исследователя средневековых исторических источников такое описание не всегда может полностью удовлетворить. Для него важны не только общие замечания о палеографии оригинала источника, но и конкретные детали – деление текста на листы и строки, графические особенности начерка букв (особенности почерка), дефекты и повреждения оригинала. Поэтому при разметке источников нельзя ограничиться только детальным представлением метаинформации как таковой. При воспроизведении самого текста на него необходимо «наложить» разметку (условно ее можно назвать палеографической), благодаря которой можно создать максимально точное представление о реальном внешнем виде – внешних признаках источника, компоновке текста, особенностях почерка, т.е. осуществить не просто дипломатическую, но и палеографически точную публикацию источника¹⁸⁰. Технических ограничений здесь, в сущности, нет¹⁸¹. Эта разметка также служит для репрезентации своего рода метаданных, «внетекстовой» информации об историческом источнике.

Разметка, обозначающая распределение текста по листам, на которых написан источник, окончания строк, а также разделение текста на слова (если в тексте источника отсутствует словоделение), составляет как бы нижний уровень, на который «накладывается» деление текста на логические – семантические и структурные – фрагменты. В

¹⁷⁹ См. подробнее: CEI. <http://www.cei.lmu.de/> (07.08.2012).

¹⁸⁰ В задачу настоящей монографии не входит участие в дискуссии историков и лингвистов о необходимой степени точности воспроизведения палеографических особенностей рукописей в изданиях русских средневековых исторических источников. Историки предлагают упрощенный вариант передачи текстов (в примерах далее этот вариант использован в передаче текстов источников XVII века по истории Динабурга). См.: *Правила издания исторических документов в СССР...* Лингвисты же считают, что воспроизведение текстов, особенно древнерусских грамот XII–XIV вв., должно быть более точным (по этим правилам в примерах воспроизведены акты и их фрагменты на древнерусском языке из комплекса смоленско-рижских грамот). См.: *Правила лингвистического издания...* И все же отметим, что несогласованность правил издания источников вряд ли можно считать оправданной: уже в 60-е годы XX века некоторые исследователи об этом писали. См.: Котков С. И. О совместном издании древнерусских скорописных памятников лингвистами и историками. В кн.: *Лингвистическое источниковедение*. Москва, 1963. С. 5–23. По мнению авторов, воспроизведение текстов средневековых русских источников по правилам, разработанным лингвистами, позволит расширить круг задач, которые может решать компьютерное источниковедение. К тому же аргументы в пользу предельно упрощенной передачи текстов, на взгляд авторов, не выдерживают критики. См., например: Корнева И. И. Основные принципы передачи текстов документов при публикации и некоторые вопросы воспроизведения источников ранних периодов (XIV–XVIII вв.). В кн.: *Труды МГИИИ*. Том. VIII. Москва, 1957. С. 43–65.

¹⁸¹ См., например, воспроизведение памятников древнерусской письменности на сайте «Манускрипт. Славянское письменное наследие». <http://manuscripts.ru/> (06.08.2012).

результате соответствующим образом размеченная текстовая информация источника окажется как бы «вложенной» (включенной) в данную разметку. Такое органичное сочетание различных уровней разметки необходимо для точной локализации (лист, строка, слово) любого фрагмента текста источника в ходе источниковедческой критики.

Для разметки строк и листов разумно применять пустые теги. Например, в схемах разметки TEI и CEI используются теги <lb/> – конец строки, <pb/> – начало очередной страницы. Поскольку в архивном деле пронумерованы, как правило, листы, при указании номера страницы приходится употреблять дополнительные обозначения – «recto» для лицевой стороны и «verso» для оборотной стороны: <pb n="674" type="recto"/> – обозначает начало текста на лицевой стороне 674-го листа в архивном деле. Отметим, что выделение строк необходимо, прежде всего, для электронного представления оригинального древнерусского текста; для источников XVII века эти элементы палеографической разметки не обязательны. Текст отдельно взятого источника при этом включается в пространство (поле), ограниченное тегами <body>...</body> (TEI) или <tenor>...</tenor> (CEI). Соответственно, метаинформация об отписке С. Л. Стрешнева (документ № 1) может быть размечена следующим образом в соответствии со схемой разметки TEI. В данную разметку «вкладывается» разметка собственно текста исторического источника.

```
<doc id="doc1">
  <head>...</head> <sourceNote>...</sourceNote>
  <body>
    <pb n="674" type="recto"/>
      Текст первой строки<lb n="1"/>
      Текст второй строки<lb n="2"/>
      ...
    ...
  </body>
</doc>
```

В разметке CEI ниже представлена общая модель разметки послания архиепископа рижского (грамота № 6):

```
<charter n="6">
  <abstract>...</abstract>
  <issued>...</issued><chDesc>...</chDesc>
  <tenor>
```

```
Текст первой строки<lb n="1"/>
Текст второй строки<lb n="2"/>
...
</tenor>
</charter>
```

Описанные выше операции по выделению и компьютерной репрезентации метаинформации об источнике можно считать предварительными. Качество их проведения, бесспорно, влияет на точность критики текстов источников, и особенно – на фиксацию ее результатов.

Более сложная проблема – палеографически точное воспроизведение текстов письменных памятников, особенно на древнерусском языке. С одной стороны, некоторые вопросы могут быть решены в рамках стандартов TEI и CEI, например, выделение выносных букв¹⁸² или разметка исправлений в тексте (в отпусках и концевых грамах), позволяющая увидеть как вариант текста до исправления, так и конечный результат правки¹⁸³.

С другой стороны, представление символов и диакритических знаков, характерных для средневековых русских текстов, требует определенных усилий. В решении задачи электронной репрезентации средневековых русских источников может помочь методика проекта Menota¹⁸⁴, согласно которой все символы, встречающиеся в текстах рукописей, должны быть представлены в первичных документах базы данных как XML-сущности (строки вида «&name;»), связанные с определенными кодами стандарта Юникод¹⁸⁵.

В Юникоде каждый символ кодируется с помощью комбинации из 16 бит – от 0 до 65535 в десятичной записи. Диапазон, состоящий из 256 кодов (от 1024 до 1279), выделен под современные и вышедшие из употребления символы, основанные на кириллице. Свои коды есть и для диакритических знаков, таких как «палатализация» и «титло». В HTML-документе символы Юникода можно включать с помощью десятичных кодов в специальном обрамлении: &#код. Например, запись Ѣ заставит браузер найти подходящий шрифт и вывести на экран заглавную букву «ять». Для XML-документа есть более гиб-

¹⁸² Так, выносные символы в схеме разметки CEI предлагается выделять тегами <sup>...<sup>.

¹⁸³ В CEI элемент ... служит для разметки исключенных (зачеркнутых) фрагментов текста, а элемент <add>...</add> – вставленного (вписанного) фрагмента или символа. Комбинация из элементов и <add> позволяет выделить исправления, сделанные писцом.

¹⁸⁴ Medieval Nordic Text Archive (Menota). <http://www.menota.org> (09.12.2012). См. также: *The Menota Handbook...* Chapter 5.

¹⁸⁵ The Unicode Consortium. <http://www.unicode.org> (09.12.2012).

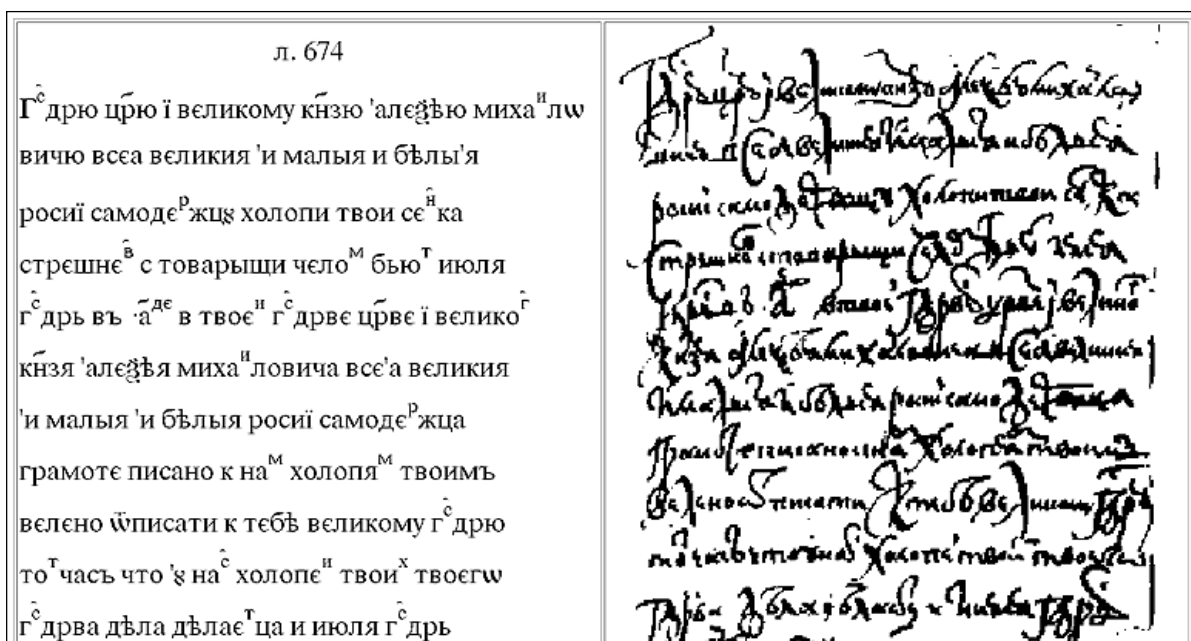


Рис. 3.2. Вариант электронной публикации документа. 1656 г.
[не ранее июля 19]. Отписка С. Л. Стрешнева царю Алексею Михайловичу
об осаде города Динабурга (воспроизведено начало документа). РГАДА.
Ф. 210. Московский стол. Столб. 270. Л. 674–675. Подлинник

Первая строка документа, представленного на рисунке 3.2, в разметке TEI с дополнительным тегом <sup> для выносных букв и с использованием XML-сущностей будет выглядеть так:

```
<w>&GLA;<sup>&slo;&pal;</sup>&dob;&rts;&ju;</w>  

<w>&tsy;&rts;&tit;&ju;</w>  

<w>&id;</w>  

<w>&ve;&est;&lu;&izhe;&ka;&on;&mys;&uk;</w>  

<w>&ka;&na;&tit;&zem;&ju;</w>  

<w>&rsq;&az;&lu;&est;&ksi;&jat;&ju;</w>  

<w>&mys;&izhe;&her;&az;<sup>&izhe;</sup>&lu;&ome;  

</w><lb n="1"/>
```

«&GLA;&slo;&dob;&rts;&ju;» означает «Гсдрю». Что касается &pal и &tit, то это диакритические знаки – палатализация и титло. Диакритические знаки следует рассматривать как отдельные символы, чтобы в максимальной степени добиться соответствия существующим в настоящее время стандартным диапазонам символов Юникода.

Для того, чтобы тексты на разных языках (что актуально для комплекса «Moscowitica–Ruthenica») могли одновременно отображаться на экране монитора, требуется создать специальный шрифт, содержащий в себе все символы, используемые в конкретном проек-

те. Не исключено, что потребуется представление одного и того же текста в разных вариантах – например, для дипломатической публикации, требующей палеографически точного воспроизведения символов, и для аналитических операций – здесь для удобства работы имеет смысл транслитерировать оригинальный текст документа и использовать современные стандартные символы.

Альтернативой использованию TrueType-шрифтов с жесткой привязкой к стандарту Юникод является применение формата SVG¹⁸⁸ для описания желаемой формы символов в виде набора кривых Безье¹⁸⁹ и создания Web-шрифтов, необходимых для определенного проекта. Каждый символ представляется в SVG-файле в виде группы объектов, снабженной идентификатором. Конечно, символ может иметь и другие характеристики, в том числе и коды близких символов Юникода. Идентификаторы, представляющие символы текста, записываются в XML-файлы. В результате мы получаем возможность выводить текст из XML-файла в окно браузера не в виде HTML с использованием TrueType-шрифта, а непосредственно в виде SVG-графики. Существуют программы для разработки таких шрифтов и их конвертации из существующих TrueType-шрифтов. Современные версии браузеров Opera, Safari, Google Chrome отображают SVG-шрифты без установки дополнительных плагинов.

В качестве примера можно рассмотреть одну из букв из древнерусского документа (рисунок 3.3):



Рис. 3.3. Графическая форма древнерусской буквы «аз»

¹⁸⁸ Scalable Vector Graphics (SVG) 1.1 (Second Edition). W3C Recommendation 16 August 2011. Part 20. Fonts. <http://www.w3.org/TR/SVG11/fonts.html> (09.12.2012).

¹⁸⁹ См.: Варфоломеев А. Г., Кравцов И. В., Филатов В. О. SVG-визуализация в цифровых библиотеках рукописных документов. В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Девятой Всероссийской научной конференции RCDL'2007*. Переславль-Залесский, 2007. С. 230–235.

Для описания этого символа с помощью кривых Безье достаточно следующего короткого фрагмента SVG-кода:

```
<g id="az2"><path d="М 113,126 С 107,78 108,79
80,85 С 66,88 51,91 47,91 С 33,91 38,76 55,63 С
74,49 84,53 71,68 С 64,76 65,78 72,78 С 94,78 100,67
93,43 С 87,25 87,17 95,8 С 103,-2 104,-2 117,12 С
124,20 130,36 131,46 С 132,57 133,75 134,86 С
136,115 130,157 123,157 С 120,157 116,143 113,126 z"
style="fill:#000000"/></g>.
```

Отметим, что использование SVG-шрифтов приводит к усилению аналитических возможностей оцифрованного комплекса источников, поскольку предоставляет пользователям возможность не только произвольно формировать множество символов для электронных текстов, но и сравнивать особенности начертания одних и тех же символов в разных источниках, так как контуры, описанные с помощью SVG, становятся доступными для количественного анализа.

И все же использование именно шрифтов в кодировке Юникод приходится признать неизбежным: хотя эти шрифты и не позволяют в полной мере отразить начерки древнерусских букв, любой другой вариант передачи существенно ограничит возможности включения комплекса источников в семантическую сеть, что является необходимым условием для эффективного применения инструментария компьютерного источниковедения.

Для палеографически точного представления древнерусских текстов может потребоваться решать нестандартные задачи. Например, при использовании Юникод-шрифтов при воспроизведении древнерусских букв «семантика» древнерусских символов (букв) может вступить в конфликт с современным значением («семантикой») букв, имеющих ту же графическую форму, что и древнерусские буквы. Так, графическая форма буквы «и» в рукописях XII–XIII веков соответствует графической форме современной буквы «н», а древнерусская буква «н» похожа на латинскую букву «N» (см. рисунок 3.4). Поэтому, при палеографически точном воспроизведении графики букв могут быть затруднены аналитические и синтетические операции по поиску или семантическому связыванию информации источников, так как определенные лексемы для компьютера могут просто утратить (или изменить) свое смысловое содержание.

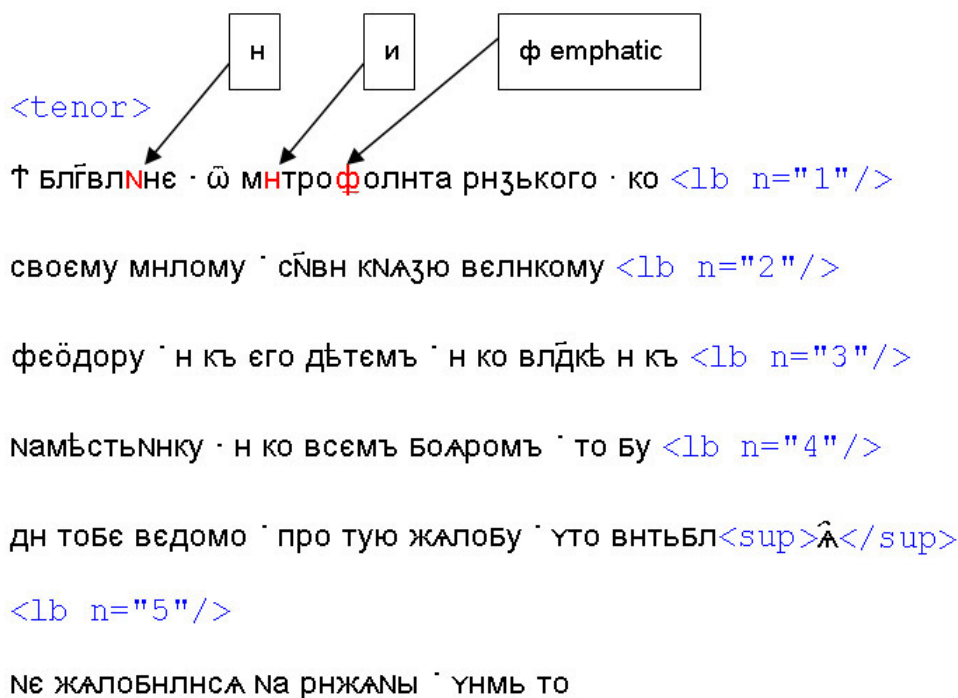


Рис. 3.4. Палеографическое воспроизведение древнерусского текста. [1284–1297 гг.] Грамота архиепископа рижского смоленскому князю Федору Ростиславичу (акт № 6). ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3. Современный список

Смысловое значение («семантика») букв может быть представлено разными способами. Так, стандарт TEI предлагает специальные элементы – глифы (<glyph>), которые используются для описания букв, имеющих особые, нетипичные графические формы. Эти элементы должны использоваться в сочетании с элементом <g>, чтобы вставить букву в текст. Подобная разметка, в сущности, преобразует устаревшие (нетипичные) графические формы древнерусских букв в стандартные буквы русского алфавита:

```
<charDecl>
  <glyph xml:id="n1">
    <glyphName>SMALL CAPITAL CYRILLIC N LIKE
    LATIN N</glyphName>
    <mapping type="standardized">н</mapping>
  </glyph>
</charDecl>
.....
<g ref="#n1">N</g>
```

Те же операции могут быть произведены при помощи языков программирования, например, PHP, JavaScript и др. Причем, в зависимости от целей, замена букв может выполняться одновременно в двух направлениях. С одной стороны, замена нестандартных букв стандартными может делаться в случае необходимости проведения операций по поиску и связыванию информации. С другой стороны, замена стандартных букв на аутентичные позволяет на экране монитора показать палеографически точно воспроизведенный текст древнерусской грамоты.

Итак, технология XML предоставляет все необходимые средства и инструменты для дипломатически и палеографически точного воспроизведения текстов источников, а также для разметки и репрезентации метаинформации. Данный уровень разметки является основой, на которую «накладываются» другие уровни разметки, которые рассматриваются далее в настоящей главе¹⁹⁰.

Информационные модели структуры и содержания отдельных источников и их комплексов

Выявление структуры текстов исторических источников, как правило, является лишь вспомогательным средством для решения более существенных задач аналитической и синтетической критики, направленной на раскрытие содержательной стороны источников и использование их свидетельств для реконструкции прошлого¹⁹¹. При этом тесная взаимосвязь структуры и содержания источников очевидна; тем не менее, она редко становится предметом самостоятельного и глубокого изучения в общем историческом источниковедении¹⁹². Пожалуй, лишь в исследованиях по исторической дипломатике проблема взаимосвязи и взаимозависимости структуры (внутренней формы, формуляра) актов источников и смысловой нагрузки структурных частей их текстов приобретает особое значение в силу специфики метода дипломатики – формулярного анализа¹⁹³. В настоящее время есть все основания предполагать, что данная проблема будет неизбежно

¹⁹⁰ Однако в последующих примерах разметка палеографическая, а также разметка метаданных не будут сохранены по причине громоздкости XML-файлов. Полная многоуровневая разметка сделала бы примеры трудными для визуального восприятия.

¹⁹¹ Разумеется, не отрицается и самостоятельное значение источниковедческого изучения структуры источников.

¹⁹² Варфоломеев А. Г., Иванов А. С. Модели структуры и содержания исторических источников...

¹⁹³ См. подробнее: Лаппо-Данилевский А. С. *Очерк русской дипломатики частных актов*. СПб., 2007. С. 143–164; Каштанов С. М. *Русская дипломатика...* С. 169–193; Каштанов С. М. *Очерки русской дипломатики*. Москва, 1970. С. 53–149.

актуализирована в связи со стремительным развитием информационных технологий, компьютерного источниковедения и появлением новых подходов к репрезентации информации исторических источников в семантических публикациях, так как представление письменного источника в семантических сетях предполагает его предварительную разметку (в терминологии дипломатики – деление или «членение» формуляра на части и компоненты), выявление логической последовательности структурных частей, увязывание формальной и семантической структуры текстов. Соответственно, в этом разделе рассматриваются не только опирающиеся на стандартизированные подходы и операции технологические процедуры по разметке (т.е. членению) текстов документальных исторических источников, связыванию и сравнению размеченных фрагментов текстов, но и проблемы корреляции структуры и содержания источников при выборе схемы разметки, а также возможности через смоделированную информационную структуру текстов раскрыть определенные (лишь некоторые) содержательные аспекты источников.

Принципиальная модель разметки, нацеленной на выделение структурных частей текстов источников, а также сам алгоритм связанного изучения информационной структуры совокупностей источников, относящихся к одной разновидности, заимствованы из исторической дипломатики¹⁹⁴, что вполне оправданно, закономерно и даже неизбежно: именно в дипломатике был детально разработан и всесторонне апробирован метод формулярного анализа, правда, применительно только к актовым источникам.

Прежде всего здесь выделяются наиболее крупные части формуляра – протокол (protocol, Protokoll), основной текст/«контекст» (context, Kontext) и эсхатокол (eschatocol, Eschatokoll)¹⁹⁵. Эти три основные части, в свою очередь, подвергаются более мелкому делению. Так, в протоколе выделяются следующие компоненты: invocatio (посвящение Богу), intitulatio (обозначение лица, от которого исходит документ), inscriptio (обозначение адресата), salutatio (приветствие)¹⁹⁶ и т.д. Впрочем, есть и несколько отличные подходы к делению условного

¹⁹⁴ Поэтому данный уровень разметки источников можно с полным основанием называть дипломатическим.

¹⁹⁵ *Vocabulaire Internationale de Diplomatique...*; Goetz, H.-W. *Proseminar Geschichte: Mittelalter*. 3., überarbeitete Auflage. Stuttgart, 2006. S. 141–147 (Form und Aufbau der Urkunde). Ср.: Каштанов С. М. Дипломатический состав древнерусского акта. В кн.: *Вспомогательные исторические дисциплины*. II. Ленинград, 1969. С. 143–159.

¹⁹⁶ Каштанов С. М. *Русская дипломатика...* С. 170.

формуляра акта¹⁹⁷. Нельзя не отметить, что идентификация данных структурных блоков, зачастую, весьма условна, так как опирается на интерпретацию текстов актов. Это означает, что при выделении формальной структуры акта в дипломатике одновременно происходит связывание структуры с содержанием источников. Соответственно, и процесс разметки – то же членение актов, но с применением современной технологии – также предполагает семантическую детерминированность своих элементов и интерпретацию структурных частей текстов. Особые проблемы возникают в ходе деления текстов древнерусских грамот, поскольку, в отличие от западноевропейских актов, их формуляр был менее унифицирован и стандартизирован. Исследователи формуляров русских средневековых актов при этом отмечают, что «деление условного формуляра на части и компоненты есть схема, не все звенья которой имеются в конкретных документах, а порядок расположения компонентов в схеме далеко не всегда соответствует их расположению в конкретных документах...»¹⁹⁸. Соответственно, при выделении структурных частей в актах нужно учитывать следующее методологическое положение дипломатики: «... при расчленении текста акта на образующие его клаузулы следует, по возможности, соотносываться с теми моментами, которые сами составители акта стремились различить в его составе, не подвергая его искусственному дроблению или более тонкому разграничению позднейшей юридической мысли...»¹⁹⁹.

В дипломатике выделяются и более глубокие уровни членения актового материала. Прежде всего выделяются статьи, являющиеся законченными по мысли выражениями²⁰⁰, которые, в свою очередь делятся на сложные обороты. В оборотах могут быть вычленены так называемые «отделения», а в них – характеристики, описания и даже отдельные лексемы. В результате деление текста документа в традиционной дипломатике приобретает достаточно громоздкий вид

¹⁹⁷ Например, А. Л. Рогачевский выделяет «вводные положения» (вводную часть), к которым можно отнести, с одной стороны, протокол и, с другой стороны, такие компоненты контекста (основного текста), как *promulgatio* (публичное объявление), *narratio* (изложение обстоятельств дела) и *agens* (мотивы составления акта). См.: Рогачевский А. Л. *Кульмская грамота – памятник права Пруссии XIII в.* Санкт-Петербург, 2002. С. 118–119.

¹⁹⁸ Каштанов С. М. *Русская дипломатика...* С. 173. См. также: Каштанов С. М. *Очерки русской дипломатики...* С. 32–34.

¹⁹⁹ Лаппо-Данилевский А. С. *Очерк русской дипломатики...* 2007. С. 146.

²⁰⁰ Каштанов С. М. *Русская дипломатика...* С. 173. Иногда этот элемент текста именуется «клаузулой». См.: Лаппо-Данилевский А. С. *Очерк русской дипломатики частных актов.* Петроград, 1920. С. 135. Клаузулы четко выделяются в актах. Выделение же статей и клаузул в делопроизводственном документе вряд ли оправданно, так как степень стандартизации и юридической унификации информации в приказном документе незначительна.

(см. рис. 3.5). В классической дипломатике этим частям присваиваются в высшей степени формализованные индикаторы²⁰¹. В сущности, эти фрагменты представляют различные стилистические, синтаксические, лексические, семантические элементы и характеристики текста источника, а также содержат информацию об исторических реалиях и «исторических объектах» (имена лиц, названия учреждений, географических объектов, даты, ссылки на другие документы и т.п.), формулы (устойчивые выражения и штампы, переходящие из одного документа в другой), этикетные элементы (обращения, этикетные приложения и т.п.), описания (отступающие от стандартных формул и этикетных элементов индивидуальные выражения, характеризующие специфику описываемой ситуации). Выделение и, соответственно, разметка этих уровней основываются, главным образом, на творческой интерпретации источников²⁰².

| | | | | |
|------------------|-----------|--------------------|------------------|--------------------|
| О _{a-1} | | – Э ₁ | | – Благословение |
| | | – Э ₂ | | – от владыкы |
| О _{a-2} | – Отд. I | – Э ₁ | | – поклонъ |
| | – Отд. II | Э _{1-a} | | – от посадника |
| | | Э _{1-б} | | – Михаила |
| | | Э _{2-a} | | – и от тысяцьскаго |
| | | Э _{2-б} | | – Кондрата |
| | | Э ₃ | – X ₁ | – и от всехъ |
| | | | X ₂ | – соцьскихъ |
| | | Э ₄ | – X ₁ | – и от всехъ |
| | | | X ₂ | – старейшихъ |
| | | Э ₅ | – X ₁ | – и от всего |
| | | | X ₂ | – Новагорода |
| О _б | | – Э _{1-a} | | – къ князю |
| | | Э _{1-б} | | – Ярославу. |

Рис. 3.5. Деление текста статьи-обращения из договорной грамоты Новгорода с князем Ярославом Ярославичем, 60-е гг. XIII в.
в исследовании С. М. Каштанова
(Каштанов С. М. *Русская дипломатика*. Москва, 1988. С. 178)

²⁰¹ Формулярный анализ предусматривает высокую степень формализации и унификации источниковой информации. Такая особенность данного метода имеет и положительные стороны: формулярный анализ документов позволяет отделить индивидуальные описания, характеристики, обороты конкретного источника от формул, устойчивых штампов, этикетно-стилистических оборотов, обязательных для данной разновидности документов. В любом же отступлении индивидуального формуляра от сложившегося абстрактного формуляра содержится ценная историческая информация.

²⁰² При разметке документов было бы целесообразным обороты, элементы и характеристики выделять тегами, которые раскрывают также смысл данных структурных компонентов. Впрочем, при разметке нет препятствий и для совмещения разметки формальной и разметки, раскрывающей смысл фрагментов текста.

Наконец, на самом нижнем уровне членения актового материала находятся символы, составляющие текст.

Одной из задач изучения актов в дипломатике является «исследование содержания актов через актовый формуляр»²⁰³, поэтому в центре внимания дипломатики – структура формуляра актового источника²⁰⁴. Методика изучения формуляров актовых источников достаточно подробно разработана в дипломатике, а также в исторической лингвистике, частично использующей методику, заимствованную из дипломатики²⁰⁵. Не вызывает сомнений, что эффективное применение формулярного анализа требует углубленного изучения индивидуальных формуляров значительного числа (репрезентативной выборки) документов одной разновидности, т.е. одновременно и тщательного сравнения, вплоть до отдельных лексических единиц, внешней и внутренней формы большого числа источников. Если же при этом не ограничиваться сравнением структуры источников, а поставить цель тотального сопоставления и многоаспектного анализа всех выделенных (размеченных) структурно-семантических компонентов (фрагментов) текстов, то исследовательская задача становится исключительно трудной, возможно, даже невыполнимой без применения инструментария компьютерного источниковедения. Наверно, поэтому классическая дипломатика задачу выявления специфики отражения исторической реальности в актовом материале ставит весьма неопределенно, в самых общих чертах²⁰⁶.

Действительно, долгое время, вплоть до второй половины XX века, реализация всех стадий исследования актового материала – членение текста, связывание частей различных документов, построение конкретного и абстрактного формуляра – осуществлялась пре-

²⁰³ Каштанов С. М. *Очерки русской дипломатики...* С. 12. См. также: Пронштейн А. П. Дипломатика как научная дисциплина. *История СССР*. 1971. № 2. С. 179. При этом, как отмечает С. М. Каштанов, «[п]од внешним содержанием мы понимаем те идеи, которые выражены в источнике, под внутренним содержанием – те процессы исторического развития, которые отражаются во внешней и внутренней форме, а также во внешнем содержании акта». Каштанов С. М. *Русская дипломатика...* С. 26.

²⁰⁴ По С. М. Каштанову, условный формуляр – это «наиболее общая схема построения документов в целом»; «абстрактный формуляр» – «общая схема построения документов определенной разновидности»; «конкретный» формуляр – «схема построения определенных небольших групп документов внутри разновидности»; «индивидуальный формуляр» – «схема построения отдельно взятого текста». Каштанов С. М. *Русская дипломатика...* С. 170.

²⁰⁵ Волков С. С. *Лексика русских челобитных XVII века...* С. 21–110; Зиновьева Е. И. *Записные кабальные книги Московского государства XVI–XVII веков: Структура. Лексика. Фразеология*. Санкт-Петербург, 2000. С. 68–99.

²⁰⁶ См. раздел «Изучение внешнего и внутреннего содержания актов. Проблема исторического факта и его отражения в актовых источниках» в книге: Каштанов С. М. *Русская дипломатика...* С. 196–225.

имущественно «вручную»: карточный метод, метод составления таблиц²⁰⁷, так называемый «графически-статистический метод»²⁰⁸ и другие. Такой традиционный способ исследования имеет существенные ограничения, например, в плане расширения объема информации, заносимой в таблицу. Возможно, именно поэтому в дипломатике преобладает «техническая» сторона исследования; лишь изредка акцент делается на интерпретацию результатов, полученных в результате осуществления данных операций²⁰⁹. В целом, дипломатика вплоть до недавнего времени демонстрировала стойкий консерватизм, проявлявшийся в приверженности к методам и приемам исследования, разработанным еще в конце XIX – начале XX века. Однако и в эту вспомогательную историческую дисциплину постепенно начинают внедряться методы компьютерного источниковедения²¹⁰.

При этом технологическая составляющая формулярного анализа позволяет в полном объеме применить компьютерные технологии в изучении комплексов актового материала и одновременно поднять традиционный формулярный анализ на новый, более высокий уровень, создав возможности для перехода от аналитических к синтетическим операциям.

В отличие от традиционного дипломатического метода изучения актового материала, технология XML позволяет включить в единую систему неограниченное число источников, проводить их параллельное изучение, сопоставление и сравнение по самым различным параметрам одновременно, не говоря уже об эффективном поиске разнообразных аспектов информации, которые рассредоточены в обширном комплексе документов. Для связывания информации ряда источников, а также метаинформации об этих источниках создается XML-документ (файл), куда интегрируются источники, в которых размечен как собственно текст, так и метаинформация, соотношенная с этим текстом. В результате получается единый файл, представляющий информационную модель структуры и содержания как отдельных источников, так и их комплекса. (Далее показана принципиальная информационная модель субкомплекса смоленско-рижских актов в разметке CEI):

²⁰⁷ Ср.: Пронштейн А. П. *Методика исторического исследования*. Ростов-на-Дону, 1971. С. 325–326; Каштанов С. М. *Русская дипломатика...* С. 174–175.

²⁰⁸ Лаппо-Данилевский А. С. *Очерк русской дипломатики...* 1920. С. 145–150.

²⁰⁹ См. упомянутые выше исследования С. М. Каштанова. См. также: Волков С. С. *Лексика русских челобитных XVII века...*; Зиновьева Е. И. *Записные кабальные книги...*

²¹⁰ Представление о направлениях использования компьютерных методов в дипломатике дают конференции «Digital Diplomats». См., например: <http://www.cei.lmu.de/digdipl11/> (06.08.2012).

```

<group>
  <charter n="1">
    <abstract> ... </abstract>
    <chDesc> ... </chDesc>
    <tenor> ... </tenor>
  </charter>
  <charter n="3a">
    <abstract> ... </abstract>
    <chDesc> ... </chDesc>
    <tenor> ... </tenor>
  </charter>
  ... ..
  <charter n="6">
    <abstract> ... </abstract>
    <chDesc> ... </chDesc>
    <tenor> ... </tenor>
  </charter>
</group>

```

В рамках такого файла становятся возможными автоматическое связывание размеченных информационных блоков, поиск, классификация и сравнение информации, а также визуализация результатов, полученных в ходе исследовательских операций. При этом, как уже отмечалось, все источники, включенные в такую информационную систему, сохраняют свою целостность и «индивидуальность» – могут быть «вычленены» из системы и подвергнуться специальному изучению, т.е. историк имеет возможность в любой момент получить представление об источнике как едином целом, включая метаинформацию о нем, о его палеографических особенностях, что особенно важно в изучении именно средневековых памятников письменности.

Основная проблема, возникающая при применении технологии XML, – это четкое установление структурных и смысловых фрагментов текста, которые впоследствии будут изучаться или отдельно, или в связи с другими фрагментами текста того же источника или текстов других источников. Именно такие фрагменты текста должны быть размечены, причем обозначение элементов (тегов) должно адекватно отразить содержание выделенных фрагментов. Ведь поиск, связывание и анализ источниковой информации проводятся как бы опосредованно – на основании частных интерпретаций и констатаций, которые зафиксированы в тегах. Например (см. далее пример разметки акта № 6), определение одной части текста как указание адресата и соответствующее выделение его тегами

<addressee>...</addressee>, а другой части – как указание получателя документа с соответствующим обозначением (<recipient>...</recipient>) означает, что историк уже проинтерпретировал текст акта, определил содержание конкретных фрагментов, соотнес их с другими структурными частями акта, на основе «внеисточниковой» информации провел тонкое различие между «адресатом» и «получателем», выбрал «маркеры» для обозначения фрагментов, учитывая при этом, что и в текстах других источников фрагменты текстов, выполняющие те же функции, занимающие примерно то же место в структуре текста и несущие аналогичную смысловую нагрузку, будут обозначены теми же элементами. Таким образом, разметка является не только основой для последующего проведения исследовательских процедур, включая интерпретацию источниковой информации, но и сама является результатом углубленного изучения исторического источника.

Предпосылки для решения проблемы четкого установления размечаемых фрагментов и выбора соответствующих элементов заложены в технологии XML: допустимость расширения наборов элементов; адекватное воспроизведение многоуровневой иерархической структуры текстов; возможность изменения разметки путем ее корректировки или дополнения новой разметкой, выделяющей иные фрагменты того же текста, что позволяет при необходимости углубить, детализировать или конкретизировать представление о структуре и содержании предварительно размеченного источника. При этом альтернативная разметка органично вписывается в уже проведенную разметку. Благодаря этим технологическим возможностям, источники, представленные в XML-файле и соответствующим образом размеченные, могут использоваться многократно для решения различных исследовательских задач, а также для подготовки публикаций источников.

Вследствие детальности разработки метода формулярного анализа в дипломатике, разметка актовых источников на основе стандарта CEI, в целом, не вызывает существенных вопросов и проблем. Существующие в словаре CEI элементы и соответствующий набор тегов дают возможность детально, довольно точно представить формуляр (структуру) акта, не только западноевропейского, но и русского (древнерусского). В качестве примера можно привести схему разметки уже упоминавшегося послания архиепископа рижского (акт № 6). Отметим, что данная разметка существенно упрощена; она отражает лишь основные (не все) палеографические особенности доку-

мента²¹¹. При этом набор тегов CEI был расширен. Например, тег <refToDoc>...</refToDoc> введен для того, чтобы в тексте источника обозначить имеющиеся прямые и косвенные ссылки на другие акты, входящие в субкомплекс смоленско-рижских актов (т.е. этот элемент оказывается безусловно необходимым для конструирования семантических связей между документами субкомплекса). Другие теги, отсутствующие в CEI, – <resolution>...</resolution> и <notification>...</notification> необходимы для более точного воспроизведения смысла отдельных частей текста:

```
<charter n="6">
  ...
  <tenor>
    <protocol>
      <invocatio> † </invocatio> <salutatio>
        БЛГВЛННЄ </salutatio> <intitulatio> <issuer>
        Ѡ <roleName id="Archbishop"> мнтрофолнта
        </roleName> рнзького </issuer>
        </intitulatio> <inscriptio> ко <lb n="1"/>
        <addressee> своѣму мнлomu ' снѣвн <roleName
        id="Grand Prince"> князю велнкому </roleName>
        <lb n="2"/> <persName id="Fedor
        Rostislavich Prince of Smolensk"> феодору
        </persName> </addressee> ' н къ єго дѣтемъ ' н ко
        влѣкъ н къ <lb n="3"/> <recipient> <roleName
        id="governor of Smolensk"> намѣстьнику
        </roleName> </recipient> ' н ко всемъ боаромъ '
        </inscriptio>
    </protocol>
    <context>
      <publicatio> то бу <lb n="4"/> дн тобѣ ведоמו '
      про тую жалобу ' уто внтѣбл<sup>ѧ </sup> <lb
```

²¹¹ Разметка выделяет сравнительно крупные структурно-семантические фрагменты текста, а также некоторые исторические реалии (лица, должности, географические наименования), по которым может происходить поиск и семантическое связывание информации внутри субкомплекса актовых источников по истории отношений Смоленска и Риги.

n="5"/>НЕ ЖАЛОБИЛСЯ НА РИЖАНЫ · УНМЬ ТО ХО-
те^{ли} ö <lb n="6"/> правитнса
протнву <persName id="Helmich merchant from
Munster"> гѣлмнка </persName> · </publicatio>
<narratio> н нхъ сло ^{вѡ} та <lb
n="7"/> ково · хотелн сѧ темь словомь ѡправитн · н
ре <lb n="8"/> клн такъ предъ <roleName
id="Prince of Briansk"> княземь браньскимь
</roleName> · вы <lb n="9"/> ехалн · ѿ · ть мужь
нзъ <placeName id="Riga"> рнги </placeName> ·
н ѡубнлн юо <lb n="10"/> ловека · н оузалн · ꙗ · ть
<term id="berkovets"> бѣрковѣсковъ </term>
воск ^ѡ </narratio> <lb n="11"/>
<resolution> н нынѣ ꙗ <roleName
id="Archbishop"> мнтрофолнтъ </roleName> тако
молвю · какъ то <lb n="12"/> внтѣбланѣ неправою ·
жалобилсѧ · на риж^ѧ <lb
n="13"/> ны · н нынѣ то естъ мнѣ ведомо · аже рижанѣ
<lb n="14"/> суть в томь невнноватн ·
</resolution> <publicatio> н нынѣ ꙗ тому · дн
<lb n="15"/> влюса · аже твон <roleName
id="governor of Smolensk"> намѣстѣннкъ
</roleName> слушаеть <lb n="16"/> всякого
ѹлѣка · слова · <refToDoc> а та правда естъ промѣ
<lb n="17"/> жн васъ н насъ · кдѣ сѧ тажа поунеть
ту ко <lb n="18"/> нцѧтн </refToDoc> ·
</publicatio> <notification> н нынѣ ꙗ молюса

```

вамъ · как то мо<lb n="19"/>знтє стоатн оу
<refToDoc> тон правдѣ · н оу крѣтномь · <lb
n="20"/> уєлованнн </refToDoc> · ажє нмєть
ЖАЛОБНТНСА васъ <lb n="21"/> кто на рнжаны · нлн
<persName id="Helmich merchant from
Munster"> гѣлмнко </persName> нлн кто н<lb
n="22"/>нын · н вы шлнтє к намъ · а мы правдоу да
<sup>Ѡмы </sup> <lb n="23"/> по бѣжн правдѣ ·
</notification> <lb n="24"/>
</context>
</tenor>
</charter>

```

Следует отметить, что в акте № 6 отсутствует эсхатокол, а части преамбулы в довольно-таки вольном порядке вошли в основной текст; возможно, некоторая структурная нечеткость данного документа обусловлена тем, что перед нами концепт послания, а не его окончательный вариант.

Однако при более детальной разметке структурных блоков, особенно основного текста («контекста») – содержательной части акта, выделение отдельных клаузул и смысловых единиц в тексте, формализованное определение семантической нагрузки этих частей и формулировка тегов потребуют значительных усилий и творческого подхода, поскольку именно этот уровень разметки принципиально важен и для последующих поисковых и аналитических операций. Так, клаузулы могут считаться относительно устойчивыми и формализованными структурными частями акта, но при этом у них есть смысл, который можно изложить. Этот смысл как бы растет снизу вверх, складываясь из смысла отдельных характеристик, элементов, оборотов. Из смысла оборотов складывается смысл клаузул (статей). Далее его можно определить для более крупных блоков, пересказывая их смысл на основе смысла входящих в них клаузул. Таким образом, корректное выделение и точная маркировка (разметка) статей и более дробных структурных фрагментов (характеристик, элементов и оборотов) позволят через схему разметки адекватно воспроизвести информационное поле актового источника, а не только его формальную структуру.

Основные принципы, на которых основывается метод формулярного анализа, безусловно, применимы и к выделению структурных и семантических блоков и компонентов в текстах других видов и разновидностей документальных исторических источников. Однако возможные ограничения и специфика применения метода, например, к документам приказного делопроизводства XVII века требуют дополнительного изучения, особенно учитывая тот факт, что в российском источниковедении дается лишь очень обобщенная, приблизительная характеристика структуры приказных документов²¹².

Дипломатика, определяя свой предмет и объекты изучения, достаточно четко отграничивает себя от других специальных (вспомогательных) исторических дисциплин: «[д]ипломатика – это не какой-то один метод исследования, применимый ко всем источникам. Вернее видеть в ней совокупность многих методов и приемов в приложении к актовому материалу...»²¹³. Нельзя спорить с таким определением дипломатики, однако вряд ли можно согласиться с тезисом, что специфически дипломатический метод формулярного анализа применим только к актовому материалу²¹⁴. Необходимо учитывать генетическую общность всех документов, проявляющуюся как в общности их функций, так и во внутренней форме, упорядоченности и стандартизованности структуры²¹⁵. Сторонники обособления дипломатики по объекту изучения отмечают, что отсутствует «непроходимая грань между актами и неактами»²¹⁶, но при этом утверждают, что «применение к неактовым источникам собственно “дипломатических” приемов анализа не дает положительных результатов»²¹⁷. Вряд ли можно согласиться с этим утверждением, так как есть исследования, в которых метод формулярного анализа вполне успешно и плодотворно

²¹² Тихомиров М. Н. Приказное делопроизводство в XVII веке... С. 362–370.

²¹³ Каштанов С. М. *Русская дипломатика...* С. 27. См. также: Каштанов С. М. Дипломатика как специальная историческая дисциплина. *Вопросы истории*. 1965. № 1. С. 40; Пронштейн А. П. Дипломатика как научная дисциплина... С. 178–180 и др.

²¹⁴ «Акты – документы юридического характера, фиксирующие правовые отношения между двумя и более контрагентами. В качестве контрагентов могут выступать физические и юридические лица, а также государство». *Источниковедение: Теория. История. Метод...* С. 385.

²¹⁵ Илюшенко М. П., Кузнецова Т. В., Лившиц Я. З. *Документоведение. Документ и системы документации*. Москва, 1977. С. 9–13, 33; Шепелев Л. Е. Проблемы источниковедческого изучения делопроизводственных документов государственных учреждений XIX – начала XX вв. В кн.: *Источниковедение отечественной истории: Сб. статей*. 1975. Москва, 1976. С. 251–253, 266.

²¹⁶ Каштанов С. М. *Очерки русской дипломатики...* С. 10. Иногда отмечается, что «к актам в широком смысле слова относятся все тексты, выполняющие функции документов». См.: *Источниковедение: Теория. История. Метод...* С. 249. Конечно, это утверждение спорное.

²¹⁷ Каштанов С. М. Дипломатика как специальная историческая дисциплина... С. 41.

творно был применен, например, к изучению челобитных, которые в группу актового материала не входят²¹⁸.

По нашему мнению, при минимальной модификации (в основном, в области терминологии) метод формулярного анализа применим к изучению приказной документации, равно как и любой делопроизводственной документации вообще, поскольку приказные делопроизводственные документы имеют достаточно четкую структуру, а задачи их комплексного изучения, во многом, совпадают с задачами изучения актовых источников – то же детальное изучение внутренней формы документов, связывание и сопоставление информации документов одной разновидности в пределах естественного комплекса, отделение уникальной информации от шаблонных частей формуляра и т.д.²¹⁹

Трудности же применения метода формулярного анализа к изучению делопроизводственных документов связаны, во-первых, с менее детальной структурированностью их текстов и с меньшей степенью унифицированности и стандартизованности их языка, по сравнению с актами, с преобладанием описаний над формулами, а во-вторых, – с неразработанностью терминологического аппарата для обозначения структурных частей документов приказного делопроизводства, что затрудняет как их членение, так и идентификацию структурных блоков в этих источниках. В документоведении при характеристике формуляров документов выделяются реквизиты, которые вряд ли могут играть роль основных объектов источниковедческого анализа документальных источников²²⁰. В свою очередь, опыт лингвистов по формулярному анализу памятников русской деловой письменности следует оценить положительно²²¹. Как ни странно, наименьшую помощь в определении структурных частей документов оказали именно труды историков, в которых дается весьма приблизительное описание формуляров приказных документов. Например, отписка характеризуется следующим образом: «... Начинается она с обращения ... Обычно в отписке после титула имеются две части. Первая часть – изложение причин, почему написана отписка. Иногда

²¹⁸ Волков С. С. *Лексика русских челобитных XVII века...*

²¹⁹ См. подробнее: Иванов А. К вопросу о повышении информативной отдачи документов приказного делопроизводства... С. 24–25.

²²⁰ См.: Илюшенко М. П., Кузнецова Т. В., Лившиц Я. З. *Документоведение...* С. 33–55. См. также: Митяев К. Г. *История и организация делопроизводства в СССР...* С. 19–20 (здесь выделены некие, весьма неопределенные, «признаки документа», «важнейшие из которых используются для составления так называемого формуляра документа»).

²²¹ См.: Волков С. С. *Лексика русских челобитных XVII века...* С. 21–110; Зиновьева Е. И. *Записные кабальные книги Московского государства...* С. 65–109; Орешников А. С. О применении формализации к описанию деловых текстов... С. 150–156.

в этой части дается изложение царской грамоты, которая поступила раньше и на которую надо дать ответ. Иногда излагается поступившая челобитная. Иногда – другие причины, вызвавшие написание отписки. А дальше идет распорядительная [? – А. И., А. В.], резолютивная часть ... Заканчивается отписка чаще всего словами: “И о том нам, холопом твоим, что ты, государь, укажешь...”²²². Подобное описание дает возможность идентифицировать данную разновидность приказной документации, но не позволяет четко определить элементы формулярного анализа, т.е. разметить текст.

Условно, единственно с целью апробации инструментария компьютерного источниковедения²²³, для обозначения структурных частей документов приказного делопроизводства за основу принята терминология, используемая для членения формуляра акта, дополненная идентифицирующими обозначениями тех структурных частей приказных документов, которые отсутствуют в актах. При этом формулярный анализ фокусируется на изучении соотношения индивидуальных и абстрактных формуляров: если «индивидуальность», уникальность и специфика источника находит свое воплощение в индивидуальном формуляре, то его генетическая связь с источниками той же разновидности полнее всего раскрывается в абстрактном формуляре. Поэтому сопоставление индивидуальных и абстрактных формуляров позволяет делопроизводственные документы изучать одновременно и как уникальные, и как типические явления.

Для разметки приказных документов комплекса источников по истории Динабурга XVII века, в отличие от смоленско-рижских актов, использован стандарт ТЕІ (словарь которого, разумеется, дополнен рядом элементов), чтобы созданная в ходе разметки информационная модель делопроизводственных документов не оказалась внешне тождественной информационной модели актов. Во-вторых, схема разметки ТЕІ более гибкая, по сравнению со схемой разметки СЕІ; она, в принципе, позволяет размечать любые письменные источники и не навязывает им унифицированную модель структуры и содержания. В-третьих, использование в данном случае схемы ТЕІ позволяет не только выявить ее потенциальные возможности, но и сравнить их с возможностями схемы разметки СЕІ.

К сожалению, в рамках одной главы детально представить результаты тотального изучения всего комплекса документального ма-

²²² Тихомиров М. Н. Приказное делопроизводство в XVII веке... С. 364–365.

²²³ В задачу монографии не входит детальная разработка принципов членения документов приказного делопроизводства, равно как и терминологии для обозначения структурных частей документов.

териала по истории Динабурга XVII века нереально. Поэтому дальнейшая демонстрация возможностей связывания информации, а также способов реализации запросов к XML-документу (в следующем разделе данной главы) не является в полной степени репрезентативной. В то же время, осуществление операций по разметке отдельных источников вполне адекватно отражает основные особенности и принципы применения стандарта TEI и технологии XML, в целом, к изучению внутренней формы и, частично, содержания приказных документов.

Из всего комплекса делопроизводственных документов по истории Динабурга 1656–1667 годов выбраны три документа одной разновидности. Это отписки воевод, в которых отразилась ситуация в городе как в начальный, так и в конечный период русской власти: уже упомянутая отписка С. Л. Стрешнева (документ № 1, 1656 год); отписка бывшего борисоглебского (динабургского) воеводы Никиты Степановича Бухвостова царю Алексею Михайловичу о невозможности передать новому воеводе города Борисоглебова Ивану Никитичу Сумарокову денежную казну и хлебные запасы без расписки (документ № 46, не ранее 20 ноября 1663 года)²²⁴; отписка воеводы Ивана Никитича Сумарокова из Опочки царю Алексею Михайловичу о бедственном положении гарнизона города Борисоглебова (Динабурга), о бегстве солдат и о действиях литовских войск (документ № 50, не ранее 14 апреля 1664 года.)²²⁵.

В структуре текстов отписок воевод (равно как и большинства других документов приказного делопроизводства) можно выделить три основные части. Во-первых, это «протокол» (`<div type="protocol">`), включающий более дробные части, например, упоминание адресата (`<addressee>`), адресанта (`<sender>`, в CEI – `<issuer>`), «приветствие» (`<salutatio>`) и др. Во-вторых, это основной текст (`<div type="Main_Text">`, в CEI – `<context>`), в котором, в свою очередь, выделяются структурные части, как показано далее. В-третьих, это заключительная часть, весьма условно названная «эсхатоколом» (`<div type="eschatocol">`), в которой, как правило, запрашиваются распоряжения о дальнейших действиях.

Сопоставление формуляров отписок в рамках всего комплекса размеченных документов показало, что внутренняя форма (структура) данной разновидности делопроизводственной документации в XVII веке приобрела особую устойчивость. Ср. структуру индивидуальных формуляров трех упомянутых выше отписок:

²²⁴ Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов...* Часть 1. № 46. С. 154–155 (РГАДА. Ф. 210. Белгородский стол. Столб. 518. Л. 103).

²²⁵ Там же. № 50. С. 161–163 (РГАДА. Ф. 210. Московский стол. Столб. 365. Л. 270–271, 270 об.).

| | | |
|--|---|---|
| <doc id="doc1"> <docBody> | <doc id="doc46"> <docBody> | <doc id="doc50"> <docBody> |
| <div type="protocol"> <addressee>...</addressee> <sender>...</sender> <salutatio>...</salutatio> </div> | <div type="protocol"> <addressee>...</addressee> <sender>...</sender> <salutatio>...</salutatio> </div> | <div type="protocol"> <addressee>...</addressee> <sender>...</sender> <salutatio>...</salutatio> </div> |
| <div type="Main_Text"> <doc_ref idref="dr1">... </doc_ref> <doc_ref id="dr2">... </doc_ref> <situation>...</situation> <activities>... ...</activities> </div> | <div type="Main_Text"> <doc_ref idref="dr27">... ...</doc_ref> <activities>... ...</activities> </div> | <div type="Main_Text"> <doc_ref idref="dr31">... ...</doc_ref> <situation>...</situation> </div> |
| <div type="eschatocol">... </div> | <div type="eschatocol">... </div> | <div type="eschatocol">... </div> |
| | | <div type="Main_Text"> <situation>...</situation> </div> |
| </docBody> </doc> | </docBody> </doc> | </docBody> </doc> |

Как видим, о высокой степени унификации и стандартизации данной разновидности делопроизводственных документов свидетельствуют как устоявшаяся, по преимуществу – трехчленная, структура, так и состав основных частей документов, особенно протоколов. Ср. тексты протоколов трех отписок:

| | | |
|--|---|--|
| <doc id="doc1"> <docBody> | <doc id="doc46"> <docBody> | <doc id="doc50"> <docBody> |
| <div type="protocol"> <addressee><titleForm> Г(о)с(у)д(а)рю ц(а)рю і великому кн(я)зю <per- son idref="p1"> Алеѣю Михаил(овичю </person>, всаа Великия и Малыя и Бѣлыя Росіі самодержцу </titleForm> </addressee>, <sender>холопи твои <person idref="p2"> рег="Стрешнев, Семен Лукьянович"> Сенка Стрешнев</person> с то- варьщи</sender> <salutatio> челом бьют </salutatio>. </div> | <div type="protocol"> <addressee> <titleForm> Г(о)с(у)д(а)рю ц(а)рю і великому кн(я)зю <person idref="p1"> Алеѣю Михаиловичю </person>, всаа Великия и Малыя и Бѣлыя Росіі самодержцу </titleForm> </addressee>,<sender> холоп твои <person idref="p3"> рег="Бухвостов, Никита Степанович"> Микитка Бухвостовъ </person></sender> <salutatio> челом бьет </salutatio>. </div> | <div type="protocol"> <addressee> <titleForm> Г(о)с(у)д(а)рю ц(а)рю і великому кн(я)зю <person idref="p1"> Алеѣю Михаил(овичю </person>, всаа Великия и Малыя и Бѣлыя Росіі самодержцу </titleForm></addressee>, <sender> холопъ твои <person idref="p4"> рег="Сумароков, Иван Никитич"> Івашко Сумороков </person> </sender> <salutatio> челом бьет </salutatio>. </div> |
| </docBody> </doc> | </docBody> </doc> | </docBody> </doc> |

Для основной части текстов документов также характерна значительная степень унификации: ссылки на предыдущие документы, описание ситуаций и действий воевод.

Стандартный набор компонентов в формулярах отписок позволяет построить абстрактный формуляр отписки, отражающий в наиболее общем виде схему построения документа данной разновидности в середине XVII века:

```
<docBody>
  <div type="protocol">
    <addressee>...</addressee>
    <sender>...</sender>
    <salutatio>...</salutatio>
  </div>
  <div type="Main_Text">
    <doc_ref ref="dr1">...</doc_ref>
    <situation>...</situation>
    <activities>...</activities>
  </div>
  <div type="eschatocol">
    ...
  </div>
</docBody>
```

Хотя построение абстрактного формуляра осуществлено лишь на основании изучения индивидуальных формуляров отписок из одного локального комплекса документов, в перспективе, технология XML позволяет выборку сделать практически совпадающей со всей совокупностью документов данной разновидности, причем такая выборка может охватывать и больший промежуток времени, скажем, XVI и XVII века. Однако даже построение абстрактного формуляра документа на основе сопоставления ограниченного числа отписок вполне наглядно демонстрирует возможности технологии XML по структурированию и сравнению информации. Увеличение выборки вряд ли изменит логику исследования и сущность операций по применению данной технологии, а лишь сделает результаты исследования более точными, объективными и репрезентативными, в то же время, – более детализированными.

Однако установление структуры текстов документальных источников – это не только способ изучения эволюции внутренней формы определенной разновидности документов, но и одно из предварительных условий для выявления информационного потенциала источников, прежде всего, – скрытой информации. И дейст-

вительно, конструирование абстрактного формуляра отписки позволяет выявить отклонения в структуре конкретных отписок от общепринятого шаблона. На основе констатаций подобных отклонений появляется возможность сделать и ряд наблюдений, помогающих дать более глубокую и точную интерпретацию источников.

Наиболее близок к стандарту формуляр отписки С. Л. Стрешнева (документ № 1). Единственное отклонение – ссылка на два документа, по поводу которых составлена отписка. Интерпретировать подобное отклонение можно следующим образом: в середине – второй половине июля 1656 года поток документации (указов и наказов царя, отписок воевод) был особенно интенсивным, что было вызвано сложной, непрерывно изменяющейся оперативной ситуацией, складывавшейся в ходе «Ливонского похода» царя Алексея Михайловича (лето – осень 1656 года), стремительного продвижения его войск к Риге²²⁶. В результате на некоторые указные грамоты царя ответы давались с запаздыванием. То есть сама совокупность документов, их взаимосвязь, некоторая хаотичность и несогласованность документооборота в целом, нашедшая косвенное отражение и в данной отписке, динамически отразили как всю ситуацию, так и мельчайшие изменения в обстановке на театре военных действий²²⁷.

Индивидуальный формуляр отписки воеводы Н. С. Бухвостова (документ № 46) также близок к абстрактному формуляру отписки. Отсутствие в основной части документа описания ситуации объясняется целью составления данной отписки, которая ясно обозначена в завершающей части документа. На требование передать город и казенное имущество новому воеводе И. Н. Сумарокову воевода Н. С. Бухвостов, подробно перечислив (и, очевидно, преувеличив) свои заслуги в хозяйственно-административной деятельности (<activities>...</activities>), в заключение отмечает:

<div type="eschatocol"> А без росписке мнѣ, холопу твоему, твоеи, великого г(о)с(у)д(а)ря, казны отдат(ь) ему, Ивану, не возможно. И о томъ, великии г(о)с(у)д(а)рь, как укажешь </div>²²⁸.

²²⁶ Ср.: Дополнения к Тому III-му Дворцовых Разрядов, издаваемых по высочайшему повелению II-м Отделением Собственной Его Императорского Величества канцелярии. Санкт-Петербург, 1854. Столб. 57, 62–63.

²²⁷ См. также: Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов...* Часть 1. № 1–10, с. 67–88.

²²⁸ РГАДА. Ф. 210. Белгородский стол. Столб. 518. Л. 103.

Нетипичность эсхатокола отписки может быть интерпретирована различно. Во-первых, Н. С. Бухвостов боится, что его заслуги останутся незамеченными; это следует из самого текста документа. Во-вторых, Н. С. Бухвостов опасается, что новый воевода города постарается обвинить предшественника в злоупотреблениях реальных и мнимых, тем более, что ряд других документов комплекса дают косвенную информацию о проступках Н. С. Бухвостова²²⁹. В-третьих, Н. С. Бухвостов пытается затянуть передачу дел И. Н. Сумарокову, к которому он был назначен «в товарищи», чтобы остаться на должности, приносящей очевидные выгоды.

Наиболее существенные отклонения от абстрактного формуляра наблюдаются в отписке И. Н. Сумарокова (документ № 50). В этом документе, фактически, не три, а четыре части – отписка не завершается «эсхатоколом»:

```
<div type="eschatocol"> И по тем, г(о)с(у)д(а)рь,
Микиты Бухвостова вестям ты, великийї
г(о)с(у)д(а)рь ц(а)рь і великийї кн(я)зь Алеѣи
Михаиловичь, всеа Великиа и Малыя и Бѣлыя Росіи
самодержецъ, что мне, холопу твоему – стоят(ь) с
твоими, великого г(о)с(у)д(а)ря, ратными людьми и
промысль чинить над неприятел(ь)скими людьми, или
обърегать лиюлянских городов, Борисоглѣбова и Лю-
тина </div>230.
```

За этой традиционной заключительной частью отписки следует повторное, более драматизированное описание ситуации в городе, который оказался отрезанным от русских земель. Данное описание как бы продолжает основной текст отписки, несмотря на то что находится на ненадлежащем месте:

```
<div type="Main_Text"> <situation> А у меня, хо-
лопа твоего, в полку малолюдно: твоих, великого
г(о)с(у)д(а)ря, ратных людеі – рейтар и драгун –
с полчетверта ста ч(е)л(о)в(е)къ. Ис тѣх,
г(о)с(у)д(а)рь, половина пѣших. И апрѣля,
г(о)с(у)д(а)рь, в розных числех с Опочки рейтар и
драгун збѣжало во Псковъ і в Великийї Новгород с
пят(ь)десят ч(е)л(о)в(е)къ. И достал(ь)ные,
г(о)с(у)д(а)рь, рейтары и драгуны хотят всѣ бе-
```

²²⁹ Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов...* Часть 1. № 44, 47, 49, с. 151–152, 155–157, 159–161.

²³⁰ РГАДА. Ф. 210. Московский стол. Столб. 365. Л. 271.

жат(ь) по грѣхом, г(о)с(у)д(а)рь, страшны[м] от литовских людей </situation> </div>²³¹.

Интерпретация этого отклонения индивидуального формуляра от абстрактного не вызывает затруднений: И. Н. Сумароков (кстати, сам он находился в то время не в Динабурге, а Опочке) максимально сгущает краски в описании ситуации, что, возможно, и отражает реальное положение дел (о тяжелой, драматичной ситуации в городе свидетельствуют и другие документы комплекса), но также возможно и то, что воевода пытается обосновать свое нежелание покинуть родные места, отправиться к новому месту службы в осажденный город, терпеть лишения и подвергаться опасностям и превратностям войны.

Как видим, сопоставление индивидуальных формуляров документов между собой при посредстве абстрактного формуляра данной разновидности документов позволяет не только изучить структуру отдельных разновидностей делопроизводственных документов, но и предложить интерпретацию, в основе которой косвенная (скрытая) информация источников. В результате формализованная репрезентация структуры исторических источников позволяет одновременно раскрыть и некоторые (далеко не все) особенности отражения исторической реальности в приказных документах, а также вплотную подойти к воссозданию исторической действительности и реконструкции исторических фактов. Бесспорно, потенциальные возможности формулярного анализа становятся особенно убедительными при применении технологии XML, позволяющей проводить системную обработку и сравнение больших массивов источников.

Те же возможности технологии XML были выявлены (и тем самым подтверждены) в ходе изучения актового материала – небольшого субкомплекса полоцких грамот XV века из комплекса «Moscovitica–Ruthenica»²³². В данном случае основное внимание было также уделено построению конкретного формуляра полоцких грамот 60-х – 70-х годов XV века, который существенно не отклоняется от абстрактного формуляра актовых источников (см. ниже).

Протокол

Богословие (invocatio)

Адресат (inscriptio)

²³¹ РГАДА. Ф. 210. Московский стол. Столб. 365. Л. 271.

²³² См.: Ivanovs, A., Varfolomejev, A. Editing and Exploratory Analysis of Medieval Documents by Means of XML Technologies... P. 159; Варфоломеев А. Г., Иванов А. С. XML-технологии в анализе информации полнотекстовых баз данных...

Адресант (intitulatio)
Приветствие (salutatio)

Основной текст / контекст

Оборот 1 (Объявление)
Оборот 2 (Распоряжение)
Оборот 3 (Ретроспекция)
Оборот 4 (Просьба, обращение, требование адресанта)
Оборот 5 (Действия адресанта)
Оборот 6 (Описание текущей ситуации)
Оборот 7 (Предложение, требование адресанта адресату)
Оборот 8 (Выполнение адресатом распоряжения)

Эсхатокол

Место и время написания (datum)

Следует отметить, что цель изучения полоцких грамот в исследованиях, проведенных еще в 2005–2006 годах, была чисто техническая: применение технологии XML для воссоздания конкретного формуляра полоцкой грамоты и выявление закономерности в следовании смысловых фрагментов в основной части. При этом оценивались возможности специализированного редактора в разметке компонентов основной части актов, а также возможности XSLT-преобразований в визуализации структуры документов, получении статистических характеристик текстов, выявлении похожих структурных компонентов в основных частях документов, определении частоты встречаемости и обычной локализации логических фрагментов грамот. Задача целенаправленного выявления содержательных аспектов источников в ходе осуществления данных операций не ставилась. Данная задача решалась при изучении комплекса динабургских документов и субкомплекса смоленско-рижских актов²³³.

Как уже отмечалось выше, информационный потенциал документальных исторических источников частично раскрывается уже в ходе формулярного анализа. Однако технология XML, а именно стандарт TEI, может одновременно использоваться и для целенаправленного выделения в текстах «исторических объектов», смысловых фрагментов, описаний ситуаций и действий, предпринятых должностными лицами, а также прочих исторических реалий²³⁴.

²³³ О выявлении и семантическом связывании информации смоленско-рижских актов см. в главе четвертой настоящей монографии.

²³⁴ Весьма поверхностно и при этом частично такая возможность XML была показана (но не прокомментирована) выше, на примере разметки акта № 6 на основе стандарта CEI.

Здесь также необходимы дополнительные теги, например: <titleForm> – титул адресата – царя; <org> – название упоминаемого учреждения – приказа; <subject> – изложение содержания упоминаемого или цитируемого в источнике документа; <situation> – описание ситуации; <activities> – описание действий, предпринятых адресантом и т.п. Количество дополнительных введенных тегов напрямую зависит конкретных задач и от степени детализации источниковедческого анализа, от параметров и признаков, по которым исследователь предполагает проводить семантическое связывание информации источников, входящих в комплекс. Причем степень детализации можно увеличить, так как формирующаяся система разметки, как уже отмечалось, в любом случае будет оставаться «открытой», позволяющей в уже размеченном документе выделить более мелкие или, напротив, более крупные смысловые фрагменты. Приведенные выше обозначения тегов даются условно, чтобы, по возможности, приблизить их обозначение к логике и стилю, принятому в системе кодирования текстов ТЕІ. В качестве примера углубленной разметки, можно привести фрагмент текста отписки С. Л. Стрешнева (документ № 1):

```

<doc id="doc1">
...
<docBody>
...
  <situation>
    А рекою, <titleForm> г(о)с(у)д(а)рь
  </titleForm>, <placeName ref="pl2"> Двиною
  </placeName>, твоим г(о)с(у)д(а)р(е)вым служи-
  лым людем ныне стругами мимо <placeName
  ref="pl1"> города </placeName> проити не мочно,
  для тогw, что s <placeName ref="pl1"> города
  </placeName> и против <placeName ref="pl1"> го-
  рода </placeName> на друго[и] <pb n="675"
  type="recto"/> сторонъ <placeName ref="pl2">
  реки Двины </placeName> на берегу учинены были
  шанцы, и на <placeName ref="pl2"> реку на Двину
  </placeName> з берег[a] стреляют ис полуторных
  пыщалеи беспрестанно
  </situation>.
  <activities> И мы, холопи твои, м(и)л(о)стию
  б(о)жиею из-sa <placeName ref="pl2"> Двины реки
  </placeName> из шанецъ сбили. И пр[о]мышляемъ
  мы, холопи твои, чтоб струг[и] с твоими

```

```

г(о)с(у)д(а)р(е)выми служилыми людьми мимо
<placeName ref="p11"> города </placeName> про-
пустить вскоре
...
</activities>
...
</docBody>
</doc>

```

Отдельный вопрос – возможность «русифицированного» обозначения тегов для содержательной детальной разметки текстов средневековых русских источников. Разумеется, смысл элементов разметки и, соответственно, тегов может быть выражен и на русском языке, причем можно использовать и кириллицу. Не вызывает сомнений, что такой подход позволяет более четко выявить как структуру документа, так и смысловую нагрузку отдельных фрагментов – описаний –, содержание которых будет выражено на языке оригинала, т.е. с меньшей потерей смысла (хотя частичная потеря смысла неизбежна и в этом случае, так как разметка – унифицированная и несколько упрощенная модель содержания источника). Обозначение тегов на русском языке облегчает и сам процесс разметки и в дальнейшем позволяет более гибко проводить аналитические и синтетические операции. Как пример, можно привести общую схему разметки отписки борисоглебского воеводы Василия Григорьевича Фефилаьева царю Алексею Михайловичу о приезде курмышан к городу Борисоглебскому (не ранее 9 августа 1656 года, документ № 7)²³⁵.

```

<вводная_часть>
  <адресат>
    Г(о)с(у)д(а)рю ц(а)рю і великому кн(я)сю Алеѣю
    Михайловичю| всеа Великия и Малыа и Бѣлыя Росіи
    самодержцу|
  </адресат>
  <адресант>
    холопъ твои Вас[ь]ка Фефилат(ь)евъ
  </адресант>
  <приветствие>
    челом бьет.
  </приветствие>
</вводная_часть>

```

²³⁵ Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов...* Часть 1. № 7, с. 81–82 (РГАДА. Ф. 210. Московский стол. Столб. 285. Л. 91а–91б).

```

<основная_часть>
  <преамбула>
    <изложение_документа>
      <дата>
        В н(ы)нешнем, | г(о)с(у)д(а)рь, во [7]164-
        м году августа въ 7 день
      </дата>
      <адресат>
        писал ко мнѣ, | холопу твоему,
      </адресат>
      <адресант>
        Иван Прончищев,
      </адресант>
      <содержание>
        что курмышана, дворяна | и дѣти боярские,
        ѣдут из моровых мѣсть.
      </содержание>
    </изложение_документа>
  </преамбула>
  <отчет_о_действиях>
    <действие номер="1">
      И я, холоп твоѣ, | поставил заставу того ж
      числа.
    </действие>
  </отчет_о_действиях>
  <изложение_обстоятельств>
    И августа ж, г(о)с(у)д(а)рь, въ 9 день | курмышаня
    приѣзжали к заставе.
  </изложение_обстоятельств>
  <отчет_о_действиях>
    <действие номер="2">
      И у заставе, г(о)с(у)д(а)рь, тѣ | курмышаня че-
      рез шгонь спрашиваны.
    </действие>
    <действие номер="3">
      А что в роспро | се, г(о)с(у)д(а)рь, шни,
      курмышаня, сказали, и тое, г(о)с(у)д(а)рь,
      роспис(ь) в Бо | рисогльбском городе велѣл я пе-
      реписати. И тое роспис(ь) послал я, | холоп
      твои, к тебѣ, г(о)с(у)д(а)рю.
    </действие>
  <запрос_указаний>

```

```

И по тои ѿписке указу мнѣ, холо|пу твоему, нѣтъ.
</запрос_указаний>
</отчет_о_действиях>
<изложение_документа вид="роспись">
    А в росписи, г(о)с(у)д(а)рь, их в прежней на-
    пи|сано,
    <сведения_об_отправке>
    что послана под ѿпискою к тебѣ,
    г(о)с(у)д(а)рю:
    </сведения_об_отправке>
    <содержание>
    Ъхали де| ѿни ис Курмышу к Москве. И, не
    доѣхав Москвы, посы|лали челобитчика к Моск-
    ве. И их де к Москвы не пустили, | и обрубил
    в лѣсу за восемь версть от Москвы. А на двух
    де| человек в тѣ поры были язвы. И с Москвы
    при|сылали ихъ ѿсматриват(ь). И, досмотря,
    тѣх двух человек| тут в обрубе покинули. А имь
    де здоровым всѣмъ| велено ѡхат(ь) на твою
    г(о)с(у)д(а)р(е)ву службу в полки. И от Моск-
    вы де|| поѣдучи, на них, на курмышанях, ни на
    комь язвы нѣт.| А с ними де ѡхал Василей Лу-
    ковъ. И на нем на самомъ| и на человекѣ ево
    язвы были. А тово де ѿни не| вѣдают - тѣ ли
    язвы или нѣтъ. И ѿн де стоит поста|влен в По-
    лоцке.
    </содержание>
</изложение_документа>
</основная_часть>
<заключительная_часть>
    <запрос_указаний>
    И ѿ том, г(о)с(у)д(а)рь, что мнѣ холо|пу своему
    укажеш(ь).
    </запрос_указаний>
</заключительная_часть>

```

В более углубленной разметке большая часть тегов также снабжается атрибутами “id” и “ref”, позволяющими идентифицировать топонимы, людей, учреждения и т.п., а также проводить поиск упоминаний о данных реалиях в рамках как одного документа, так и

всего комплекса в целом. Разница между этими атрибутами в том, что значения “id” должны быть уникальными в рамках одного комплекса документов, а значения “ref” – ссылки на идентификатор – могут повторяться. Этот более глубокий уровень разметки органично включается («вкладывается») в проведенную ранее разметку метаинформации и палеографических особенностей источника. При этом, как отмечалось выше, возможна и дальнейшая детализация разметки – выделение стилистических, синтаксических, лексических элементов и характеристик текста источника²³⁶.

Несмотря на очевидное удобство и «естественность» разметки посредством тегов, обозначенных на русском языке²³⁷, все-таки придется признать, что эти преимущества проявляются только при источниковедческом изучении локального комплекса средневековых русских документов. В случае же интеграции размеченного комплекса в открытую семантическую сеть, в которой могут быть представлены разнообразные комплексы источников, к тому же написанные на разных языках (например, комплекс «Moscovitica–Ruthenica»), преимущества могут превратиться в недостатки: такие теги могут существенно затруднить связывание информации. Поэтому в качестве перспективного инструмента разметки следует признать английский вариант схемы TEI (для актового материала – схемы CEI).

Следует подчеркнуть, что процесс разметки имеет последовательный, стадийный характер, что особенно важно при совместной и распределенной работе группы исследователей, изучающих один комплекс источников²³⁸. Сначала создается файл с размеченной метаинформацией и первичной (палеографической) разметкой текста. На основе этого файла создается связанный с ним XML-документ с текстом на современных языках (это особенно актуально в изучении комплекса «Moscovitica–Ruthenica», в который входят документы на различных языках). Затем поверх первичной XML-разметки исследователи создают свои разметки, выделяя в тексте логические фрагменты, упоминания персоналий и географических

²³⁶ Хотя проблемы разметки текстов рассматривались применительно к документальным историческим источникам, по мнению авторов, те же самые принципы разметки в полной мере применимы и к нарративным источникам по истории России. Однако здесь выделение структурных частей текста будет иметь второстепенное значение, больший акцент будет делаться на выделении чисто смысловых аспектов текста, описаний и исторических реалий.

²³⁷ Действительно, несколько «неестественно» выглядит внедрение в русский текст англоязычных маркеров, которые несут семантическую нагрузку, «чуждую» оригинальным источникам.

²³⁸ Такой подход был апробирован в ходе создания и совершенствования системы «Источник» в Петрозаводском государственном университете. См. <http://istochnik.karelia.ru> (09.12.2012). О распределенной работе сетевого сообщества см. подробнее в разделе «Виртуальные среды для научных исследований и «communities of practice» четвертой главы.

названий, лингвистические элементы и т.д. Наконец, оформляются результаты исследования и дается описание методики в специальном XML-формате, делая их проверяемыми и воспроизводимыми для других исследователей.

XML как технология агрегирования, связывания, анализа и репрезентации источниковой информации

Применение технологии XML требует не только проведения многоуровневой разметки текстов источников (в сущности, это задача историка или источниковеда), но и создания (или применения уже имеющегося) эффективного инструментария, приспособленного для многоуровневого и многоаспектного перекрестного сравнения информации источников, входящих в один комплекс, связывания и агрегирования информации источников по различным параметрам и признакам, верификации свидетельств отдельных источников, восполнения пробелов в источниковой информации, моделирования аспектов реальности, не нашедших полноценного воплощения ни в одном из источников, в итоге – «сводного» описания фрагментов реальности, которое с полным основанием можно назвать исторической реконструкцией, являющейся результатом источниковедческого синтеза. Технологически такое всестороннее и многоаспектное изучение источников предполагает не только количественный анализ их информации, основанный на замене текстов или их частей математическими объектами (векторами, графами) и операциях с этими объектами, но и качественный анализ, основанный на выборе формы представления текстов, удобной для обнаружения исследователем скрытых закономерностей, проведения других операций источниковедческой критики. Возможности для этого предоставляет технология XML, в некоторых случаях – в сочетании с другими компьютерными технологиями. Это означает, что ценность XML состоит в его универсальности, благодаря чему различные XML-форматы можно использовать в рамках одного проекта для различных целей, произвольно их комбинируя и при этом всегда оставаясь в рамках стандарта.

Технология XML может быть реализована с помощью стандартных и общедоступных программных средств. Для создания XML-документа не требуется ничего, кроме текстового редактора. При этом данные из XML-файла могут читаться различными программами, или преобразовываться в документы других форматов.

Для этого используются языки XSLT²³⁹ и XSL-FO²⁴⁰, также созданные на основе XML. Предназначение их в том, чтобы преобразовывать одно XML-дерево в другое, а также переводить XML-документы в документы формата HTML, PDF или RTF, представляющие документ на экране монитора или формирующие оригинал-макет для издательства. Преобразования могут охватывать не весь XML-документ, а только некоторую его часть, удовлетворяющую тем или иным условиям.

Так, с помощью преобразований, описанных на языке XSLT, на основе первичного XML-документа можно создавать XML-файлы, содержащие разметку только одного уровня. Такие файлы будут служить основой для процесса анализа (интерпретации) текста, который должен заключаться в ручном или автоматическом распределении объектов текста по различным типам, определяемым исследователем, в подсчете частоты встречаемости объектов (признаков) тех или иных типов, в установлении связи между фрагментами разных документов, в комментировании или переводе этих фрагментов. Результаты такого анализа, проведенные для одного или нескольких текстов, также можно оформить с помощью разметки XML, что позволит исследователям обмениваться друг с другом как своими выводами, так и научными методиками.

Поиск нужной информации, а также ее визуализация на экране в удобной для анализа форме²⁴¹ может осуществляться с помощью создания различных XSLT-запросов к XML-документу. Помимо XSLT, запросы могут быть сформированы с помощью и других высокоуровневых средств, таких как XQuery, или написаны в виде серверных скриптов на языках PHP, Java и т.д., напрямую извлекающих информацию из XML-деревьев. В любом случае, с точки зрения хранения, запросы представляют собой текстовые структурированные файлы, которые могут быть отображены в виде XML. Результаты таких запросов, как правило, тоже имеют XML-форму.

²³⁹ Язык XSLT (Extensible Stylesheets Language for Transformations) предназначен для описания шаблонов (templates) – правил преобразования некоторых элементов XML-документа в текст на языке HTML. См. подробнее: Холзнер С. *XSLT*. Санкт-Петербург, 2002.

²⁴⁰ Grigoriev, N. XSL Formatting Objects Tutorial. <http://www.renderx.com/tutorial.html> (19.11.2012).

²⁴¹ Следует отметить, что визуализацию можно использовать в качестве инструментального средства формирования новой информации и знаний. Людям, как правило, намного проще оперировать визуальными образами при формировании своих умозаключений, при построении взаимосвязей между объектами или упорядочивании их, чем использовать для этого таблицы или размеченные тексты.

В ходе реализации запросов к XML-документам историк и источниковед должны получить полный набор легко идентифицируемых фрагментов текстов источников и метаинформации о них (разумеется, если в ходе разметки все фрагменты были идентифицированы и обозначены надлежащим образом). Как отмечалось выше, такое связывание информации источников позволяет предложить «сводное» описание фрагментов реальности по многим параметрам – действующие лица, место действия, содержание – одновременно. Разумеется, сводное описание аспектов исторической реальности нуждается в творческой интерпретации, что дает возможность историку или источниковеду убедительно продемонстрировать свои профессиональные возможности интерпретатора и рассказчика о событиях прошлого.

Возможности реализации XSLT-запросов были изучены в ходе источниковедческого изучения комплекса приказных документов по истории Динабурга XVII века. Например, для вывода содержимого трех отписок, разметка которых была подробно рассмотрена в предыдущем разделе (документы № 1, 46, 50), в виде трех колонок в окне Web-браузера Internet Explorer (начиная с версии 6.0) был создан текстовый файл с расширением “.xsl”, в котором находится такой шаблон:

```
<xsl:template match="/">
<html><body><table border="2" bgcolor="yellow"
align="center" width="100%">
<tr valign="top">
<td><xsl:value-of select="docGroup/
doc[@id='doc1']/docBody"/></td>
<td>
<xsl:value-of select="docGroup/
doc[@id='doc46']/docBody"/></td>
<td>
<xsl:value-of select="docGroup/
doc[@id='doc50']/docBody"/></td>
</tr></table></body></html>
</xsl:template>
```

Здесь используются два тега языка XSLT: `<xsl:template>` содержит шаблон (их может быть несколько в одном запросе), атрибут “match” указывает, к каким элементам применяется шаблон (значение “/” обозначает корневой элемент XML-документа). Второй тег – `<xsl:value-of>` – позволяет извлечь из элемента, указанного в атрибуте “select”, текстовое содержимое. Как видно, это текстовое со-

держимое обрамляется HTML-тегами <td>...</td>, которые определяют ячейку таблицы.

Если возникает необходимость раскрасить различные логические части формуляров документов в разные цвета для облегчения процесса визуального анализа формуляра, то этого можно достичь, описав еще несколько шаблонов, например, шаблон

```
<xsl:template match="addressee">
  <span style="color:red"><xsl:value-of
  select="." /></span>
</xsl:template>
```

окрашивающих текстовое содержимое элементов <addressee> в красный цвет, и заменив в шаблоне для корневого элемента теги <xsl:value-of /> на теги <xsl:apply-templates /> с теми же значениями атрибута "select" (см. рис. 3.6).

| doc1 | doc46 | doc50 |
|---|---|---|
| Г[о]с[у]д[а]рю ц[а]рю ї великому кн[я]зю Алексію Михаиловичю, всеа Великия и Малыя и Бѣлыя Росіи самодержцу, холопи твои Сенка Стрешнев с товариши челом бьют. | Г[о]с[у]д[а]рю ц[а]рю ї великому кн[я]зю Алексію Михаиловичю, всеа Великия и Малыя и Бѣлыя Росіи самодержцу, холоп твои Микитка Бухвостовъ челом бьет. | Г[о]с[у]д[а]рю ц[а]рю ї великому кн[я]зю Алексію Михаиловичю, всеа Великия и Малыя и Бѣлыя Росіи самодержцу, холопъ твои Івашко Сумороков челом бьет. |
| Июля, г[о]с[у]д[а]рь, въ 1 де[нь] в твои, г[о]с[у]д[а]р[е]в[е] ц[а]р[е]в[е] ї | Н[ы]нешняго, г[о]с[у]д[а]рь, [7] 172-г[о] году ноября въ 20 де[нь] | В н[ы]нешнем, г[о]с[у]д[а]рь, во [7]172-м году апрѣля въ 14 де[нь] писал ко мнѣ, |
| <i>д[а]р[е]в[ь]м служилым людем н[ы]не стругами мимо города проити не мочно, для того, что з города и против города на друго[й] сторонѣ реки Двина на берегу учинены были шанцы, и на рехуна Двина уз берег[а] стреляют ис полугорных гвщалеи бестрестанно. И мы, холопи твои, м[и]л[о]стию Б[о]жию</i> | прииму учинил тебе, великому г[о]с[у]д[а]рю, прибыл[ь] с неприятел[ь]ской стороны хлѣбных запасов безденежно четыреста девяносто одна пора с полупурою ржі, шестьсот пят[ь] | Борисоглѣбов. <i>А какъ я, холопъ твои, из Борисоглѣбова пошел, и по то, г[о]с[у]д[а]рь, число в Борисоглѣбова было твоих, великого г[о]с[у]д[а]ря, ратных людеи всякого чину с триста ч[е]л[о]в[е]къ. А н[ы]не, г[о]с[у]д[а]рь, и тех половина бол[ь]ших. А твоим, великого г[о]с[у]д[а]ря,</i> |

Рис. 3.6. Сравнение формуляров трех отписок (визуализация).

Слева: Отписка С. Л. Стрешнева (документ № 1). РГАДА. Ф. 210. Московский стол. Столб. 270. Л. 674–675. Посередине: Отписка Н. С. Бухвостова. РГАДА. Ф. 210. Белгородский стол. Столб. 518. Л. 103. Справа: Отписка И. Н. Сумарокова из Опочки. РГАДА. Ф. 210. Московский стол. Столб. 365. Л. 270–271

На рисунке 3.6 представлены выдержки из начала и середины документов, причем для наглядности цветовая визуализация логических фрагментов заменена на оформление их с помощью шрифтов и оттенков серого цвета (элемент <addressee> отображается серым

цветом, элемент `<sender>` – жирным шрифтом, `<salutatio>` и `<situation>` – курсивом разного размера, `<activities>` – шрифтом Courier). Данный пример визуализации особенностей формуляров трех документов наглядно демонстрирует возможности технологии XML в сравнительном анализе индивидуальных формуляров.

Еще одну возможность визуализации и, одновременно, связывания информации предоставляет технология SVG, основанная на XML-описании векторной графики²⁴². Преимущества использования SVG-визуализации заключаются в том, что алгоритмы визуализации могут быть технически реализованы в виде XSLT-преобразований из XML в SVG, а значит, открыты для модификаций.

В ходе глубокого и всестороннего источниковедческого изучения исторических источников исследователю зачастую необходимы не только их тексты, снабженные полноценной информацией о палеографических особенностях оригиналов, но и сами оригиналы источников, их фотокопии или растровые изображения, например, для проверки правильности передачи текстов в ходе транскрипции источников и создания XML-документа. Поэтому, наряду с XML-файлами, в электронной библиотеке исторических источников обязательно должны храниться их растровые изображения. Технологии SVG позволят связать XML-файл с изображением-оригиналом. Для этого создается дополнительный SVG-файл, содержащий в себе объекты-прямоугольники, отмечающие расположение словоформ, надстрочного текста и диакритических знаков на изображении документа. Связь прямоугольников с соответствующими элементами XML-разметки реализуется с помощью атрибутов-идентификаторов. Прямоугольники соединяются в группы объектов, обозначающие расположение строк и предложений. Впоследствии, при добавлении в XML-файл логической и семантической разметки, она также может быть визуализирована на изображении-оригинале, поскольку любой фрагмент текста состоит из набора словоформ, местоположение которых определено в SVG-файле.

SVG-связывание XML-разметки текста с изображением исторического источника позволяет решить две важные задачи. Во-первых, на стадии транскрипции или проверки результатов воспроизведения текста источника заменить ручное распознавание всего текста в целом распознаванием значения отдельных элементов изо-

²⁴² Scalable Vector Graphics. <http://www.w3.org/Graphics/SVG/> (09.12.2012). См. также: Варфоломеев А. Г., Кравцов И. В., Филатов В. О. SVG-визуализация в цифровых библиотеках рукописных документов...

бражения – словоформ, которое поддается автоматизации. Для этого распознанные словоформы записываются не только в XML-файл, но и в базу данных словоформ, и затем, при вводе очередной словоформы, самые подходящие из них появляются в виде динамически изменяющегося списка подсказок, реализованного с помощью технологии AJAX²⁴³, что значительно облегчает трудоемкий процесс распознавания текста. Во-вторых, SVG-связывание позволяет визуализировать результаты выполнения запросов к XML-документам, используя для этого их изображения. Речь идет о запросах на поиск определенных словоформ, персоналий, географических названий, логических частей документов. Вывод результатов таких запросов в виде соответствующих фрагментов изображений-оригиналов обеспечивает дополнительные удобства в работе исследователя, имитируя работу с реальными документами в архиве.

Апробация данного подхода проводилась на источниках, входящих в субкомплекс полоцких грамот из комплекса «Moscowitica–Ruthenica». На рис. 3.7 представлен фрагмент одного из документов данного субкомплекса с выделенными на нем двумя прямоугольными областями, которые соответствуют двум словоформам.

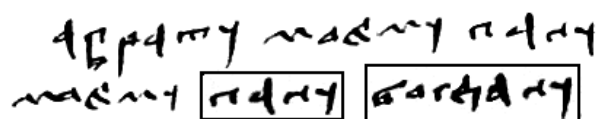


Рис. 3.7. Пример выделения областей на растровом изображении источника для SVG-связывания с XML-файлом. [1471 г.] августа 30.

Грамота полоцкого воеводы Олехны Судимонтовича.

ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 7. Л. 2

Фрагмент соответствующего SVG-файла для рис. 3.7 выглядит следующим образом:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.0//EN"
"http://www.w3.org/TR/SVG/DTD/svg10.dtd">
<svg version="1.0"
xmlns="http://www.w3.org/2000/svg"
xmlns:xlink="http://www.w3.org/1999/xlink"
zoomAndPan="enable"
width="1280" height="1280">
```

²⁴³ Крейн Д., Паскарелло Э., Джеймс Д. *Ajax в действии*. Москва, 2006.

```

<image x="0" y="0" width="900" height="997"
xlink:href="005_MIN.JPG">
<g id="markup" stroke="black" stroke-width="0.5"
fill="black"
fill-opacity="0.1">
....
  <rect id="r104" x="322" y="296" width="95"
height="27">
  <rect id="r105" x="420" y="299" width="111"
height="26">
....
</g>
</svg>

```

В приведенном SVG-коде теги `<rect>` задают прямоугольные области, тег `<g>` – группу прямоугольников, которая может быть строкой или законченной фразой. Тег `<image>` используется для вывода растрового изображения исходного документа.

Данному SVG-файлу соответствуют два XML-файла, содержащие точное воспроизведение оригинального древнерусского текста и текст в современной транскрипции. В обоих файлах вместо прямоугольных областей используются тэги словоформ `<w>`, имеющие такое же значение атрибута “id”, что и прямоугольники. Так, например, выглядят две выделенные на рис. 3.7 словоформы в XML-файле с адаптированным написанием:

```

<w id="r104">пану</w>
<w id="r105">Богдану</w>

```

Так как все способы представления одной словоформы связаны одним и тем же идентификатором, исследователи всегда могут перейти от одного способа к другому²⁴⁴. Выделять логические фрагменты, сравнивать структуру исторических источников удобнее с использованием адаптированного текста, но сами тексты часто тре-

²⁴⁴ Следует отметить, что упоминаемые в главе 4 система TextGrid и редактор EditMom также содержат в себе средства связывания изображения документа с распознанным текстом в XML-формате, что подтверждает актуальность и общепризнанность данного принципа. См. подробнее: Aschenbrenner, A., Gietz, P., Küster, M. W., Ludwig, Ch., Neuroth, H. TextGrid – a Modular Platform for Collaborative Textual Editing. In: *Proceedings of the International Workshop on Digital Library Goes e-Science (DLSci06)*. Alicante, 2006. Pp. 27–36. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/ecdl06_textgrid.pdf (06.08.2012); Küster, M. W., Ludwig, Ch., Aschenbrenner, A. TextGrid as a Digital Ecosystem. In: *IEEE DEST. 2007, 21–23 February, Cairns*. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/DigitalEcosystem07_CameraReady-1.pdf (07.08.2012); EditMom (Monasterium’s XML Editor). <http://www.mom-wiki.uni-koeln.de/MOM-CA1/EditMOM/> (09.12.2012).

буется отображать в форме, максимально близкой к оригиналу, т.е. в палеографической транскрипции.

Рассмотренный выше способ визуализации и связывания информации исключительно важен и полезен для дипломатически и палеографически точной электронной репрезентации источника, которая служит надежной основой для проведения источниковедческой критики. В свою очередь, для эффективного проведения аналитических и синтетических операций могут быть использованы и другие современные средства визуализации источниковой информации.

Прежде всего, это так называемые «облака тегов» (*tag clouds*), получаемые при запросах на поиск определенных элементов в XML-документах, – текстовая визуализация семантических категорий, отражающая накопленные знания о текстах в выбранном разрезе. Визуализация строится с помощью применения различного цвета и размера, а также взаимного расположения категорий. Облака состоят из SVG-прямоугольников и показывают местоположение интересующих исследователя элементов в историческом источнике. С помощью такой визуализации можно сравнивать, например, структуру различных актов источников между собой. Во-вторых, это построение графа документа, который показывает порядок следования логических и семантических элементов (фрагментов текста) и установленные исследователем связи между элементами внутри XML-документа. В-третьих, можно упомянуть граф комплекса источников, показывающий связи между источниками – как явные (например, прямые ссылки одного источника на другой), так и неявные (упоминание одних и тех же событий, явлений, объектов, лиц). В-четвертых, это так называемая «временная линейка» (*history timeline*) для размещения источников комплекса на оси (шкале) времени в соответствии с их абсолютной или относительной датировкой.

Для построения визуализации «временная линейка» необходимо с помощью запроса получить XML-файл, содержащий даты создания источников, входящих в комплекс, для которых строится визуализация. Также он должен содержать элемент, определяющий начало и конец периода, для которого строится визуализация. Такой файл может выглядеть так:

```
<report>
  <period>
    <start>1657</start>
    <end>1665</end>
  </period>
  <doc doc_id="d146" date="13.08.1658" type="1">
```

```

...
<doc doc_id="d148" date="17.02.1662" type="2">
...
</report>

```

Имея такой документ, достаточно несложно с помощью XSLT-преобразования построить SVG-документ, воспроизведение которого в окне браузера даст необходимую визуализацию (см. рис. 3.8).

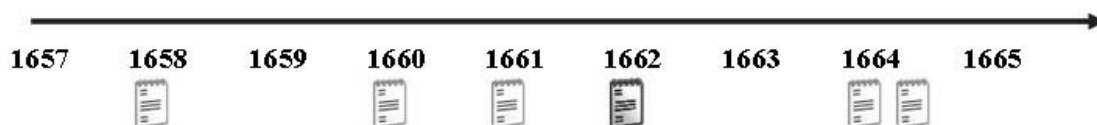


Рис. 3.8. Временная линейка.

Комплекс приказных документов XVII века по истории Динабурга

Построенные SVG-схемы можно преобразовывать в XML-файл, отражающий результаты исследования. В итоге получают наглядные информационные модели содержания отдельных исторических источников, на основе которых можно проводить дальнейшие исследовательские операции, в том числе и с совокупностями взаимосвязанных источников²⁴⁵.

Помимо поиска, связывания и визуализации информации технология XML предоставляет возможность обобщать и совершенствовать технику источниковедческого исследования. Так, использование различных XML-форматов открывает возможность соединять тексты источников, размеченные разными способами, с другими текстами, содержащими в себе описание методики и самого процесса исследования источников, а также сделанных в итоге выводов. Эти описания также могут быть структурированы с помощью XML-разметки, а значит, доступны для различных запросов.

В ходе изучения применимости компьютерных методов к источниковедческим исследованиям был разработан один из вариантов схемы разметки для описания методики и результатов исследования комплекса отписок, посланных русскими воеводами из Динабурга в Москву в ответ на царские грамоты²⁴⁶. XML-файл, создаваемый в информационной системе в соответствии с предлагаемой схемой разметки, состоит из трех основных частей: постановки за-

²⁴⁵ Более подробно вопросы репрезентации семантики текстов источников в семантической сети рассмотрены в четвертой главе монографии.

²⁴⁶ Варфоломеев А. Г., Кравцов И. В. Использование технологии XML для публикации методики и результатов исследования...

дачи, описания метода исследования и формулировки результатов. При формировании описаний задачи и метода ее решения помимо словесных формулировок необходимо также отнести задачу и метод к одному из классов. Списки классов могут расширяться; они должны храниться в отдельном XML-файле, общем для всех пользователей. Кроме того, в той же информационной системе имеется возможность дать необходимые ссылки на печатные и электронные публикации, в том числе и на описания других исследований.

Формулировка результатов исследования включает перечень всех источников, которые наличествуют либо упоминаются в данном комплексе. Для каждой отписки указаны: дата создания, ссылки на «документ-предшественник», а также на те документы, в которых упоминается данный текст. Для несохранившихся источников приведено возможное (гипотетически реконструированное) их содержание. Кроме того, для всех сведений об исторических источниках, составляющих результат исследования, есть возможность указать источник информации и степень ее надежности и достоверности.

Сохранение и доступность методик и результатов ранее проведенных исследований не только способствуют совершенствованию инструментария компьютерного источниковедения, но и в перспективе смогут оказать существенное влияние на развитие методологии исторического и гуманитарного исследования в целом. Доступность описаний этапов исследования позволит научному сообществу проверять обоснованность выводов, поскольку сами источники тоже будут доступны, таким образом, принцип воспроизводимости результатов эксперимента, принятый в естественных науках, окажется реально применимым и в исторической науке. С другой стороны, необходимость для самого исследователя структурировать, а значит, в определенной степени формализовать описание проделанной работы приведет к развитию алгоритмической и логической составляющих стиля мышления специалиста в области гуманитарных наук.

На первый взгляд, работа с XML-документом лишь воспроизводит основные этапы аналитической и синтетической критики источника в традиционном источниковедении, поэтому естественные пределы применения технологии обусловлены уровнем развития традиционного исторического источниковедения и других специальных исторических дисциплин.

Однако впечатление, что применение технологии XML – лишь современный аналог классического источниковедения, обманчиво,

так как технология XML позволяет проводить все эти операции на более высоком технологическом уровне, что неизбежно приводит к расширению и корректировке устоявшихся подходов. Во-первых, источниковед, пожалуй, впервые получает возможность проводить полноценное комплексное изучение практически неограниченных совокупностей источников²⁴⁷. Во-вторых, любые операции могут быть гибко модифицированы в зависимости от задач конкретного исследования. В-третьих, XML-документ является динамической базой данных, которая может «приспосабливаться» к любой программе исследования. К тому же технология XML способна значительно упростить и ускорить труд историка-источниковеда, а возможность визуализации временной и пространственной информации стимулирует творческое воображение историка. В результате технология XML повышает степень репрезентативности и комплексности исследования.

«Опасность» же применения технологии XML видится в том, что традиционное «самоограничение» историков, источниковедов и археографов, применяющих технологию XML, – увлечение чисто технологической стороной (разметка текстов, создание обширных баз данных и т.п.) – может привести к забвению конечной цели – извлечение содержательных аспектов информации и ее интерпретация.

При этом основными недостатками технологии XML являются громоздкость и медленная обработка XML-файлов, а также невозможность *смыслового* связывания размеченных фрагментов текстов источников с «внешней» (внеисточниковой) информацией²⁴⁸ – описаниями и образами лиц, организаций, географических объектов и т.п. Для преодоления данного ограничения, накладываемого технологией XML, следует использовать семантический подход к репрезентации источниковой информации²⁴⁹.

²⁴⁷ Существенно, что расширение источниковой базы с одновременным увеличением числа параметров для источниковедческого анализа не приводит к потере какого-либо значимого аспекта информации.

²⁴⁸ Здесь имеется в виду информация, содержащаяся в исторических источниках, которые не входят в комплекс размечаемых источников – XML-документ.

²⁴⁹ См. четвертую главу. См. также: Varfolomeyev, A. Soms, H., Ivanovs, A. Knowledge-Based Information Systems...; Varfolomeyev, A., Ivanovs, A., Knowledge-Based Scholarly Environment Project for Regional Historical Studies. In: *Interactive Systems and Technologies: The Problem of Human-Computer Interaction*. Vol. III. Ulyanovsk, 2009. Pp. 273–276; Варфоломеев А. Г., Иванов А. С. Виртуальные среды для научных исследований в арсенале историка: тенденции и перспективы. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 36. Москва, 2010. С. 5–7.

Глава 4

Технологии семантического Web и возможности непосредственного связывания информации исторических источников

Виртуальные среды для научных исследований и «communities of practice»

Сетевой сервис под названием WWW с самого начала создавался как среда для оперативных научных публикаций, связанных между собой гиперссылками. Что же касается вычислительных сетей в целом, и Интернета в частности, то задача, стоящая перед ними, была и есть намного серьезнее – организация произвольных коммуникаций между людьми, программами и данными. При этом потенциальные, лишь частично реализованные возможности сети в историческом источниковедении и, в целом, в исторических исследованиях гораздо шире. Общая логика развития Всемирной паутины, представленная на рис. 4.1, диктует необходимость перехода к территориально распределенным исследованиям²⁵⁰, в качестве актуальных задач ставит разработку специализированных семантических сетей для репрезентации и изучения больших комплексов исторических источников и создание семантических публикаций памятников письменности и других типов источников.

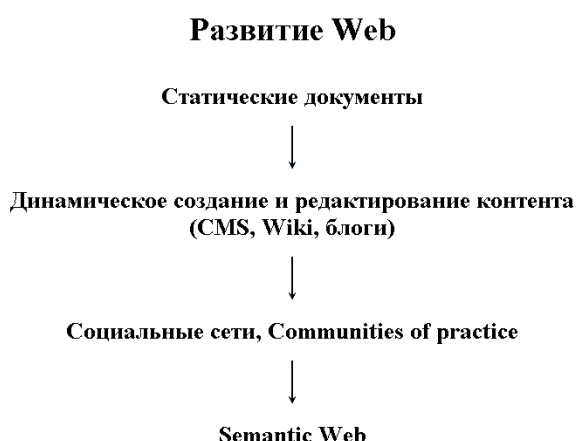


Рис. 4.1. Тенденции в развитии Всемирной паутины

²⁵⁰ Бородкин Л. И. Приоритеты современной исторической информатики: технологии e-Science. В кн.: *Круг идей: Междисциплинарные подходы в исторической информатике*. Москва, 2008. С. 5–15.

Реальные возможности распределения компьютерных вычислений среди множества машин, соединенных через сеть, появились только в конце 90-х годов XX века в связи со стремительным развитием Интернет-технологий. Способ распределенных исследований стал применяться для задач, требующих больших вычислительных мощностей либо использующих значительные объемы территориально распределенных данных. Технология организации распределенных вычислений в сети Интернет получила название Grid-технологии²⁵¹, а само направление развития научных исследований с использованием распределенных вычислений стало называться *e-Science*. В рамках Grid появился термин «виртуальная организация», обозначающий комплекс аппаратных, программных и организационных решений, а также человеческие ресурсы, обеспечивающие распределенные вычисления²⁵². В то же время возникли так называемые «виртуальные лаборатории», сначала для компьютерных экспериментов в учебных целях, затем – и для научных исследований, особенно в области биоинформатики. Основой таких лабораторий были специализированные программы, запускаемые через web-интерфейс и реализующие алгоритмы статистического анализа, сравнения последовательностей, построения филогенетических деревьев и т.д.

Виртуальные лаборатории быстро обрастали сопутствующими ресурсами, такими как тезаурусы, словари, библиотеки. С появлением технологий Web 2.0²⁵³ стали создаваться инструменты для совместного редактирования и аннотирования научных текстов, обмена библиографическими описаниями и ссылками. Появился и широко распространился термин «виртуальная среда для научных исследований» (Virtual Research Environment, VRE)²⁵⁴. Методы и технологии организации VRE стали составной частью *e-Science*.

Выход *e-Science* за рамки Grid-технологий привел к тому, что сначала в США и Великобритании, а затем и в других странах стали делаться попытки использования методологии *e-Science* в социальных, а затем – и в гуманитарных науках. В этих науках тоже возникают задачи, требующие широкомасштабных распределенных вычислений и обработки больших массивов оцифрованных данных,

²⁵¹ См.: Шевель А. Технология GRID. *Открытые системы. СУБД*. 2001. № 2. С. 36–39.

²⁵² В США получил широкое распространение близкий по смыслу термин «киберинфраструктура».

²⁵³ O'Reilly, T. What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. Sept. 2005. <http://oreilly.com/web2/archive/what-is-web-20.html> (09.12.2012); Musser, J., O'Reilly, T. *Web 2.0 Principles and Best Practices*. [S.I.] O'Reilly Media, Inc., 2006. 101 p.

²⁵⁴ Blanke, T., Hedges, M., Palmer, R. Restful Services for the e-Humanities – Web Services that Work for the e-Humanities Ecosystem. In: *DEST'09. 3rd IEEE International Conference on Digital Ecosystems and Technologies*. [S.I.], 2009. Pp. 637–642.

например, коллекций текстов в корпусной лингвистике, растровых изображений и электронных текстов письменных источников в историческом источниковедении, фотографий предметов, привязанных к планам раскопок, в археологии, аудиозаписей в устной истории и фольклористике. В данных случаях обычно применяются Grid-технологии. Однако в настоящее время более распространенными являются задачи исследования или разметки различных текстов, перевода растровых изображений источников или печатных изданий в электронные (машиночитаемые) тексты, управления коллекциями графических файлов, например, изображений археологических находок или музейных экспонатов. Такие задачи могут эффективно решаться в рамках VRE.

Первые виртуальные среды для гуманитариев создавались на основе программ для лингвистического анализа текста – лемматизаторов, морфологических анализаторов и т.д. Библиотеки таких программ широко представлены в Интернете – например, проекты GATE²⁵⁵ и TAPoR²⁵⁶. Так возникла среда MONK (Metadata Offer New Knowledge)²⁵⁷. Пользователи этой виртуальной среды имеют возможность с помощью набора программных средств проводить исследования коллекций текстов, размеченных в формате TEI. Сначала тексты с помощью токенизации и лемматизации превращаются в последовательности слов, затем подсчитываются статистические характеристики текстов, служащие основой для их сравнения и классификации. Результаты исследований могут быть записаны и сохранены в среде как составные части проектов. Однако среда не рассчитана на повторное использование методик и результатов другими исследователями.

Более широкой функциональностью обладает виртуальная среда TextGrid²⁵⁸, ориентированная на историко-филологические исследования и позволяющая работать не только с самими текстами источников, но и с их графическими изображениями, предоставляя среду для полного цикла их изучения. Лингвистические средства TextGrid ориентированы на немецкий язык. В отличие от MONK, данная среда не

²⁵⁵ General Architecture for Text Engineering. <http://gate.ac.uk/> (28.11.2012).

²⁵⁶ Text Analysis Portal for Research. <http://portal.tapor.ca> (28.11.2012).

²⁵⁷ Metadata Offer New Knowledge (MONK). <http://www.monkproject.org/> (28.11.2012).

²⁵⁸ TextGrid Project. <http://www.textgrid.de/home/> (09.12.2012). См. также: Gietz, P. at al. TextGrid and e-Humanities. In: *Second IEEE International Conference on e-Science and Grid Computing (e-Science'06). December 4–6, 2006, Amsterdam, Netherlands. 2006.* <http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/TextGrid-Amsterdam-2006-final.pdf> (05.12.2012); Aschenbrenner, A., Gietz, P., Küster, M. W., Ludwig, Ch., Neuroth, H. TextGrid – a Modular Platform for Collaborative Textual Editing...

доступна через Интернет, но ее можно развернуть в рамках одной или нескольких организаций для коллективного использования. Для масштабного проекта Monasterium, начатого немецкими историками совместно с коллегами из Австрии, Венгрии и других стран и посвященного созданию электронного архива документов из архивов монастырей Центральной Европы²⁵⁹, в университете Кельна (Германия) разрабатывается специализированный редактор EditMom для совместной распределенной работы по оцифровке и ручному распознаванию текстов средневековых актов²⁶⁰. Ряд виртуальных сред предназначен для размещения, просмотра и интерпретации изображений – примером может служить среда Virtual Vellum²⁶¹ для работы с изображениями средневековых рукописей.

Указанные среды могут способствовать решению серьезных задач в компьютерном источниковедении²⁶². Однако сравнение этих сред с современными виртуальными лабораториями биоинформатики показывает, что ряд перспективных тенденций пока себя не проявили в гуманитарных науках. Так, виртуальная среда myExperiment²⁶³ позволяет не только проводить исследования, но и сохранять так называемые workflows – блок-схемы экспериментов, которые можно сравнивать между собой и повторно воспроизводить. Активно развиваются управляемые словари и онтологии предметных областей. Разрабатываются системы идентификаторов объектов, позволяющие добиваться однозначности смысла при употреблении имен и терминов. Развиваются средства семантических аннотаций, позволяющие формировать базы знаний об объектах исследования и использовать их потом для извлечения новых знаний.

Отмеченные тенденции связаны не только с биоинформатикой; скорее они отражают магистральную линию развития Web-технологий. Поэтому постепенно эти тенденции начинают находить свое воплощение в виртуальных средах для историков. Учет этих тенденций при разработке новых или развитии уже существующих виртуальных сред позволит существенно обогатить методы истори-

²⁵⁹ Monasterium.Net. [Monasterium Project]. <http://monasterium.net/> (09.12.2012).

²⁶⁰ См., например: EditMom (Monasterium's XML Editor). <http://www.mom-wiki.uni-koeln.de/MOM-CA1/EditMOM/> (09.12.2012). См. также: Burkard, B., Gruner, S., Vogeler, G. Informatics for Historians. Tools for Medieval Document XML Markup and Their Impact on History-Science. *Journal for Universal Computer Science*. 2008. Vol. 14, no. 2. Pp. 193–210.

²⁶¹ Virtual Vellum Overview. <http://www.shef.ac.uk/hri/projects/projectpages/virtualvellum> (28.11.2012).

²⁶² Это подтверждает и опыт по созданию и использованию информационной системы «Источник» – виртуальной среды для публикации и анализа исторических источников (см. далее).

²⁶³ <http://www.myexperiment.org> (09.12.2012).

ческого исследования, в целом, и компьютерного источниковедения, в частности²⁶⁴.

Развитие VRE подразумевает появление особых сообществ исследователей, использующих программные средства и эмпирический материал, предоставляемый средами. Такие сетевые научные сообщества хорошо укладываются в популярную концепцию *communities of practice*²⁶⁵, а также в более широкое понятие «социальная сеть», что является основой для развития Web 2.0 с ее новыми формами взаимодействия пользователей и способами создания контента. Технологии Web 2.0 – Wiki-системы совместного создания текстов, системы управления контентом, блоги, Web-сервисы – активно развиваются и готовы к применению в научной инфраструктуре, наряду с традиционными технологиями *e-Science*. Деятельность сетевых сообществ исследователей, содержание которой определяется использованием распределенных систем накопления, представления, обработки информации и генерации научных знаний, целиком и полностью соответствует идее эпохи семантического Web²⁶⁶. В этом русле естественным выглядит следующий шаг – предоставить в распоряжение сообщества не только данные и инструменты анализа, но и описания проведенных ранее исследований. Технологии для достижения поставленной цели существуют.

В рамках современного направления семантического Web разрабатываются стандарты представления так называемых «бизнес-правил»²⁶⁷, в виде которых могут быть записаны и научные выводы, гипотезы, формулы или алгоритмы. Примерами таких стандартов являются язык PMML²⁶⁸, служащий для записи регрессионных и других моделей анализа данных²⁶⁹, форматы группы МКМ²⁷⁰ для обмена ма-

²⁶⁴ Одним из перспективных вариантов движения в этом направлении является, например, создание виртуальной среды, обеспечивающей работу с семантической сетью для региональных исторических исследований (см. в заключительном разделе данной главы).

²⁶⁵ См. подробнее: Ardichvilli, A. Page, V. Wentling, T. Motivation and barriers to participation in virtual knowledge sharing in communities of practice. In: *Journal of knowledge management*. 7 (1): 64–77. 2003. doi:10.1108/13673270310463626; Cox, A. What are communities of practice? A comparative review of four seminal works. In: *Journal of Information Science*. 31 (6): 527–540. 2005. doi:10.1177/0165551505057016.

²⁶⁶ Varfolomeyev, A., Soms, H., Ivanovs, A. Knowledge-Based Information Systems in Research of Regional History...

²⁶⁷ Rule Interchange Format Working Group. http://www.w3.org/2005/rules/wiki/RIF_Working_Group (09.12.2012).

²⁶⁸ Predictive Model Markup Language. <http://www.dmg.org/pmml-v3-2.html> (09.12.2012).

²⁶⁹ Pechter, R. Conformance Standard for the Predictive Model Markup Language. In: *DMSSP'06. Proceedings of the 4th International Workshop on Data Mining Standards, Services and Platforms*. 2006. Pp. 6–13. doi: 10.1145/1289612.1289613.

тематическими результатами, а также язык RuleML²⁷¹, имеющий универсальный характер. В источниковедческих исследованиях письменных источников формализация методик и результатов в виде бизнес-правил также может быть осуществлена. К сожалению, пока нельзя привести ни одного примера электронных коллекций текстов, предоставляющих подобную информацию. На наш взгляд, причиной является то, что пока подавляющее большинство текстовых коллекций ориентированы на предоставление данных для локальной работы индивидуальных исследователей и не предусматривают обмен формализованными методиками и результатами в рамках сетевых сообществ.

Итак, территориальное распределение сбора и хранения данных, разработка стандартов для свободного обмена данными, а также сервисов, позволяющих с ними работать, особую актуальность приобретают в гуманитарных науках²⁷². При этом в центре внимания оказывается социальная составляющая *e-Science* – сама организация работы сетевого научного сообщества²⁷³. Перспективы развития сетевых сообществ исследователей в гуманитарных науках уже привели к выделению направления *e-Humanities*²⁷⁴ в рамках *e-Science*. Можно также сказать, что развитие таких сообществ – это постоянный процесс обучения, в котором его члены получают новые знания как в явном формализованном виде информационных сообщений друг другу, так и в неявном виде – за счет совместного освоения инструментария и методик работы, передачи опыта между участниками сообщества. Работа в сообществе мыслится как работа в групповой сетевой операционной системе.

Одним из наиболее подходящих направлений для реализации принципов *e-Humanities* является компьютерное источниковедение, призванное изучать и вводить в научный оборот обширные комплексы исторических источников, прежде всего – памятников письменности. Можно утверждать, что чем обширнее эти коллекции и

²⁷⁰ The MKM (Mathematical Knowledge Management) Interest Group. <http://www.mkm-ig.org> (09.12.2012).

²⁷¹ The Rule Markup Initiative. <http://ruleml.org/> (09.12.2012).

²⁷² Anderson, S. E-science for the Arts and Humanities. A Discussion Paper. *AHRB E-Research Expert Seminar*, 28th April 2004. <http://www.ahds.ac.uk/e-science/e-science-discussion-2004.pdf> (04.12.2012).

²⁷³ Паринов С. И. О тенденциях и ориентирах развития открытых электронных библиотек (из опыта создания и эксплуатации системы Соционет). *Электронные библиотеки*. 2004. Т. 7. № 6. С. 1–14.

²⁷⁴ Aschenbrenner, A., Blanke, T., Dunn, S., Kerzel, M., Rapp, A., Zielinski, A. Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur. *Bibliothek: Forschung und Praxis*. 2007. Bd. 31, N. 1. S. 11–21. http://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2007/Jg_31-Nr_1/Jg_31-Nr_1_Aufsaeetze/Jg_31-2007-Nr_1-S_11-21.pdf (05.12.2012); Gietz, P. at al. TextGrid and e-Humanities...

комплексы, тем более необходима организация распределенной работы с ними²⁷⁵. Охарактеризованные выше виртуальные среды для научных исследований (TextGrid, Monasterium Project и др.) могут служить и уже служат основой для формирования сетевых сообществ исследователей – историков, источниковедов и археографов.

Опыт авторов монографии по разработке в Петрозаводском государственном университете программной среды «Источник»²⁷⁶ позволяет сформулировать основные принципы²⁷⁷ создания информационных систем, предназначенных для работы сетевых сообществ археографов и исследователей исторических источников. Причем эти принципы апробировались и, частично, корректировались в процессе развития и совершенствования системы: первоначально она возникла как сетевой сервис для подготовки на основе технологии XML печатных изданий средневековых исторических источников, прежде всего комплекса «Moscovitica–Ruthenica». И в дальнейшем электронная публикация средневековых грамот из данного комплекса использовалась и в настоящее время используется в качестве модельной электронной коллекции. Поскольку тексты источников, один раз размеченные, могут быть повторно использованы, причем не только для публикации, но и для проведения источниковедческих операций, возникла идея разрабатывать многофункциональный сетевой сервис для организации распределенных исследований, включающий в себя различные инструменты источни-

²⁷⁵ Cossard, P. K. Electronic Medievalia: Global Warming for Humanities Computing? Strategic Changes in the Economic Forecast. *The Heroic Age. A Journal of Early Medieval Northwestern Europe*. October 2006, Issue 9. <http://www.mun.ca/mst/heroicage/issues/9/em.html> (05.12.2012).

²⁷⁶ Источник. Информационная система для работы сообществ исследователей текстов. <http://istochnik.karelia.ru> (09.12.2012). См. также: Филатов В. О., Кравцов И. В., Варфоломеев А. Г. Информационная система для работы с полнотекстовыми базами данных...; Кравцов И. В., Филатов В. О. Информационная система для работы с коллекциями рукописных исторических документов. *Информационные технологии моделирования и управления*. 2007. № 2 (36). С. 188–196; Иванов А. С., Варфоломеев А. Г. Проект издания комплекса документов «Moscovitica–Ruthenica»...; Варфоломеев А. Г., Кравцов И. В. Приобретение и представление знаний в сетевом сообществе исследователей текстов...; Кравцов И. В., Варфоломеев А. Г. Принципы организации информационного пространства сетевого сообщества исследователей рукописных текстов. В кн.: *Информационное общество. Интеллектуальная обработка информации. Информационные технологии*. Москва, 2007. С. 383–386; Каргинова Н. В., Кравцов И. В., Москин Н. Д., Варфоломеев А. Г. Проект электронной библиотеки методик и результатов исследований текстовых коллекций для системы «Источник». В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Десятой Всероссийской научной конференции «RCDL'2008»*. Дубна, 2008. С. 239–245.

²⁷⁷ Кравцов И. В., Варфоломеев А. Г. Принципы организации информационного пространства...; Иванов А. С., Варфоломеев А. Г. Публикация и анализ рукописных исторических документов...

коведческой критики²⁷⁸. В результате система «Источник» трансформировалась в систему поддержки работы сетевого сообщества исследователей произвольных текстов с возможностью накопления размеченных текстов и отчетов о проведенных исследованиях в виде своеобразной «базы знаний» сообщества²⁷⁹. Можно утверждать, что в настоящее время эта система нацелена на приобретение и представление знаний²⁸⁰.

Первый принцип создания информационных систем для организации работы сетевых сообществ исследователей письменных исторических источников – это ориентация на «сплошную» репрезентацию²⁸¹ и изучение комплексов источников. Этот принцип опирается на одну из основных парадигм компьютерного источниковедения – комплексный подход к историческим источникам, подробно охарактеризованный во второй главе монографии. Еще раз отметим, что обширность естественных исторических комплексов источников обуславливает необходимость распределить различные этапы работы по реконструкции и репрезентации комплексов, а также системному изучению самих источников между участниками сетевого научного сообщества, что может значительно сократить сроки выполнения проектов.

Второй принцип – это модульность системы. В поддержку современных тенденций миграции приложений в Интернет, а также оформления их в виде специализированных сервисов, как показывает опыт развития системы «Источник», информационные системы должны строиться из набора различных компонент – небольших специализированных редакторов, социальных сервисов и информационных менеджеров. Использование различных наборов инструментов для разных целей позволяет строить произвольные исследовательские пространства. Передача информации между компонентами осуществляется в XML-формате, что позволяет прозрачно обмениваться ею не только внутри пространства системы, но и с другими системами, а также включать компоненты сторонних разра-

²⁷⁸ Филатов В. О., Кравцов И. В., Варфоломеев А. Г. Информационная система для работы с полнотекстовыми базами данных...

²⁷⁹ Там же. См. также: Кравцов И. В., Варфоломеев А. Г. Принципы организации информационного пространства...

²⁸⁰ Т.е. систему «Источник» можно с полным основанием определить как knowledge-based information system.

²⁸¹ Причем электронные тексты источников должны быть связаны с их растровыми изображениями, для чего, как отмечалось в третьей главе, может быть использована технология SVG. Изображение источника можно условно считать «оригиналом документа» в информационной системе.

ботчиков и модификации, созданные членами сообщества. (Структурная схема работы системы представлена на рис. 4.2.)

Информационная система поддержки научных исследований в виде Web-приложения включает в себя следующие компоненты: редактор подготовки полнотекстовых публикаций источников в виде XML-файлов с первичной разметкой; редактор для создания XML-файлов интерпретаций текстов со вторичной логико-семантической разметкой; среду для апробации полученных интерпретаций со встроенными средствами автоматического сравнения и визуализации разметок; среду для разработки, апробации и сохранения средств анализа текстов в виде XSLT-запросов, PHP-скриптов и т.д.; среду для совместной подготовки отчетов о результатах и методике исследования²⁸²; «хранилище знаний», состоящее из всех созданных ранее XML-файлов, инструментов анализа и сопутствующей базы данных для быстрого поиска информации, уже извлеченной из текстов в ходе проведенных исследований. Основное назначение инструментов – добавление к входящей информации новых знаний, либо трансформация информации и представление ее в другом виде.

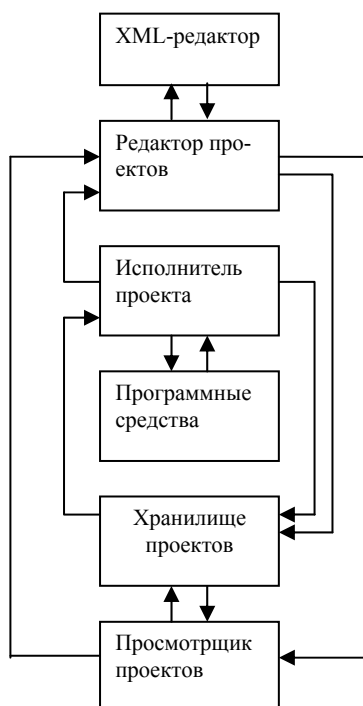


Рис. 4.2. Функциональная структура информационной системы

²⁸² Для удобства дальнейшей работы эти отчеты также желательно представлять в виде стандартизованных XML-документов.

Современные Web-приложения строятся, как правило, из наборов независимых компонент, решающих различные задачи. Так, система TextGrid включает в себя XML-редактор, редактор метаданных, интерфейс с OCR-программами, так называемые «токенизаторы» и «лемматизаторы» для лингвистической обработки текстов на различных языках, сервис для сравнения вариантов редактирования одного и того же текста и ряд других модулей²⁸³. Так же и система «Источник» строится из набора различных компонентов²⁸⁴, включая, в том числе, специализированный XML-редактор, специальный SVG-редактор для анализа и визуализации информации, RDF-редактор для описания моделей текстов, RuleML-редактор для представления алгоритмов исследования, модули сравнения и классификации моделей текстов (подключаемые пользователями) и модуль для выполнения запросов к множеству формализованных описаний методик и результатов исследований. Важным компонентом системы является конструктор экранных и печатных публикаций текста источника, который при помощи XSLT или XSL-FO преобразований позволяет приводить документ с первичной разметкой к печатному виду на экране монитора и сохранять его в формате HTML, RTF или PDF, используя подходящий шрифт. В конструкторе публикаций текст исторического источника соединяется с метаинформацией, необходимой для печатной публикации. Данный компонент работает как с отдельными источниками, так и со всем комплексом в целом, позволяя, например, сформировать для публикации справочный аппарат: лист содержания, указатели, вступительные статьи и послесловия, текстуальные примечания и комментарии, библиографические ссылки и т. д.

Третий принцип – ориентация на многоцелевое использование информации, представленной в системе. Соответственно, информационное пространство сетевого сообщества должно иметь как можно больше измерений, чтобы охватить запросы всех потенциальных потребителей информации. В этой связи можно выделить несколько вариантов обращений к информационным ресурсам системы. Прежде всего, это поиск необходимых для работы исторических источников и информации о них. Поэтому потенциальные возможности информационной системы обуславливаются, главным образом, наличием в ней исторических источников, как в их оригинальном виде

²⁸³ Küster, M. W., Ludwig, Ch., Aschenbrenner, A. TextGrid as a Digital Ecosystem...

²⁸⁴ См.: Варфоломеев А. Г. Алгоритмическое и программное обеспечение работы сетевых сообществ исследователей текстов...

(растровые изображения), так и в виде соответствующим образом размеченных электронных текстов и метайнформации. Информационная система может быть интересна также предлагаемым инструментарием, на базе которого историк может провести собственное исследование с произвольным материалом. К тому же, формально описанные и представленные в системе результаты исследований сами по себе являются уникальными информационными ресурсами. Благодаря им система дает возможность проводить исследования как на основании самих источников, так и используя уже полученные знания о них. И, наконец, сетевое сообщество само по себе – его состав и развитие, ход совместных исследований, деятельность рабочих групп, способ разделения ресурсов и информационных компонентов и т.д. – может стать объектом самостоятельного изучения, которое уже интересно само по себе – как результат взаимодействия его участников в рамках социальной сети: совместные исследования, рабочие группы, разделение ресурсов и информационных компонентов; т.е. можно анализировать эволюцию документа в системе, а можно рассматривать состав и развитие сообщества. Все эти запросы равноправны и должны найти отражение в реализации информационного пространства.

Четвертый принцип создания информационных систем, предназначенных для работы сетевых сообществ исследователей исторических источников, – это ориентация на накопление и представление знаний. Знаниями в информационной системе, ориентированной на репрезентацию и изучение письменных исторических источников, можно считать совокупность методов и результатов аналитической и синтетической критики источников, полученных и принятых сетевым научным сообществом²⁸⁵.

Поскольку репрезентация исторического источника в системе – транскрипция текста, многоуровневая разметка, представление метайнформации – является, в сущности, его интерпретацией, файлы с первичной разметкой тоже можно считать знаниями, при этом основа для их проверки – файлы с растровыми изображениями оригиналов – сохраняется в системе. Кроме задокументированных явных знаний, в процессе совместной работы и воспроизведения ее хода и результатов передаются так называемые неявные (*tacit*) знания об источниках, их комплексах и методах исследования, а также опыт экспертов. Работает правило персонального управления знаниями: «то, что я знаю,

²⁸⁵ Варфоломеев А. Г., Кравцов И. В. Приобретение и представление знаний в сетевом сообществе...

кого я знаю, и что знают те, кого я знаю». Сама же система является инструментом, на базе использования которого формируется сетевое сообщество исследователей – *community of practice*.

Если рассматривать технологии извлечения знаний, то в современных информационных системах выделяют две фазы аналитической обработки²⁸⁶. Сначала с помощью так называемых ETL-процессов²⁸⁷ структурирования контента формируется хранилище исходной и аналитической информации. Эта первичная информация используется на второй фазе извлечения знаний, к которой относятся OLAP-технологии, Text Mining и Data Mining, а также различные методы генерации гипотез на основе правдоподобных суждений. В системе «Источник» первая фаза представлена процессами разметки и связывания документов, результатами которых являются XML-документы, на основе чего формируется информационное хранилище. Таким образом, основным методом приобретения знаний сетевым сообществом является создание XML-разметок исходных документов, запросов к XML-файлам и интерпретаций этих запросов в едином формате, позволяющем накапливать и передавать знания. Вторая фаза извлечения знаний может быть представлена различными модулями прикладных исследований²⁸⁸.

Еще одним вариантом извлечения знаний из исторических источников является выделение граничных объектов (*boundary objects*), словоформ исходных текстов, которые ссылаются на внешние объекты (сущности)²⁸⁹. Этими объектами могут быть другие источники комплекса, энциклопедические статьи базы знаний системы, статьи обсуждения, а также ресурсы остального Интернета, например, партнерские информационные системы, электронные библиотеки и коллекции. Тем самым, система способна в полной мере использовать гипертекстовую природу Интернета.

Пятый принцип – открытость системы. Однако вопрос о степени открытости проектов для других участников сообщества является трудным вопросом. С одной стороны, исследователи могут быть

²⁸⁶ Ильин Н., Киселев С., Рябышкин В., Танков С. Технологии извлечения знаний из текста. *Открытые системы. СУБД*. 2006. № 6. <http://www.osp.ru/os/2006/06/2700556/> (09.12.2012).

²⁸⁷ Extract, transfer, load — извлечение, преобразование, загрузка.

²⁸⁸ На сегодняшний день в информационной системе «Источник» не представлено мощных средств анализа информации в хранилище, однако они постоянно разрабатываются параллельно с наполнением системы исходной информацией.

²⁸⁹ Добавлять такие связи в тексты можно не только после публикации, но и во время ее подготовки, как это делается, например, в Wiki-системах. Для этого приходится создавать ссылки на несуществующие или виртуальные объекты, например, документы в базе знаний проекта, которые впоследствии должны создаваться или наполняться информацией по результатам дальнейших исследований.

не заинтересованы в доступе других к незавершенным проектам. С другой стороны, чем больше проектов будет открыто, тем эффективнее будет происходить обмен знаниями в сообществе. Поэтому метаинформация о составе комплекса источников и описание методики должны быть открыты у любого проекта. По завершении проекта исследователи должны получить доступ к его результатам. И, наконец, исходные данные, если только они не опубликованы библиотекой или архивом для всеобщего использования, должны быть доступны другим лишь при согласии их владельца.

Сформулированные выше общие принципы создания информационных систем для организации работы сетевых сообществ исследователей исторических источников могут быть применены на практике благодаря технологиям Web 2.0, при этом ожидаемый результат деятельности сетевых сообществ исследователей – подготовка высококачественных семантических публикаций исторических источников и системная репрезентация источниковой информации в семантических сетях. Таким образом, создание виртуальных сред для научных исследований является технологической и организационной (инфраструктурной) основой для развития и совершенствования инструментария современного компьютерного источниковедения.

Семантические публикации исторических источников: концепция и ее реализация (SMW и ACE)

Репрезентация исторических источников в сети Интернет уже имеет свою историю. С середины 90-х годов XX века широкое распространение получили Web-проекты, цель которых – публикация исторических источников, текстов литературных произведений, а также электронных копий печатных изданий. Похожие по содержанию и целям Web-проекты объединяются в «консорциумы»²⁹⁰, в результате деятельности которых формируются сетевые архивы электронных публикаций²⁹¹ исторических источников, а также распределенные цифровые научные библиотеки²⁹².

²⁹⁰ Text Encoding Initiative. <http://www.tei-c.org/index.xml> (07.08.2012); Charters Encoding Initiative (CEI). <http://www.cei.lmu.de/> (07.08.2012); The Model Editions Partnership. <http://wyatt.elasticbeanstalk.com/mep/home> (17.11.2012).

²⁹¹ Следует отметить, что к электронным публикациям относятся также публикации на компакт-дисках (CD), в подготовке которых Web-технологии не используются. Удельный вес таких изданий незначителен.

²⁹² Burkard, B., Gruner, S., Vogeler, G. Informatics for Historians...; Aschenbrenner, A., Gietz, P., Küster, M. W., Ludwig, Ch., Neuroth, H. TextGrid – a Modular Platform...

Как уже отмечалось в главе 3, научные Web-публикации исторических источников могут иметь вид коллекций сканированных изображений с метаданной об источнике, либо размеченных текстов исторических источников, либо соединять в себе и тексты, и изображения. Различные способы представления источников в сети Интернет обусловлены, с одной стороны, выбором определенной технологии для подготовки публикации, с другой – целями публикации: или это просто обеспечение отдаленного доступа к памятникам письменности (что само по себе ценно), или же это попытка снабдить публикацию каким-либо инструментарием для источниковедческой критики источников. Технологические ограничения, существовавшие до широкого внедрения технологий Web 2.0 в процесс репрезентации источников в сети Интернет, заставляли специалистов ставить и решать вопросы о принципах и способах подготовки электронных публикаций исторических источников, а также их предназначении и целевой аудитории²⁹³. В настоящее время технологии семантического Web эти вопросы, в сущности, снимают. На передний план выдвигается задача создания сетевыми сообществами исследователей полноценных и всесторонних *семантических* публикаций комплексов источников, снабженных инструментами их аналитической и синтетической критики.

Термин «семантические публикации» в последнее время является общепризнанным и широко распространенным. Их можно определить как электронные публикации текстов в сети Интернет, снабженные дополнительными информационными слоями, в большей или меньшей степени воспроизводящими содержание и смысл текстов, а также знания о текстах в формализованном виде, что позволяет автоматически проводить операции по репрезентации, агрегации, связыванию, классификации и анализу текстов. Слово «смысл» является здесь наиболее туманным понятием, охватывающим все, что может понимать человек, читающий текст. Поскольку знания и цели людей, обращающихся к семантическим публикаци-

²⁹³ Иванов А. С., Варфоломеев А. Г. Публикация и анализ рукописных исторических документов с помощью технологии XML... С. 70; Варфоломеев А. Г., Кравцов И. В. Аналитические Web-публикации исторических документов. В кн.: *Научный сервис в сети Интернет: многоядерный компьютерный мир*. Москва, 2007. С. 389–390. См. также: Robinson, P. Current Issues in Making Digital Editions of Medieval Texts – or, Do Electronic Scholarly Editions Have a Future? *Digital Medievalist*, 1 (Spring 2005). <http://www.digitalmedievalist.org/journal/1.1/robinson/> (05.12.2012); Ciula, A., Spence, P., Vieira, J. M., Poupeau, G. Expressing Complex Associations in Medieval Historical Documents: The Henry III Fine Rolls Project. In: *Digital Humanities 2007: The 19th Joint International Conference of the Association for Computers and the Humanities, and the Association for Literary and Linguistic Computing*. Urbana-Champaign, 2007. Pp. 34–37.

ям, разные, понимание текста может включать в себя несколько уровней – от восприятия отдельных семантических аспектов текста до построения логических конструкций на основе его информации. Это приводит к существенным различиям в способах реализации семантических публикаций.

Основные виды семантических публикаций – это публикации научных статей в электронных научных журналах²⁹⁴ и цифровых библиотеках²⁹⁵. Семантические публикации научных статей, а также результатов научных экспериментов уже стали объектом специального изучения в литературе по компьютерным технологиям.

Так, на основе одной отдельно взятой статьи, посвященной вопросам биоинформатики, Дэвид Шоттон провел углубленный анализ так называемых «семантических усилителей» (semantic enhancements) – таблиц, представляющих исходные данные исследования, интерактивных диаграмм и графиков (которые могут быть трансформированы по запросу пользователя), разработанных на основе этих данных, ссылок на терминологические словари и т.п.²⁹⁶ В этом случае понятие «семантическая публикация» означает публикацию, снабженную определенным набором инструментов, которые помогают человеку (но не компьютеру) точнее и полнее воспринимать смысл статьи.

В развитии концепции семантических публикаций следует особо отметить роль сотрудника исследовательской лаборатории издательства Elsevier Аниты де Ваард, которая выделяет три степени формализации семантики и, соответственно, три семантических слоя, которые «накладываются» на традиционную электронную публикацию научной статьи²⁹⁷. Первый слой включает в себя термины и концепты, выделенные в тексте в процессе его разметки. Эти термины связаны с тезаурусами и онтологиями – контролируруемыми словарями терминов, собственных имен, географических названий и т.п. Второй слой представляет различные факты о тексте, другими словами – метаинформацию, выраженную в форме триплетов «субъект-предикат-

²⁹⁴ Shotton, D., Portwin, K., Klyne, G., Miles, A. Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Computational Biology*. 2009. Vol. 5, no. 4. e1000361. doi:10.1371/journal.pcbi.1000361; Shotton, D. Semantic Publishing. The Coming Revolution in Scientific Journal Publishing. *Learned Publishing*. 2009. Vol. 22, no. 2. Pp. 85–94.

²⁹⁵ Baruzzo, A., Casoto, P., Challapalli, P., Dattolo, A., Pudota, N., Tasso, C. Toward Semantic Digital Libraries: Exploiting Web2.0 and Semantic Services in Cultural Heritage. *Journal of Digital Information*. 2009. Vol. 10, no. 6. <http://journals.tdl.org/jodi/article/viewArticle/688/576> (07.08.2012).

²⁹⁶ Shotton, D., Portwin, K., Klyne, G., Miles, A. Adventures in Semantic Publishing...

²⁹⁷ de Waard, A. From Proteins to Fairytales: Directions in Semantic Publishing. *IEEE Intelligent Systems*. 2010. Vol. 25, no. 2. Pp. 83–88.

объект», так как именно триплеты являются основным способом записи данных в семантических сетях. Третий слой служит для репрезентации информации о ходе, методике и результатах исследовательского процесса как такового, т.е. содержит данные о постановке исследовательских задач, выдвижении гипотез, дает описание экспериментов, воспроизводит ход и логику дискуссий и т.п. В сущности, посредством трех семантических слоев происходит выделение в тексте и в метаинформации о тексте иерархии смысловых блоков и отношений между ними. Первый слой «привязывает» текст к существующим онтологиям предметной области. Благодаря второму слою образуется семантическая сеть вокруг текста, выражающая знания о нем. В совокупности, первый слой и второй слой формируют формализованный смысловой «образ» текста в семантической сети. Третий слой отражает процесс генерации научной статьи или исследования, т.е. представляет «историю» текста. Существенно и то, что эти три семантических слоя могут восприниматься не только человеком, но и компьютером, однако при этом должны использоваться формализованные языки для репрезентации информации.

Следует особо подчеркнуть, что представленные в литературе модели семантических публикаций фокусируются, как правило, на иерархической репрезентации метаинформации о текстах в семантических слоях, которые «накладываются» на электронную публикацию исследовательских текстов или текстов исторических источников.

Преимущества и потенциальные возможности семантических публикаций очевидны²⁹⁸. Во-первых, они предоставляют новые возможности и повышают качество поиска и связывания информации, так как семантические публикации позволяют сблизить компьютерные алгоритмы поиска с особенностями формулировки запросов человеком. Например, если в онтологии определен термин связан с синонимом, поиск самого термина и его синонима может быть осуществлен одновременно, так как нет необходимости упоминать синоним в запросе. Во-вторых, появляются возможности использования семантических публикаций как баз знаний, порождающих новые знания или гипотезы для дальнейших исследований с помощью процедур автоматического вывода.

В настоящее время в теоретической и практической плоскости идет работа по применению основных принципов семантических

²⁹⁸ Ahonen, E., Hyvönen, E. Publishing Historical Texts on the Semantic Web...; Mirzaee, V., Iverson, L., Hamidzadeh, B. Computational Representation of Semantics in Historical Documents. In: *Humanities, Computers and Cultural Heritage*. Amsterdam, 2005. Pp. 199–206.

публикаций к репрезентации и изучению культурного наследия²⁹⁹, в том числе и памятников письменности³⁰⁰, однако общепринятые концептуальные и технологические решения пока не найдены. Это означает, что семантические публикации исторических источников все еще находятся на стадии апробации. Далее в данном разделе показано, как основные принципы семантических публикаций могут быть применены в репрезентации комплексов исторических источников. Следует также отметить, что сам процесс создания дополнительных семантических слоев может одновременно способствовать более глубокому проникновению источниковеда и историка в смысловое поле исторических источников, поскольку является проявлением творческой интерпретации источниковой информации. При этом такие интерпретации, выражением которых становится репрезентация информации в определенном семантическом слое, могут быть частично верифицированы на основе формализованных процедур автоматического вывода.

Концепция семантических публикаций может быть реализована различными способами. Один из очевидных способов – это представление семантической структуры исторических источников в виде графов³⁰¹, являющихся «образом» семантической сети как таковой.

Текст исторического источника является многоплановым, многоуровневым объектом, между компонентами которого могут быть определены различные системы отношений, то есть одному и тому же тексту может быть поставлено в соответствие несколько различных структур³⁰². Подобные структуры естественно описывать в виде

²⁹⁹ Там же. См. также: Huvönen, E., Mäkela, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S. MuseumFinland – Finnish Museums on the Semantic Web. *Journal of Web Semantics*. 2005. Vol. 3, no. 2. Pp. 224–241; Witte, R., Gitzinger, T., Kappler, T., Krestler, R. A Semantic Wiki Approach to Cultural Heritage Data Management. In: *Language Technology for Cultural Heritage Data (LaTeCH 2008), Workshop at LREC 2008, June 1st, 2008, Marrakech, Morocco*. 2008. <http://www.l3s.de/web/upload/documents/1/latech08.pdf> (05.12.2012).

³⁰⁰ Иванов А. С., Варфоломеев А. Г. Репрезентация содержания исторических источников в семантических сетях...; Varfolomeyev, A., Ivanovs, A. Wiki Technologies for Semantic Publication of Old Russian Charters...; Ivanovs, A., Varfolomeyev, A. Semantic Publications of Charter Corpora...; Варфоломеев А. Г., Иванов А. С. Семантические публикации комплексов исторических источников...; Варфоломеев А. Г., Иванов А. С. Семантические публикации информации исторических источников на основе технологии Wiki... и др.

³⁰¹ Этот подход был апробирован на выборке фольклорных текстов. См.: Варфоломеев А. Г., Каргинова Н. В., Кравцов И. В., Москин Н. Д. Применение RuleML для представления и вывода знаний о семантической структуре фольклорных текстов, полученных на основе их теоретико-графовых моделей. В кн.: *Одиннадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2008)*. Т. 2. Москва, 2008. С. 183–191. См. также: Варфоломеев А. Г. Алгоритмическое и программное обеспечение...

³⁰² Новиков А. И. *Семантика текста и ее формализация*. Москва, 1983.

моделей – графов, которые представляют собой совокупность некоторых объектов и отношений между этими объектами.

При анализе подобных графов можно использовать различные методы³⁰³. Так, визуализация графов позволяет оценивать сложность рассматриваемой структуры и ее основные особенности. В свою очередь, аппроксимация графов дает возможность обобщить структуру графа, отбросив лишние, несущественные связи, и при этом уменьшить сложность задачи при дальнейшем анализе³⁰⁴. Большую роль играют методы сравнения и классификации графов. Один из подходов связан с использованием различных числовых инвариантов графа (число вершин и ребер, максимальная степень вершины, характеристики связности графа, распределение объектов и связей на группы и т.д.). Второй подход основан на использовании мер сходства или различия графов³⁰⁵. На множестве графов задается мера, которая позволяет оценить, насколько те или иные структуры «похожи» друг на друга.

Следует отметить, что понятие «теоретико-графовая модель текста» довольно широкое, в него входят и рассмотренные в третьей главе XML-деревья, и даже линейные модели, представляющие текст в виде цепочки следующих друг за другом структурных блоков.

В процессе анализа теоретико-графовых моделей семантической структуры текстов источников исследователь обнаруживает новые закономерности, которые обычно представляют собой логические заключения, поэтому их удобно хранить с помощью правил вывода³⁰⁶. Для записи правил (т.е. результатов исследования) и вывода знаний о семантической структуре текстов исторических источников, полученных на основе их теоретико-графовых моделей, подходящим средством является стандарт RuleML³⁰⁷, соответствующий особенно-

³⁰³ Москин Н. Д. *Теоретико-графовые модели структуры фольклорных текстов, алгоритмы поиска закономерностей и их программная реализация*. Автореф. дисс. на соиск. уч. степени к.т.н. Петрозаводск, 2006.

³⁰⁴ Москин Н. Д. Методы анализа теоретико-графовых моделей текстов. *Вестник Поморского университета. Серия «Естественные и точные науки»*. № 3. 2006. С. 105–108.

³⁰⁵ Bunke, H. Graph Matching: Theoretical Foundations, Algorithms, and Applications. *International Conference on Vision Interface, Montreal, Quebec, Canada*, 2000. Pp. 82–88. <http://www.ai.rug.nl/ki2/literature/graphmatch-bunke.pdf> (05.12.2012).

³⁰⁶ Правило представляет собой некоторую инструкцию, с помощью которой можно получить новую информацию на основе уже имеющейся. В общем виде ее можно записать в виде предложений «Если (условие), то (вывод)».

³⁰⁷ The Rule Markup Initiative. <http://ruleml.org/> (09.12.2012). Стандарт включает в себя такие языки, как PMML (для описания моделей анализа данных), MathML (для записи функциональных зависимостей), XSLT (для правил преобразований из XML в XML). Но ядром RuleML является XML-вариант языка Datalog. Он позволяет записывать факты и правила в форме продукций.

стям семантического Web. С помощью данного языка можно обмениваться правилами, созданными в рамках разных систем и задач.

Особую сложность представляет реализация заложенной в семантических публикациях возможности автоматической генерации новых закономерностей как гипотез для последующих исследований³⁰⁸. Так, возможны различные варианты автоматической генерации гипотез в зависимости от полученных результатов исследования и уже имеющихся результатов других исследований. Например, если исследователь получает какой-либо вывод на основании определенного правила, то информационная система может попробовать применить это правило к другим похожим коллекциям и предложить сравнить результаты. Если есть несколько исследований с похожими целями или методиками, проведенных на одном комплексе исторических источников, система может самостоятельно сравнить полученные результаты и сделать выводы об их сходстве или различии. Если проводится повторное исследование комплекса или отдельного источника, система может сообщить все полученные ранее результаты, чтобы исследователь мог их подтвердить или опровергнуть.

Для того чтобы подобная система работала, необходима достаточно высокая степень формализации процесса исследования комплексов исторических источников, включенных в семантическую публикацию. При этом процесс исследования можно представить в виде последовательности этапов. Так, при проведении формулярного анализа документальных исторических источников тексты источников сначала разбиваются на структурные, логические или семантические фрагменты (блоки), например, на основе технологии XML. Затем структуру текста источника необходимо представить в виде некоторого объекта, с которым будет в дальнейшем работать система. В самом простом варианте это может быть список (вектор), в котором каждый элемент будет являться идентификатором соответствующего блока в разметке. Структуру текста можно представить в виде таблицы (см. таблицу 1), где номер – это порядковый номер блока в структуре текста, тип блока – значение из заданного набора типов блоков выбранной разметки (например, протокол, контекст, эсхатокол); характеристики блока могут использоваться для хранения дополнительной информации о том фрагменте текста, который относится к данному блоку. В качестве характеристик могут выступать объем фрагмента тек-

³⁰⁸ Способы автоматической генерации гипотез частично апробированы. См.: Каргинова Н. В., Кравцов И. В., Москин Н. Д., Варфоломеев А. Г. Проект электронной библиотеки...

ста, вычисленный в словах или символах, или процентное соотношение объема блока к общему объему текста и т.п.

| № | Тип блока | Характеристика блока |
|-----|-----------|----------------------|
| 1 | A | 20 |
| 2 | B | 25 |
| ... | ... | ... |

Таблица 1. Репрезентация линейной структуры текста источника в виде таблицы

Приведенная таблица адекватно передает линейную структуру текста, которая является основным объектом изучения при проведении формулярного анализа документов или контент-анализа массовых источников³⁰⁹.

После представления структуры текста в выбранном стандартном виде можно проводить анализ при помощи правил вывода.

Правила в системе делятся на заданные изначально и формируемые в ходе работы исследователей. Заданные изначально правила обеспечивают анализ исследований, получение новых выводов и применение уже существующих закономерностей к текущему исследованию. Также изначально может быть задан набор фактов, устанавливающих, например, что какая-то структура является общей для текстов группы источников или что две структуры похожи друг на друга:

*структура 1 – общая для структур
(структура 2, ..., структура n)
структура 1 похожа на структуру 2 на n %.*

Некоторые факты могут быть представлены в виде функций (алгоритмов). Например, функция *compare_structure* с двумя аргументами – структурами возвращает их сходство в процентах:

compare_structure(структура 1, структура 2) = n %.

К заданным изначально правилам также относятся шаблоны, с помощью которых могут записываться и новые правила, полученные в ходе исследования. Ниже приведены примеры шаблонов.

³⁰⁹ В случае использования более сложной иерархической или графовой структуры таблица представления результатов тоже должна быть усложнена.

| |
|--|
| <p>ЕСЛИ</p> <p><i>текст 1 — структура 1</i></p> <p><i>текст 2 — структура 2</i></p> <p>....</p> <p><i>текст n — структура n</i></p> <p><i>(текст 1, ..., текст n) – тип 1</i></p> <p><i>структура k – общая для структур (структура 1, ..., структура n)</i></p> |
| <p>ТО</p> <p><i>структура k – общая структура для типа 1</i></p> |

Шаблон 1

В соответствии с данным шаблоном, если для текста каждого источника из определенной их совокупности была получена определенная структура и известно, что все источники относятся к одному типу (виду или разновидности), то можно сделать предположение об общей структуре для текстов данного типа.

| |
|--|
| <p>ЕСЛИ</p> <p><i>текст 1 — структура 1,</i></p> <p><i>текст 2 — структура 2,</i></p> <p>....</p> <p><i>текст n — структура n,</i></p> <p><i>(текст 2, ..., текст n) имеют тип 1,</i></p> <p><i>(структура 1, структура 2..., структура n)</i></p> <p><i>похожи на $m\%$,</i></p> <p><i>m больше, чем пороговое значение</i></p> |
| <p>ТО</p> <p><i>текст 1 имеет тип 1</i></p> |

Шаблон 2

В соответствии с этим шаблоном, если имеется несколько текстов источников достаточно близкой структуры и про часть из них известно, что они относятся к определенному типу, можно предположить, что оставшиеся тексты также относятся к этому типу. Следует отметить, что шаблоны могут отражать вовсе не только дедук-

тивные правила вывода, но, как в указанных примерах, «правдоподобные» рассуждения, основанные на индукции и аналогии³¹⁰.

Правила, формирующиеся в ходе работы исследователей, представляют собой формализованную запись выводов исследователей. Они могут формироваться как конкретизация имеющихся в системе шаблонов. В таком случае на место свойств, типов и структур подставляются конкретные значения. Эти правила могут быть сохранены в библиотеке системы и использованы в дальнейших исследованиях. Для каждого правила должны задаваться область видимости и коэффициент доверия. По умолчанию исследователь видит только свои правила, но при желании может ознакомиться с правилами других исследователей. Если в ходе исследования какое-то уже сохраненное в системе правило подтверждается или опровергается, это может вызвать изменение его коэффициента доверия. Таким образом, в дальнейшем исследователи могут выбирать наиболее подтвержденные правила.

Как для записи результатов исследования семантической структуры текстов исторических источников, полученных на основе их теоретико-графовых моделей, так и для записи правил, полученных в ходе формулярного анализа, целесообразно использовать стандарт RuleML в качестве формата представления знаний. Ниже приводится пример описания фрагмента шаблона 2 на языке RuleML:

```
<Implies>
  <head>
    <Atom>
      <Var>text 1</Var>
      <Rel>type of text</Rel>
      <Var>type 1</Var>
    </Atom>
  </head>
  <body>
    <Atom>
      <Var>text 2</Var>
      <Rel>type of text</Rel>
      <Var>type 1</Var>
    </Atom>
    <Atom>
      <Var>text 1</Var>
      <Rel>structure of text</Rel>
    </Atom>
  </body>
</Implies>
```

³¹⁰ Вагин В. Н., Головина Е. Ю., Загорянская А. А., Фомина М. В. *Достоверный и правдоподобный вывод в интеллектуальных системах*. Москва, 2004.

```

        <Var>structure 1</Var>
    </Atom>
    <Atom>
        <Var>text 2</Var>
        <Rel>structure of text</Rel>
        <Var>structure 2</Var>
    </Atom>
    <Atom>
        <Var>structure 1</Var>
        <Rel>is like</Rel>
        <Var>structure 2</Var>
        <Ind>m %</Ind>
    </Atom>
</body>
</Implies>

```

Для обеспечения логического вывода на основе имеющихся фактов необходимо использовать машины логического вывода, работающие с форматом RuleML³¹¹. Примерами таких машин являются: Bossam³¹², которая позволяет строить приложения в рамках концепции Semantic Web, а также ОО jDREW³¹³ – объектно-ориентированная дедуктивная машина вывода для RuleML. ОО jDREW представляет собой библиотеку, написанную на языке Java. Таким образом, можно написать набор приложений на Java, которые будут реализовывать функции системы и осуществлять логический вывод, используя функции этой библиотеки.

Также существует система DR-DEVICE³¹⁴, позволяющая осуществлять рассуждения в условиях неполной и противоречивой информации. Это актуально в условиях семантического Web: например, противоречия могут возникнуть при объединении онтологий.

³¹¹ Другой способ – использование какого-либо внутреннего формата для хранения фактов и правил, например, форматов представления знаний таких сред, как CLIPS или SWI-Prolog. Программы, написанные в обеих этих средах, могут быть интегрированы с программами на других языках, в том числе с Java, что позволит создать web-интерфейс для системы. Сама программа, осуществляющая логический вывод, может выполняться на сервере, а передача входных и выходных данных может осуществляться через web-интерфейс. Стандарты XML, RDF и RuleML будут играть роль форматов для представления результатов и методики исследования другим сообществам исследователей и другим системам. Для преобразования из внутренних форматов во внешние необходимо будет использовать уже имеющиеся программы (например, SWI-Prolog содержит библиотеку для разбора данных в формате RDF) или реализовать собственную.

³¹² Bossam Rule/OWL Reasoner. <http://bossam.wordpress.com> (09.12.2012).

³¹³ Object Oriented jDREW. <http://www.jdrew.org/ooidrew> (09.12.2012).

³¹⁴ DR-DEVICE. A Defeasible Logic Reasoner for the Semantic Web. <http://lpis.csd.auth.gr/systems/dr-device.html> (09.12.2012).

Система поддерживает использование фактов, правил и приоритетов правил. DR-DEVICE позволяет загружать RDF документы из Интернета и использовать их в качестве входной информации – фактов – для программы. Правила могут быть записаны либо на языке, подобном языку CLIPS, либо с использованием синтаксиса RuleML. Полученные результаты можно выгружать в Интернет в виде RDF документов.

Описанные выше алгоритмы исследовательских операций ориентированы, прежде всего, на углубленное изучение семантической структуры исторических источников, их репрезентацию, сравнение и получение результатов исследования на основе автоматического вывода. Однако репрезентация комплексов источников в семантических сетях подразумевает также широкие возможности непосредственного связывания источниковой информации с уже имеющейся в сети Интернет информацией, что является условием для полноценной аналитической и синтетической критики в рамках компьютерного источниковедения. Это означает, что выделенная (размеченная) информация исторических источников может и должна быть связана (соотнесена) не только с информацией других источников, входящих в комплекс, но и с «внешней» информацией, представленной, прежде всего, в онтологиях. Именно поэтому основой семантических публикаций исторических источников являются онтологии, которые могут видоизменяться в широких пределах – от простых терминологических словарей или систем классов для описания объектов до сложных структур, отражающих разнообразные знания о предметной области в виде классов, объектов, отношений, ограничений, логических правил «если – то» и т.д.

За последние годы было создано множество онтологий, в том числе и для предметных областей, связанных с историей и источниковедением. В качестве наиболее разработанной и универсальной часто рассматривается онтология CIDOC (le Comité International pour la Documentation) CRM (Conceptual Reference Model), созданная для описания предметов в музейных коллекциях, но включающая в себя также онтологию исторических событий с персоналиями, их ролями, географическими местами, моментами и периодами времени³¹⁵. На этой основе были разработаны другие исторические онтологии, ори-

³¹⁵ The CIDOC Conceptual Reference Model. <http://www.cidoc-crm.org/> (07.08.2012). См. также: Doerr, M. The CIDOC CRM. An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*. 2003. Vol. 24, no. 3. Pp. 75–92.

ентированные на более узкие цели³¹⁶, в том числе – в рамках проектов Pearl Harbor Project³¹⁷ в США и CultureSampo Project³¹⁸ в Финляндии. К сожалению, указанные выше онтологии лишь частично отвечают задачам источниковедческого изучения письменных источников, так как в них не отражены особенности этих источников. К тому же данные онтологии созданы на основе так называемого событийно-ориентированного подхода. Авторы монографии в создании специализированных онтологий, предназначенных для семантической публикации комплексов письменных исторических источников, предлагают иной, документно-ориентированный подход для описания исторических знаний и так называемых «исторических объектов» – документов, лиц, географических объектов, учреждений, конкретных событий и т.п. Здесь события выступают не узлами сети, а только ее связями, порожденными свидетельствами источников³¹⁹.

Этот подход нашел свое воплощение при создании специализированной онтологии для подготовки прототипа семантической публикации субкомплекса смоленско-рижских актов из комплекса документальных источников «Moscovitica–Ruthenica». Данная онтология отражает два вида семантических связей, которые устанавливаются в семантической публикации. Во-первых, это связи между исторически и тематически взаимосвязанными актами. Эти связи возникли в период создания и оборота документов в контексте своего времени, т.е. в XIII веке, когда данные акты выполняли свои прямые функции по документированию отношений между Смоленском и Ригой. Во-вторых, это связи с другими историческими источниками, которые не входят в субкомплекс смоленско-рижских актов, представленный в прототипе семантической публикации. В последнем случае, «внутренняя» информация субкомплекса смоленско-рижских актов связывается с «внешней» информацией, которую напрямую предоставляют другие исторические источники³²⁰ или, опосредованно, научные труды, эн-

³¹⁶ Pasin, M., Motta, E., Keynes M. PhiloSURFical: Browse Wittgensteins Tractatus with the Semantic Web. In: A. Pichler, H. Hrachovec, eds. *Wittgenstein and the Philosophy of Information. Proceedings of the 30th International Ludwig Wittgenstein Symposium*. Vol. 1. [S.l.]: Ontos Verlag, 2008. Pp. 319–333.

³¹⁷ Ide, N., Woolner, D. Historical Ontologies. In: K. Ahmad, Chr. Brewster, M. Stevenson, eds. *Words and Intelligence*. Vol. II: *Essays in Honor of Yorick Wilks*. Dordrecht: Springer, 2007. Pp. 137–152.

³¹⁸ Ahonen, E., Huvönen, E. Publishing Historical Texts on the Semantic Web...

³¹⁹ Varfolomeyev, A., Ivanovs, A. Knowledge-Based Scholarly Environment... См. подробнее в последнем разделе настоящей главы.

³²⁰ Например, другие ливонские акты. См.: Напиерский К. Э., сост. *Грамоты, касающиеся до сношений северо-западной России с Ригой и ганзейскими городами в XII, XIII и XIV веке*. СПб., 1857. № III и др.

циклопедические словари, тезаурусы и онтологии. В совокупности эти связи формируют определенную семантическую сеть, которая представляет знания о смоленско-рижских актах. В сущности, такая семантическая сеть предлагает модель интеграции и связывания обширной информации, относящейся к субкомплексу исторических источников, а поэтому она может стать достаточно надежной основой для подготовки семантической публикации этих источников.

На рис. 4.3 показан сегмент семантической сети, созданной на основе выявленных связей смоленско-рижских актов. В центре сегмента семантической сети – акт № 6 – грамота архиепископа рижского смоленскому князю Федору Ростиславичу³²¹. Тематически и, частично, текстуально этот источник связан с другими смоленско-рижскими актами – № 4 (грамота, написанная от имени смоленского князей Федора Ростиславича, о предоставлении чистого пути немецким и русским купцам), № 5 (грамота Федора Ростиславича по судному делу о немецком «колоколе»), № 3а (список А Готландской редакции договора 1229 года) и № 1 (так называемый «договор неизвестного смоленского князя»). Следует отметить, что связи между этими актами установлены на основе исследовательской интерпретации «внутренней» информации семантической публикации субкомплекса. В свою очередь, «внешняя» информация использовалась для идентификации лиц, упомянутых в актах, а также для полноценного описания других исторических объектов. В рамках семантической сети отношения между ее объектами описываются в виде триплетов: «charter written by person», «charter sent to person», «charter mentions person»³²². Как это обычно делается в онтологиях, могут быть также введены и обратные (инверсионные) отношения между объектами, например: «person is mentioned in charter». Нельзя не отметить, что данная семантическая сеть частично основывается на гипотетических данных; соответственно, гипотетический или неопределенный характер некоторых отношений обозначен или знаком вопроса, или определенной комбинацией слов: «probably refers to» вместо «refers to».

³²¹ Об актах см. во второй главе монографии. Уже отмечалось, что в семантической публикации источникам из субкомплекса смоленско-рижских актов и из комплекса документов приказного делопроизводства по истории Динабурга присвоены условные номера в соответствии с их последними печатными изданиями.

³²² Разумеется, триплеты выражены на английском языке, так как это «универсальный» язык для репрезентации и связывания информации в семантическом Web.

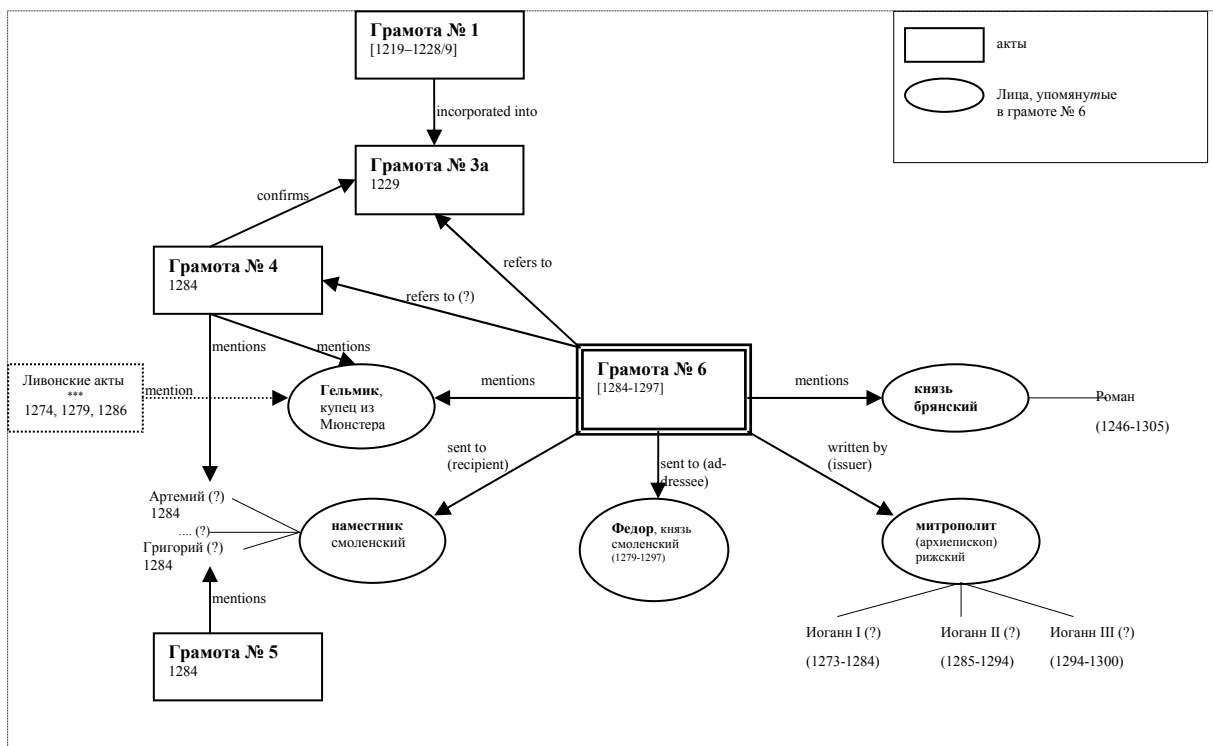


Рис. 4.3. Представление субкомплекса смоленско-рижских актов в виде открытой семантической сети (фрагмент)

Следующий пример включения конкретного источника в систему объектов, между которыми устанавливаются семантические связи, позволяет детально показать процесс генерирования семантической сети для репрезентации комплекса делопроизводственных документов по истории Динабурга в середине XVII века. В центре данного фрагмента семантической сети находится документ № 1 – отписка С. Л. Стрешнева царю Алексею Михайловичу об осаде города Динабурга, написанная не ранее 19 июля 1656 года. Далее приводится начало документа, информации которого вполне достаточно для описания процесса формирования семантической сети:

«Г(о)с(у)д(а)рю ц(а)рю і великому кн(я)зю Алеѣью
 Михаило|вичю всеа Великия и Малыя и Бѣлыя| Росіи самодержцу
 холопи твои Сен(ь)ка| Стрешнев с товарищи челом бьют.

Июля,| г(о)с(у)д(а)рь, въ 1 де(нь) в твоеи г(о)с(у)д(а)р(е)ве
 ц(а)р(е)ве і великог(о)| кн(я)зя Алеѣѣя Михаиловича всеа Великия| и
 Малыя и Бѣлыя Росіи самодержца| грамоте писано к нам, холопам

твоимъ:| велено ѿписати к тебѣ великому г(о)с(у)д(а)рю| тотчасъ,
что у нас, холопеи твоих, твоего| г(о)с(у)д(а)р(е)ва дѣла дѣлаетца.

И июля, г(о)с(у)д(а)рь,| въ 19 де[нь] писали мы, холопи твои, к
тебѣ,| великому г(о)с(у)д(а)рю, з друискимъ казакомъ| с Гришкою
Скирмонтовым, что мы, холо[п]я | твои, со всѣми твоими
г(о)с(у)д(а)р(е)выми слу|жилыми людми пришли под Диноборокъ|
июля въ 18 де[нь] и город Диноборок осадили. | И Б(о)жиимъ и твоимъ
г(о)с(у)д(а)р(е)вым ц(а)р(е)вым| і великог(о) кн(я)зя Алеѣя Михаи-
ловича | всеа Великия и Малыя и Бѣлыя| Росіи самодержца дѣлом
над городом| и над городцкими сидѣл(ь)цы учали промы|шлят(ь) –
шанцы копать и туры ставит(ь)| того ж числа [...]»³²³.

Вышеприведенный текст предполагает установление ряда семантических связей с другими историческими объектами; эти связи выходят за рамки свидетельств данного источника (на рис. 4.4 обозначен как «Doc1»). Во-первых, это исторические деятели, упомянутые в начале отписки – царь Алексей Михайлович и боярин С. Л. Стрешнев. Во-вторых, это документ, который до нас не дошел, – указная грамота царя Алексея Михайловича от 1-го июля 1565 года, на которую ссылается воевода С. Л. Стрешнев во втором абзаце документа (на схеме – DocX1³²⁴). В-третьих, это еще один документ – несохранившаяся отписка С. Л. Стрешнева царю Алексею Михайловичу, отправленная 19 июля с казаком Г. Скирмонтовым, содержание которой кратко изложено в третьем абзаце (на схеме – DocX2).

Данных объектов, в принципе, достаточно для того, чтобы можно было приступить к генерированию семантической сети (см. рис. 4.4); при этом в сеть включаются также специальные темпоральные объекты – моменты и периоды времени (timeline-objects), представленные в виде неких временных точек на хронологической шкале. Данные объекты отражают относительную датировку отписки С. Л. Стрешнева по косвенным датирующим признакам, содержащимся как в самом тексте документа, так и вне его текста. Связи между объектами в рамках генерируемой семантической сети на схеме обозначены стрелками,

³²³ РГАДА. Ф. 210. Московский стол. Столб. 270. Л. 674. См.: Иванов А., Кузнецов А. *Динабург в документах Российского государственного архива древних актов...* Часть 1. № 1. С. 67.

³²⁴ Литера «X» указывает на то, что данный исторический источник отсутствует в комплексе динабургских документов.

при этом характер связей описываются в виде триплетов: «document written by person», «document sent to person», «document refers to», «document written before [date]», «document written after [date]».

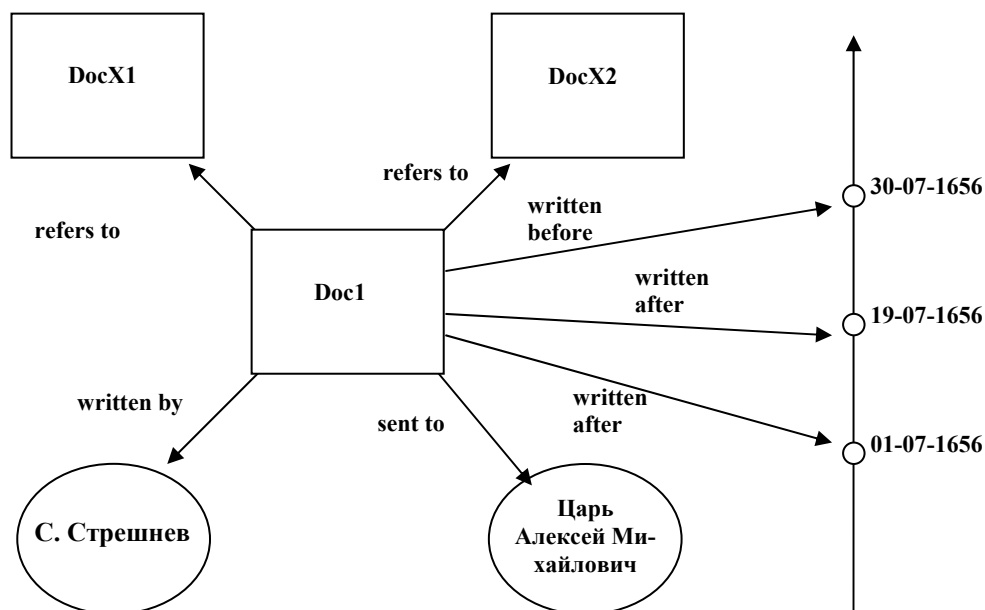


Рис. 4.4. Генерирование семантической сети на основе комплекса документов по истории Динабурга середины XVII века

Продолжение работы по генерированию семантической сети будет заключаться в росте числа объектов, связанных с данной отпиской, с одновременным увеличением интенсивности связей между ними; при этом внутренние связи внутри комплекса источников дополняются внешними связями (как это было показано на примере формирования семантической сети, связывающей смоленско-рижские акты). При этом отписка С. Л. Стрешнева по-прежнему может оставаться своего рода центром семантической сети. При включении же в эту сеть других документов, «генетически» связанных с данной отпиской, возникает семантическое поле, связи в котором будут тяготеть к взаимосвязанным и взаимозависимым семантическим центрам (узлам). В конечном итоге, в семантическое поле могут быть интегрированы все документы, образующие исторически сложившийся комплекс источников, связанных с историей Динабурга в 50–60-е годы XVII века.

В рамках семантического Web семантическая сеть, в сущности, является математической моделью, представляющей множество объектов, связанных между собой ссылками различных типов (т.е. семантическими отношениями). Как уже отмечалось, такую сеть можно выразить с помощью триплетов «субъект–предикат–объект».

Триплеты образуют основу онтологии – базы знаний о предметной области, в данном случае – об источниках, представленных в семантической публикации. Для передачи семантических связей и записи онтологий могут использоваться технологии RDF и OWL³²⁵. При этом, если онтологию удастся записать, например, с помощью выразительных средств языка OWL DL, то дальше специальная программа OWL-reasoner может производить логический вывод, то есть получать следствия из посылок, напрямую не представленные в базе знаний. В результате появляется возможность получать новые знания о предметной области. Однако непосредственное создание семантических публикаций на основе онтологий, записанных в форматах RDF или OWL, видится довольно затратным делом. Поэтому актуальным является изучение возможностей инструментов, облегчающих использование технологий семантического Web.

В качестве альтернативы RDF и OWL можно предложить технологию Wiki³²⁶, представляющую собой программные комплексы, которые размещаются на Web-сервере. В наше время Wiki-системы часто используются в качестве специфических редакторов онтологий, созданных сетевыми сообществами исследователей³²⁷. В качестве средства создания семантических публикаций средневековых русских источников может быть использована среда Semantic MediaWiki (SMW)³²⁸ – часто используемое расширение популярной системы MediaWiki.

Преимущество SMW заключается в том, что она, обладая характерными для Wiki возможностями распределенного создания и редактирования Web-страниц, дает возможность целому ряду исследователей практически одновременно участвовать в процессе формирования семантической сети, в том числе – и в редактировании заметки текстов исторических источников, включенных в эту сеть. При этом информация обо всех этапах такой распределенной работы и

³²⁵ См., например: Ciravegna, F., Greengrass, M., Hitchcock, T., Chapman, S., McLaughlin, J., Bhagdev, R. Finding Needles in Haystacks: Data-mining in Distributed Historical Datasets. In: Greengrass, M., Hughes, L., eds. *The Virtual Representation of the Past*. [S.l.], 2008, reprinted in 2010. Pp. 65–79.

³²⁶ См. подробнее о семантических Wiki: http://en.wikipedia.org/wiki/Semantic_wiki (09.12.2012).

³²⁷ Souzis, A. Building a Semantic Wiki. *IEEE Intelligent Systems*. September–October 2005. Vol. 20, no. 5. Pp. 87–91; Auer, S., Dietzold, S., Riechert, T. OntoWiki – A Tool for Social, Semantic Collaboration. In: Cruz, I. et al., eds. *The Semantic Web – ISWC 2006*. Berlin; Heidelberg, 2006. Pp. 736–749. (Lecture Notes in Computer Science, 4273).

³²⁸ Semantic MediaWiki (SMW). <http://www.semantic-mediawiki.org/> (07.08.2012).

внесенных изменениях сохраняется в базе данных³²⁹. Еще одна особенность SMW – упрощенная разметка, которая позволяет добавлять к тексту тот или иной семантический слой. Эта разметка может быть использована для выделения в текстах источников различных объектов и для дополнения источниковой информации метаинформацией³³⁰. Размеченный таким образом текст может быть преобразован в набор «фактов», например, триплетов в формате RDF. Удобство применения данной системы заключается в том, что любой исследователь может ее апробировать, не устанавливая систему на Web-сервере. Для этого необходимо зарегистрироваться на одном из сайтов, бесплатно предоставляющем доступ к SMW³³¹.

В SMW объекты представляются с помощью Wiki-страниц. Роль классов играют категории, которые можно назначать страницам, выражая тем самым принадлежность объекта классу. Категории могут вкладываться друг в друга, образуя иерархии классов. Для задания всех остальных отношений используются типизированные гиперссылки между страницами. Таким образом, в роли субъектов и объектов семантических отношений в SMW всегда выступают именно Wiki-страницы, что существенно ограничивает возможности по представлению знаний об источнике. Так, в SMW легко выразить утверждение, что данный документ адресован некоему человеку (для этого и документ, и человек должны быть представлены в виде Wiki-страниц), но невозможно адекватно отразить факты, относящиеся к другим фрагментам текста или к отдельным словам.

В ходе апробации (пилотных исследований) применимости SMW для репрезентации субкомплекса смоленско-рижских актов и документов по истории Динабурга³³² в семантических сетях оказалось, что базовая версия SMW мало пригодна для создания полноценных семантических публикаций источников по истории России. Так, в ней отсутствуют возможности по отображению текста с XML-разметкой, а также нет возможности подбирать подходящие шрифты. Поэтому на этапе апробации SMW и реализации пилотных проектов вместо адек-

³²⁹ В качестве примера применения одной из многочисленных Wiki-систем можно привести широко известную «Википедию». Это энциклопедия, созданная на основе технологии Wiki, которая изначально не была семантической.

³³⁰ Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., Studer, R. Semantic Wikipedia. *Journal of Web Semantics*. 2007. Vol. 5, no. 4. Pp. 251–261.

³³¹ Один из таких сайтов – http://www.referata.com/wiki/Main_Page (09.12.2012).

³³² В среде SMW были размещены пять взаимосвязанных между собой актов XIII–XIV веков из комплекса «Moscovitica–Ruthenica», а также несколько документов XVII века по истории Динабурга. См.: Varfolomeyev, A., Ivanovs, A. Wiki Technologies for Semantic Publication of Old Russian Charters...; Варфоломеев А. Г., Иванов А. С. Семантические публикации информации исторических источников на основе технологии Wiki... и др.

ватных лингвистических транскрипций средневековых русских источников были использованы в одном случае переводы актов на английский язык, а в другом случае – упрощенные транскрипции текстов документов на основе современного русского алфавита.

Как показывает приведенный ниже пример SMW-разметки акта № 6 – грамоты архиепископа рижского, типизированные гиперссылки на страницы позволяют отразить довольно широкий круг семантических связей данного источника с другими объектами семантической сети (см. также рис. 4.3). Эти гиперссылки обозначены в двойных квадратных скобках и обычно включают в себя три части: (1) тип отношений («written by», «sent to», «refers to», «probably refers to» и т.п.); (2) наименование объекта, т.е. название страницы, на которую указывает гиперссылка («Archbishop of Riga», «Helmich» и др.); (3) фрагмент текста источника, непосредственно относящийся к гиперссылке и поэтому включенный в нее (например, «Metropolitan's of Riga»). Страницы, содержащие тексты источников с гиперссылками, являются в данном случае основными объектами семантической сети; в свою очередь, гиперссылки выделяют отношения между объектами:

[[written_by::Archbishop of Riga|Metropolitan's of Riga]] benediction over our beloved sun, [[sent_to::Fedor Rostislavich, Prince of Smolensk|Grand Prince Fedor]], and over his children, and over [[sent_to::Bishop of Smolensk]], and over [[sent_to::Governor of Smolensk]], and over all boyars. You should know about that complaint, which the inhabitants of Vitebsk lodged against Rigans in order to set themselves right with [[mentions::Helmich]]. And their words were as follows – they wanted to justify themselves with those words – and they said to [[mentions::Prince of Briansk]] that 50 men had ridden out of Riga, and killed a man, and taken 10 berkovets of wax. And now I, Metropolitan, say that those inhabitants of Vitebsk unjustly complained on Rigans. And now I know that Rigans are not guilty of that. And now I am surprised that Governor listens to anyone. But there is [[refers_to::Charter 3a|an agreement between you and us]]: if a controversy arises, it should be settled between you and us. And now I pray that you would observe [[probably_refers_to::Charter 4|that agreement and oath]]: if anyone complains to you about Rigans, or [[Helmich]], or somebody else, you should send a missive to us, but we will render justice in accordance with divine justice.

Благодаря гиперссылкам определенные фрагменты текста акта оказываются связанными с другими страницами, содержащими тексты других актов, а также данные, относящиеся к историческим ли-

цам, географическим объектам, событиям³³³. Важно, что SMW также позволяет визуализировать эти гиперссылки.

Особо стоит отметить, что гиперссылки автоматически генерируют наборы фактов, относящихся, например, к акту № 6: «Charter 6 mentions Helmich» (акт № 6 упоминает Гельмика), «Charter 6 probably refers to Charter 4» (акт № 6, возможно, ссылается на акт № 4), «Charter 6 refers to Charter 3a» (акт № 6 ссылается на акт № 3a), «Charter 6 is written by Archbishop of Riga» (акт № 6 написан архиепископом рижским) и т.п. Эти факты «привязаны» к тексту акта; причем в некоторых случаях такая связь устанавливается на основе творческой интерпретации содержания источника.

Другие страницы семантической сети, включая страницы, содержащие информацию о лицах, географических объектах и событиях, формируются по тому же образцу. И здесь данные о лицах, местах и событиях записываются в виде последовательностей триплетов. Так, в разметке SMW текст «Helmich is a merchant from Münster. He is mentioned in Charter 6 and Charter 4» (Гельмик является купцом из Мюнстера. Он упомянут в акте № 6 и акте № 4) будет представлен следующим образом:

```
Helmich is a [[is_a::merchant]] from
[[is_from::Münster]]. He is mentioned in
[[is_mentioned_in::Charter 6]] and [[is_mentioned_in::Charter 4]].
```

В документах по истории Динабурга облегченная разметка SMW также позволила без труда вручную выделить в текстах упоминания персоналий, исторических мест, дат, а также выразить некоторые другие аспекты информации источников, например, указания на адресата и отправителя документа, ссылки в текстах на другие документы и др. В итоге, приведенный ранее фрагмент отписки С. Л. Стрешнева (документ № 1) в разметке SMW будет выглядеть следующим образом³³⁴:

```
"Отписка боярина и воеводы Семена Лукьяновича Стрешнева царю Алексею Михайловичу об осаде города Динабурга"
```

```
"Дата создания:" не ранее 19 июля 1656 г. Вероятно, не позднее 30 июля 1656 г. (день взятия Динабурга).
```

```
----
```

³³³ См. публикацию акта № 6 в разметке SMW на сайте http://histdocs.referata.com/wiki/Main_Page (09.12.2012).

³³⁴ В квадратных скобках показаны семантические связи с другими объектами семантической сети; на рис. 4.4 эти связи обозначены стрелками.

Государю царю и великому князю [[sent to::Алексей Михайлович|Алексю Михайловичу]], вся Великия и Малыя и Белыя России самодержцу, холопы твои [[written by::Семен Стрешнев|Сенька Стрешнев]] ^{[[comment Doc1.1|1]]} с товарищами челом бьют.

Июля, государь, в [[after::Jul 1 1656|1 день]] в твоей [[make reference to::DocX1|грамоте]] писано нам, что велено отписать тебе тотчас, что у нас твоего государева дела делается. И июля, государь, в [[after::Jul 19 1656|19 день]] [[make reference to::DocX2|писали]] мы, холопы твои, тебе, великому государю, с друйским казаком [[Григорий Скирмонтов|Гришкой Скирмонтовым]], что мы со всеми твоими государевыми служилыми людьми пришли к Динабургу июля в 18 день и город Динабург осадили. И Божым и твоим государевым произволением над городом и над городскими жителями начали промышлять, шанцы копать и туры ставить того же числа.

При этом непосредственно на экране монитора представлен лишь сам текст источника (см. рис. 4.5³³⁵), а семантические связи его с другими объектами обозначены в нижней части экрана под текстом отписки («Facts about Doc1»).



Рис. 4.5. Текст отписки воеводы С. Л. Стрешнева на сайте <histdocs.referata.com>

³³⁵ Данный документ представлен в Интернете: <http://histdocs.referata.com/wiki/Doc1> (09.12.2012).

Однако в обоих случаях серьезным ограничением SMW-разметки можно считать невозможность в рамках нее создавать «вложенные» или «пересекающиеся» текстовые фрагменты, а также соединять вместе структурную и семантическую разметку, использовать стандарты TEI или CEI, привязывать текст к изображению источника. Тем не менее, несмотря на ограничения, среда SMW оказалась очень удобной для быстрого создания семантической сети исторических источников, отражающей знания о взаимосвязи источников друг с другом и с другими объектами (такими, как персоналии). Разумеется, такая сеть в ее настоящем виде вряд ли может использоваться для решения задач аналитической и синтетической критики исторических источников в компьютерном источниковедении. Это всего лишь модель сети, служащая для поиска оптимального варианта репрезентации и способов связывания в семантической сети источников и внеисточниковой информации.

Следует также отметить, что возможности отображения текста с XML-разметкой и подбора подходящих шрифтов уже предусмотрены в расширениях SMW, таких как XML2Wiki и WebFonts, причем разработка расширений для SMW довольно легка для Web-программистов – их создано уже несколько сотен. Поэтому изучение возможностей уже созданных расширений и разработка новых расширений для таких систем, как SMW, являются перспективными направлениями деятельности специалистов в области исторической информатики. Таким образом, в перспективе, с помощью расширений нетрудно добиться того, чтобы SMW служила средством для создания полноценных Web-публикаций исторических источников с передачей их изображений, дипломатических транскрипций оригинальных текстов с использованием подходящих шрифтов, текстов с XML-разметкой, изложений текстов на современном языке, справочного аппарата.

Существующие и апробированные авторами монографии модели семантических публикаций создают широкие возможности для семантического связывания прежде всего метаинформации – генерированной исследователем внетекстовой информации об источниках. К сожалению, имеющиеся модели семантических публикаций не позволяют *непосредственно* представить в семантических сетях свидетельства (информацию) источников: семантическая репрезентация текстовой источниковой информации обязательно предполагает тщательную и трудозатратную предварительную разметку источников, выявление логической последовательности структурных

частей, увязывание формальной и семантической структуры текстов³³⁶. При этом любая схема разметки неизбежно упрощает содержание источника, в результате частично теряется смысл. Даже разметка текстов на основе схем разметки TEI и CEI, хоть и является довольно гибкой, все же не способна исчерпывающе представить в семантической публикации содержание источников. К тому же любая разметка текстов, в сущности, опирается на определенную, весьма субъективную интерпретацию содержания источников.

Таким образом, полноценная семантическая публикация не может ограничиться лишь репрезентацией источниковой информации, которая выделена исследователем в ходе разметки текстов документов. Необходимо также непосредственное представление в семантической сети смысла текстов исторических источников, для чего можно использовать один из так называемых «контролируемых естественных языков» (controlled natural languages)³³⁷. Тексты исторических источников на таком языке переводятся компьютером в формулы логики предикатов первого порядка, над которыми можно производить логический вывод, получая новые знания.

В настоящее время контролируемые естественные языки, служащие для представления знаний, стремительно развиваются. Для этих языков характерны жестко ограниченная семантика и упрощенный синтаксис. Можно назвать следующие контролируемые естественные языки: Controlled English to Logic Translation – CELT³³⁸, Processable English – PENG³³⁹, Computer-Processable Language – CPL³⁴⁰, а также Attempto Controlled English – ACE, который разрабатывается в университете Цюриха под руководством Норберта Фук-

³³⁶ См. также: Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения...; Варфоломеев А. Г., Иванов А. С. Модели структуры и содержания исторических источников...; Иванов А. С., Варфоломеев А. Г. Репрезентация содержания исторических источников в семантических сетях... и др.

³³⁷ Fuchs, N. E., ed. *Controlled Natural Language: Workshop on Controlled Natural Language, CNL 2009. Marettimo Island, Italy, June 2009. Revised Papers*. Berlin; Heidelberg: Springer-Verlag, 2010. (Lecture Notes in Computer Science, 5972); Rosner, M., Fuchs, N. E., eds. *CNL 2010: Second Workshop on Controlled Natural Languages. Pre-Proceedings of the Second Workshop on Controlled Natural Languages. Marettimo Island, Sicily, Italy, September 13–15, 2010*. <http://ceur-ws.org/Vol-622/> (07.08.2012).

³³⁸ Pease, A., Murray, W. An English to Logic Translator for Ontology-Based Knowledge Representation Languages. In: *Proceedings of the 2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering*. Beijing, 2003. Pp. 777–783.

³³⁹ Schwitter, R. Processing Coordinated Structures in PENG Light. In: Wang, D., Reynolds, M., eds. *AI 2011: Advances in Artificial Intelligence*. Berlin; Heidelberg, 2011. Pp. 658–667. (Lecture Notes in Computer Science, 7106).

³⁴⁰ Clark, P., Harrison, P., Jenkins, T., Thompson, J., Wojcik, R. Acquiring and Using World Knowledge Using a Restricted Subset of English. In: *The 18th International FLAIRS Conference (FLAIRS'05)*. 2005. <http://www.cs.utexas.edu/users/pclark/papers/flairs.pdf> (07.08.2012).

са³⁴¹. Attempto Controlled English – один из самых выразительных и распространенных в этом ряду, поэтому данный язык и был выбран авторами монографии для апробации возможностей репрезентации содержания источников в прототипе семантической публикации суб-комплекса смоленско-рижских актов из комплекса «Moscowitica–Ruthenica»³⁴².

Тексты на ACE могут быть переведены в DRS (Discourse Representation Structures) и, соответственно, напрямую трансформированы в формулы логики предикатов первого порядка³⁴³. Здесь следует отметить, что тексты на обычных языках не могут быть напрямую (автоматически) преобразованы в логические формулы вследствие неопределенности любого дискурса, поэтому смысл таких текстов может быть понят только при наличии определенного контекста. ACE выглядит совершенно естественным языком, хотя, на самом деле, это формальный язык, а именно – язык логики предикатов первого порядка, использующий синтаксис английского языка. Этот язык понятен и человеку, и компьютеру. В настоящее время Attempto Controlled English поддерживается рядом специализированных инструментов³⁴⁴, среди которых: синтаксический анализатор Attempto Parsing Engine (APE), переводящий тексты на ACE в DRS; модуль Attempto Reasoner (RACE), служащий для осуществления операций логического вывода на основе утверждений, записанных на ACE; редактор для корректировки текстов на ACE и др. Существенно и то, что для записи текстов на ACE могут использоваться различные специализированные словари. К тому же тексты исторических источников на ACE, обработанные при помощи модуля RACE, могут использоваться для генерирования новых гипотез на основе фактов, выявленных исследователем. Таким образом, Attempto Controlled English полностью отвечает задачам семантических публикаций исторических источников.

При этом неизбежно возникает вопрос о том, насколько приемлемо использование естественного контролируемого языка ACE для

³⁴¹ de Coi, J. L., Fuchs, N. E., Kaljurand, K., Kuhn, T. Controlled English for Reasoning on the Semantic Web. In: Bry, F., Małuszyński, J., eds. *Semantic Techniques for the Web: The Reverse Perspective*. Berlin; Heidelberg, 2009. Pp. 276–308. (Lecture Notes in Computer Science, 5500); Kuhn, T. How Controlled English can Improve Semantic Wikis. In: *Proceedings of the Fourth Workshop on Semantic Wikis: European Semantic Web Conference 2009*. [S.l.], 2009. Pp. 1–16. (SEUR Workshop Proceedings, vol. 464).

³⁴² Здесь можно отметить, что одна из Wiki-систем – AceWiki – использует ACE.

³⁴³ Kamp, H., Reyle, U. *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Publishers, 1993.

³⁴⁴ Attempto Tools. <http://attempto.ifi.uzh.ch/site/tools/> (13.08.2012).

создания семантической публикации исторических источников на древнерусском языке, ведь любой перевод оригинального текста на другой язык (это относится и к передаче древнерусских документов на современном русском языке) неизбежно опирается на интерпретацию содержания источника, что ведет к частичному изменению смысла текста, а так же к частичной утрате его содержательных аспектов, не говоря уже о стилистических и выразительных средствах выражения содержания, которые не могут быть сохранены в переводе. При этом передача даже англоязычных текстов на АСЕ не может избежать существенных упрощений вследствие жестких правил перевода текстов.

В этой связи можно отметить, что, во-первых, передача оригинальных текстов на языке АСЕ служит лишь для создания возможности многоаспектного семантического связывания информации ряда источников, включенных в семантическую сеть, и ни в коем случае не «замещает» оригинальный текст источника «суррогатным» текстом, поскольку тексты на АСЕ в семантической публикации должны быть соотнесены (связаны) с текстами на языке оригинала, представленными в рамках той же семантической сети. Т.е. АСЕ является всего лишь инструментом генерирования семантической сети. Во-вторых, тексты на АСЕ позволяют включить в систему семантических связей значительно больший объем источниковой информации, нежели схемы разметки SMW, TEI или CEI. В принципе, тексты источников, переданные на языке АСЕ, практически без потерь включаются в семантическую сеть, что значительно повышает эвристическую эффективность семантической публикации и возможности использования инструментария компьютерного источниковедения.

Для оценки применимости Attempto Controlled English для создания семантической публикации древнерусских грамот далее приводится текст одной из грамот (акт № 6), включенных в прототип семантической публикации, переданный на АСЕ.

Следует отметить, что при передаче текста на АСЕ необходимо придерживаться определенных правил, которые, в целом, выглядят довольно простыми. Так, в соответствии с этими правилами, любое существительное должно предварять определяющее слово (a, every, no, some и т.п.), за исключением имен собственных, которые пишутся с прописной буквы; могут использоваться также анафорические ссылки (He = Metropolitan), однако в других случаях такие ссылки запрещены (“Rigans are not guilty of that”) и т.д. Хотя Attempto Controlled English накладывает и другие существенные ог-

раничения на передачу текстов (например, глаголы могут использоваться только в настоящем неопределенном времени – the Present Indefinite Tense), тем не менее смысл текстов любых источников, в том числе и на древнерусском языке, может быть более-менее точно передан на АСЕ.

| Оригинальный текст в «палеографической» транскрипции | Перевод на язык АСЕ |
|---|--|
| † БЛГБАННЕ · ѿ мнтрофолнта рнзъкого · ко своему мнлому · снѣн князю келнкому феѳдору · н къ его дѣтемь · н ко влдѣуѣ н къ намѣстннкѹ · н ко всемъ боаромъ · | The metropolitan of Riga blesses the prince of Smolensk, the children of the prince of Smolensk, the bishop of Smolensk, the governor of Smolensk, and the boyars. |
| то бѹ дн тобе кедомо · про тѹю жалокѹ · что внтѣблѣне жалобннса на рнжаньн · чнмь то хотѣ ѿ пракнтнса протнкѹ гѣамнка · | The metropolitan of Riga demands that the prince of Smolensk knows about a complaint that is lodged against some inhabitants of Riga by some inhabitants of Vitebsk. The Metropolitan of Riga supposes that the inhabitants of Vitebsk want to justify their actions to Helmich. |
| н нхѣ слѡ тал коко · хотелн са темь словомь ѿ пракнтн · н ре клн такъ предъ княземь враньскѣмь · вѣ ехалн · н · тѣ мужь нзъ рнгѣ · н оубнлн ѹ ловека · н оузалн · і · тѣ вѣрковѣсковъ коск | The inhabitants of Vitebsk appeal to the Prince of Briansk. The inhabitants of Vitebsk say that 50 Rigans kill a man and grab 10 berkovets of wax. |
| н нѣне ѿ мнтрофолнтѣ тако молкю · как то внтѣблѣне неправдою · жалобннса · на рнжннѣ · | The metropolitan of Riga says that the inhabitants of Vitebsk complain on the inhabitants of Riga unjustly. |
| н нѣне то есть мнѣ кедомо · аже рнжане сѹтъ в томь невиноватн · | The metropolitan of Riga knows that the inhabitants of Riga are not guilty. |
| н нѣне ѿ тому · дн клѹса · аже тѹон намѣстннкѣ слѹшаетъ всакого ѹлѣка · слова · | The metropolitan of Riga is surprised by the credulity of the governor of Smolensk. |

| | |
|--|--|
| <p>А ТА ПРАВДА ЕСТЬ ПРОМѢЖИ ВАСЪ И НАСЪ • КДѢ СЯ ТАЖА ПОУНЕТЪ ТУ КО НЦАТИ</p> | <p>There is an agreement that is con- cluded by Riga and Smolensk. The agreement states that if a controver- sy arises then it must be settled by Riga and Smolensk.</p> |
| <p>И ПЪНІЕ А МОЛЮСА ВАМЪ • КАК ТО МО СТОАТИ ОУ ТОИ ПРАВДѢ • И ОУ КРѢТНОМЪ • УЕЛОКАНИИ •</p> | <p>The metropolitan of Riga demands that the agreement is observed.</p> |
| <p>АЖЕ ИМЕТЪ ЖАЛОБИТНСА ВАСЪ КТО НА РИЖАНИ • НАИ ГѢЛМИКО НАИ КТО ИЛИИ • И ВЪ ШАНТЕ К НАМЪ • А МЪ ПРАВДОУ ДА ПО ВЪЖИ ПРАВДѢ •</p> | <p>The agreement settles that if any- one complains-about Rigans, or Helmich, or somebody else then the prince of Smolensk must send a missive to the metropolitan of Riga. Then the metropolitan of Riga ren- ders justice in accordance with di- vine justice.</p> |

Таблица 2. Грамота архиепископа рижского смоленскому князю Федору Ростиславичу (акт № 6) на языке оригинала и на АСЕ.
ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3

Также и информация, относящаяся к источнику (факты об источнике), может быть переведена на АСЕ: «The charter_6 is written by the archbishop of Riga» (акт № 6 написан архиепископом рижским); «The charter_6 probably refers to the charter_4» (акт № 6, возможно, ссылается на акт № 4); «The charter_6 mentions Helmich and the prince of Briansk» (акт № 6 упоминает Гельмика и князя брянского) и т.п. К тому же АСЕ позволяет записывать аксиомы и правила логического вывода, например: «if something X is written by someone Y, then Y is the author of X». В результате большой объем фактического материала, относящегося к текстам источников, который вводится и обрабатывается вручную в ходе создания семантической публикации на основе, например, SMW, может быть частично обработан автоматически благодаря возможностям, которые предоставляет АСЕ. При этом, как уже отмечалось, новые знания – факты и гипотезы – об источниках и их текстах в перспективе могут быть получены на основе автоматического вывода. Поэтому язык АСЕ можно считать многообещающим инструментом современного компьютерного источниковедения.

Семантические сети для реконструкции исторической реальности (HSN)

Как и в общем историческом источниковедении, в компьютерном источниковедении завершающая стадия синтетических процессов – восполнение пробелов в свидетельствах отдельных источников и в информационном поле их комплексов, воссоздание утраченных фактов и создание алгоритма реконструкции исторической реальности в семантических сетях. Следует отметить, что эти задачи, частично, выходят за рамки компьютерного источниковедения, переходя в сферу исторического исследования. Поэтому в настоящем разделе будут рассмотрены лишь некоторые принципиальные подходы к конструированию семантических сетей, позволяющих на основе источниковой информации, представленной в таких сетях, решать вопросы реконструкции исторической реальности.

Принципиальная модель исторической семантической сети (Historical Semantic Network, далее – HSN)³⁴⁵ была разработана авторами монографии для поддержки исследований по региональной истории³⁴⁶. Очевидно, что подобная сеть должна интегрировать большое количество памятников истории и культуры, относящихся к различным типам исторических источников: письменные, изобразительные и аудиовизуальные источники (кинофотофонодокументы), источники устной истории, а также оцифрованные изображения и описания вещественных и этнографических источников. На первый взгляд, подобные базы данных не являются чем-то новым, так как репрезентация историко-культурного наследия на Web-сайтах практикуется библиотеками, архивами, научно-исследовательскими центрами и сообществами исследователей разных стран. Однако, за редким исключением, существующие в Интернете базы данных яв-

³⁴⁵ См. подробнее: Varfolomeyev, A. Soms, H., Ivanovs, A. Knowledge-Based Information Systems in Research of Regional History...; Варфоломеев А. Г., Иванов А. С., Сомс Г. Семантическая сеть как модель представления знаний по региональной истории. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 35, декабрь 2008. Москва; Барнаул, 2008. С. 158–160; Варфоломеев А., Иванов А., Сомс Х. Проект семантической сети для поддержки исследований по региональной истории. В кн.: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XVIII starptautisko zinātnisko lasījumu materiāli. Vēsture XII*. Daugavpils, 2009. 206.–211. lpp.; Varfolomeyev, A., Ivanovs, A. Knowledge-Based Scholarly Environment Project for Regional Historical Studies...; Варфоломеев А. Г., Иванов А. С. Модели для адекватной репрезентации неполной и/или противоречивой исторической информации...

³⁴⁶ Модель HSN разрабатывалась на основе существующей с 1994 года базы данных «Latgales Dati» <http://latgalesdati.du.lv/> (09.12.2012). Соответственно, цель HSN была первоначально определена следующим образом: накопление и репрезентация источниковой информации, необходимой для изучения истории Латгальского края – исторического региона в восточной части Латвии, а также предоставление историкам необходимого исследовательского инструментария для проведения аналитической и синтетической критики исторических источников, верификации их информации и получения нового знания о прошлом.

ляются ни чем иным, как «библиотеками», предоставляющими исследователям широкие возможности удаленного доступа к оцифрованным материалам. Для того чтобы базы данных стали не только хранилищами ретроспективной информации, но и эффективным исследовательским инструментом в рамках парадигмы семантического Web, необходимо решить целый ряд методологических, теоретических и практических проблем.

Наиболее сложные из них связаны не столько с извлечением информации из исторических источников, сколько со связыванием свидетельств разных источников, относящихся к одному и тому же историческому факту, оценкой надежности этой информации, а также с восполнением пробелов в источниковой информации. При этом неизбежно возникает вопрос о том, насколько математическая модель репрезентации источниковой информации в HSN соответствует характеру традиционной творческой работы историка с историческими источниками, поскольку существует вероятность, что операции с источниковой информацией в рамках HSN могут основываться на упрощенных представлениях о сущности исторического факта и исторического источника, а поэтому будут проводиться формально.

В основе модели HSN находятся не «исторические события»³⁴⁷, а реальные и исключительно многообразные «исторические объекты» – прежде всего, исторические источники различных типов, видов и разновидностей (письменные источники, разумеется, соответствующим образом размечаются для выделения их потенциальной информации), а также лица, исторические здания, организации, географические объекты и т.п.³⁴⁸

«Исторические объекты» разделяются на ныне существующие и уже исчезнувшие, о былом существовании последних можно судить только на основе свидетельств исторических источников. Поэтому,

³⁴⁷ В проекте HEML (Historical Events Markup and Linking), см.: <http://www.heml.org> (09.12.2012) предложен специальный XML-язык для репрезентации и связывания информации об исторических событиях. Исторические события характеризуются местом, временными рамками, участниками и описанием, для таких характеристик введены специальные теги. Однако недостатком подхода, предложенного в данном проекте, является центральная роль «исторического события» (event), которое является основой для структурирования информации. В действительности, «событие» есть лишь его описание, в сущности, – это интерпретация свидетельств исторических источников, зачастую неполных или противоречивых. При этом «событие» как таковое предстает как совокупность различно интерпретируемых свидетельств об определенных связях между объектами, включенными в исторический процесс и привязанными к некоторому моменту или промежутку времени.

³⁴⁸ С точки зрения методологии исторического исследования, понятие «исторические объекты» является сомнительным. Однако и в традиционном историческом дискурсе появляются понятия, которые не совсем вписываются в традиционную методологию истории. Так, «sites of memoгу» в известной и популярной концепции П. Нора довольно-таки точно соответствуют реальному содержанию «исторических объектов» в HSN.

бесспорно, ключевыми «историческими объектами» исторической семантической сети являются исторические источники, которые предоставляют информацию об остальных объектах. Именно на основе интерпретации информации исторических источников происходит реконструкция (а иногда – и конструирование) прочих объектов, их иерархическое и тематическое упорядочивание, а также выявление связей между объектами. При этом объектам семантической сети можно поставить в соответствие множество свойств (атрибутов) – имена, даты, место и т.д., – определяющих их свойства. Другая возможность – присваивать объектам только уникальные идентификаторы, а все остальное многообразие информации передавать связями между объектами, создавая таким образом семантическую сеть.

Что касается информации об исторических фактах (событиях), то она заключается, прежде всего, в связях между объектами, вовлеченными в семантическую сеть. В такой сети важную роль играют также специальные темпоральные объекты – моменты и периоды времени (timeline-objects). Синхронические и диахронические связи между рядом объектов, включая темпоральные объекты, указания на характер отношений и сами источники, в данной семантической сети являются моделью исторического факта, а точнее – представлением о факте-событии. Разнообразные связи между «историческими объектами» в несколько упрощенном виде представлены на рис. 4.6.

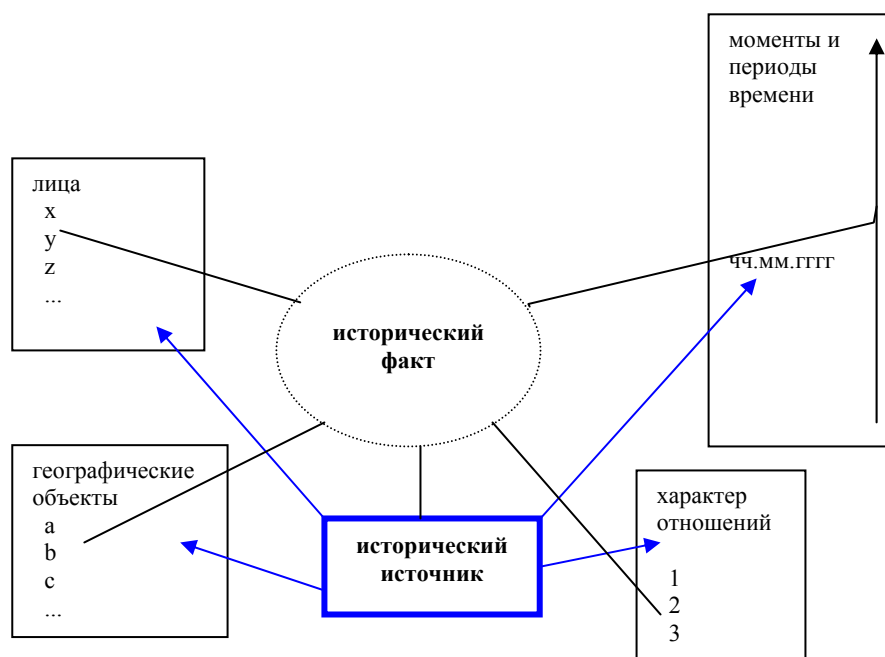


Рис. 4.6. Связи между «историческими объектами», описывающие исторический факт

Математической моделью HSN может служить гиперграф (рис. 4.7) или обычный граф, но эти представления не отражают существенной стороны концепции HSN – пространство событий не является конечным. Можно даже сказать, что оно не является дискретным, так как характеристики событий могут непрерывно меняться в зависимости от исходной информации и способов ее интерпретации. Также потенциально не являются конечными и множества объектов, которые остаются открытыми для добавления новых сущностей.

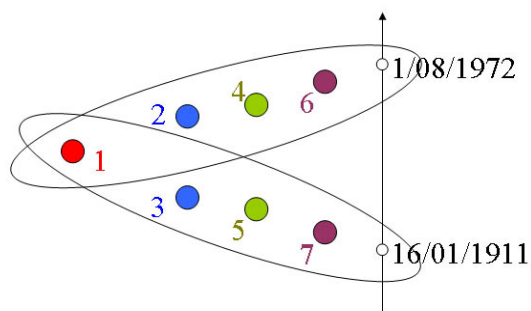


Рис. 4.7. Репрезентация исторического факта в виде гиперграфа:
 1 – лицо, 2 – дата смерти, 3 – дата рождения, 4 – место смерти,
 5 – место рождения, 6,7 – источники

В идеале HSN должна самостоятельно генерировать подобные образы исторических фактов, но при этом семантическая сеть должна предоставлять пользователю не одну версию исторического факта, а несколько, так как надежность и достоверность источниковой информации всегда находится под вопросом.

Это означает, что особенностью семантической сети является нечеткость практически всех аспектов информации, отраженной сетью. Причины нечеткости сходны с нечеткостью экспертных оценок. Свидетельства исторических источников могут быть, с одной стороны, недостаточно информативными, с другой – противоречивыми, или даже намеренно или ненамеренно ложными. В результате исторические факты, представленные в HSN как отношения «исторических объектов», будут логическими высказываниями разной степени истинности. Поэтому HSN должна в полной мере учитывать неопределенность свидетельств, извлекаемых из исторических источников.

Для выражения степени достоверности информации отношения семантической сети снабжаются весовыми переменными. Эти переменные могут быть обычными числами от 0 до 1, но более оправданным выглядит применение для них векторных значений (true,

false), где компоненты true и false независимо друг от друга принимают значения от 0 до 1. Итак, степень истинности может выражаться числом от 0 до 1 или же моделироваться нечеткой величиной, распределенной по некоторому интервалу числовой прямой (см. рис. 4.8).

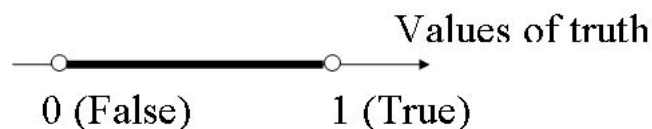


Рис. 4.8. Моделирование степени истинности высказываний

Для работы с такими логическими высказываниями, в частности, для реализации операций логического вывода одних высказываний из других, существуют различные варианты многозначных и нечетких логик. Есть среди них и такие логики, которые специально ориентированы на возможную противоречивость информации³⁴⁹. В них степень истинности высказывания выражается не одним числом, а парой чисел или двумерным вектором (true, false), то есть предполагается, что степень истинности высказывания A (которая может рассматриваться в нашем случае как степень подтверждения факта-события A свидетельствами внушающих доверие источников) не зависит от степени ложности этого высказывания (то есть от степени подтверждения источниками некоего факта-события, из истинности которого следует, что событие A не происходило). Например, если про одного и того же человека есть свидетельства двух достаточно достоверных источников о двух разных датах его рождения, то значение истинности для фактов-событий рождения данного человека, описанных в данных источниках, следует признать равным $(1,1)$, означающим, что информация о реальном факте-событии абсолютно противоречива. В то же время утверждение, что данный человек родился в некотором временном интервале, содержащем две указанные даты, следует признать имеющим истинностное значение $(1,0)$, то есть абсолютно истинным. Значение $(0,1)$ в таком случае будет соответствовать абсолютно ложному факту-

³⁴⁹ Представляется, что при моделировании HSN для адекватного отражения неполноты и противоречивости исторической информации было бы целесообразным использовать векторную логику с изменением значений истинности, созданную российским математиком Л.В. Аршинским: Аршинский Л. В. Содержательный и формальный выводы в логиках с векторной семантикой. *Автоматика и телемеханика*. 2007. № 1. С 153–162. См. также: Belnap, N. A Useful Four-Valued Logic. In: Dunn, J. M., Epstein, G., eds. *Modern Uses of Multiple-Valued Logic*. Dordrecht; Boston: Reidel, 1977. Pp. 8–37. (Series “Episteme”, vol. 2); Klir, J. R., Yuan, B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River (NJ), 1995.

событию, то есть имеющему только опровергающие его свидетельства, а значение $(0,0)$ – отсутствию какой-либо информации, подтверждающей или опровергающей факт-событие. Но степень достоверности источников тоже, на наш взгляд, может непрерывно меняться (например, от 0 до 1), а это значит, что для значений истинности событий тоже естественно предположить непрерывное изменение от $(0,0)$ до $(1,1)$ (см. рис. 4.9).

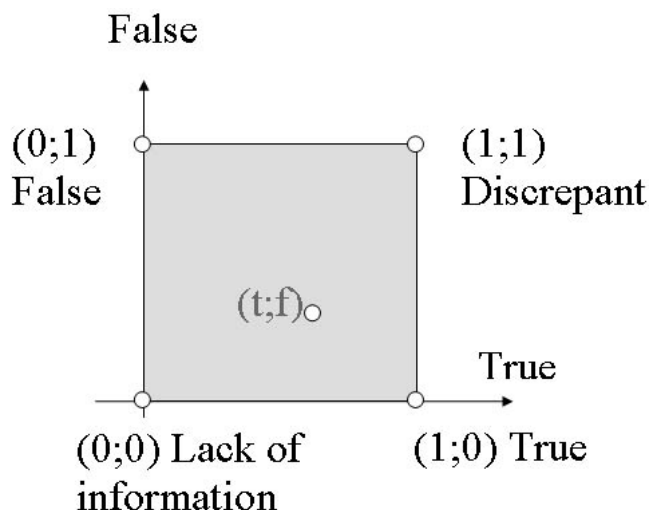


Рис. 4.9. Выражение степени истинности в векторной логике

Указанные характеристики могут задаваться пользователями, а также корректироваться в соответствии с новыми данными. Есть надежда, что вывод истинностных характеристик одних событий по другим может быть реализован и программным путем³⁵⁰. В любом случае, «измерение» истинности исторических данных производится под управлением пользователей семантической сети, результаты могут изменяться и дополняться, что отражает кооперацию исследователей в рамках сетевого сообщества.

В целом, концепция HSN предусматривает осуществление ряда операций над моделями событиями. Во-первых, необходимо реализовать поиск похожих событий (например, относящихся к одному и тому же человеку). Во-вторых, надо «научить» систему автоматически или с участием человека сравнивать свидетельства о похожих событиях, чтобы обнаруживать и оценивать противоречивость информации. В-третьих, нужна операция «снятия противоречий», создающая новое событие, обобщающее информацию различных источников. Наконец, для реализации сравнения событий необходимо предусмотреть предварительную операцию установления идентич-

³⁵⁰ Hähnle, R. *Automated Deduction in Multiple-Valued Logics*. Oxford, 1993.

ности объектов, отраженных в разных документах – степень этой идентичности тоже может непрерывно меняться от 0 до 1.

Векторная логика достаточно адекватно отражает противоречивый характер традиционного исторического синтеза – выявления исторических фактов на основе частично логических, частично – интуитивных интерпретаций свидетельств исторических источников. При этом HSN, использующая векторную логику, сможет предложить историку целый набор вариантов – «образов» исторических фактов; выбор же подходящего варианта целиком и полностью будет зависеть от историка, его компетентности, эрудиции и интуиции, степени проникновения в историческое прошлое.

Практическая работа по генерации исторической семантической сети лишь незначительно отличается от рассмотренного выше процесса создания семантической сети для репрезентации комплексов исторических источников. Так, при создании HSN требуется последовательно реализовать следующие операции с источниковой информацией: реконструкция, связывание и представление (визуализация³⁵¹). Реконструкция – это генерация взаимоотношений между объектами, которые напрямую могут быть не отражены в исторических документах. На этой стадии могут возникнуть новые (реконструированные реальные или гипотетические) исторические объекты. Связывание может быть определено как обнаружение сходных объектов, которые в рамках семантической сети (или в нескольких взаимосвязанных сетях) можно считать единым объектом; также – соединение отношений различных объектов в единое целое. Представление – это совокупность операций по поиску данных, представлению информации, содержащейся в сети, в удобном для исследователя виде и выводу результатов в форме таблиц, графиков, временных «линеек», тематических карт и др.

Для записи онтологий и представления семантических связей здесь также могут использоваться технологии RDF и OWL. Однако общепринятый подход, заключающийся в описании попарных связей сети в форме триплетов «субъект–предикат–объект», нуждается в некотором уточнении. Так, даже простейший исторический факт «человек X родился в городе Y в день Z» представляет собой четверку, а не тройку, поэтому он должен быть разбит на два триплета: «человек X родился в городе Y» и «человек X родился в день Z».

³⁵¹ См., например: Robertson, B. G. Visualizing an Historical Semantic Web with HEML. In: *Proceedings of the 15th International Conference on World Wide Web. WWW'06*. New York (NY), 2006. Pp. 1051–1052. doi:10.1145/1135777.1136010.

При этом исходный факт все-таки является единым, поскольку, по всей видимости, опирается на определенный фрагмент некоего исторического источника, в связи с чем конструкция должна стать иерархической: «фрагмент 1 говорит «человек X родился в городе Y» и «человек X родился в день Z»». В результате структура RDF может оказаться довольно громоздкой для описания связей между объектами, число которых может намного превосходить 2: например, простейший исторический факт «человек X родился в городе Y в день Z» связывает между собой три объекта.

Еще одной серьезной проблемой является объединение нескольких HSN-систем в единое информационное пространство. Такое объединение будет способствовать сотрудничеству исследователей, предоставляя удаленный доступ к историческим источникам и полученным на их основе данным, создавая условия для развития сетевого сообщества историков и источниковедов. Кроме того, объединение семантических сетей в единую сеть придает историческим знаниям намного большую актуальность и востребованность. И главное, в полной мере возможности семантической сети могут проявиться только во взаимодействии разных систем исторических объектов, и прежде всего систем, нацеленных на репрезентацию комплексов исторических источников, свидетельства которых и являются основой для генерации специализированных исторических семантических сетей.

Заключение

В монографии обобщен двенадцатилетний опыт совместной работы авторов по применению компьютерных технологий для репрезентации и изучения средневековых письменных источников по истории России. Теоретико-методологическое осмысление и практическая апробация базовых методов и подходов компьютерного источниковедения в ходе осуществления ряда пилотных проектов еще раз подтверждают тезис, что современные компьютерные технологии на новом уровне воспроизводят операции, входящие в традиционный «арсенал» источниковеда, расширяют источниковую базу исторических исследований и набор параметров для источниковедческого анализа и синтеза, но при этом принципиально не меняют ни логику источниковедческого исследования, ни содержание его методов. Это означает, что способы изучения исторических источников в классическом источниковедении ничем принципиально не отличаются от технологий репрезентации и обработки источниковой и внеисточниковой информации в компьютерном источниковедении, при этом любая исследовательская операция исторического источниковедения, все без исключения этапы аналитической и синтетической критики источников находят свой аналог в компьютерном источниковедении. Поэтому можно утверждать, что компьютерное источниковедение, в определенном смысле, «вторично» – основой его развития является методология традиционного исторического источниковедения, прежде всего, положения об информационном потенциале исторических источников и способах раскрытия этого потенциала, о комплексном подходе к изучению памятников прошлого. В то же время компьютерное источниковедение существенно расширяет инструментарий исторического источниковедения, снимает некоторые естественные ограничения на применение классических методов источниковедения и других вспомогательных исторических дисциплин.

Вследствие ограниченного объема в монографии рассмотрены лишь отдельные направления применения компьютерных технологий в источниковедческих исследованиях. За рамками работы остались, например, многие вопросы использования количественных методов в источниковедческих исследованиях, компьютерного распознавания текстов исторических источников и др. Следует при этом отметить, что авторы работы стремились выявить преимуще-

ства и обозначить перспективы *системного* применения методов компьютерного источниковедения в комплексном изучении исторических источников. Системность в применении технологий компьютерного источниковедения означает, что все они должны быть ориентированы на достижение конечной цели источниковедческой критики – обеспечение информационной основы для полноценного проведения синтетических операций, реконструкции аспектов исторической действительности. В монографии была сделана попытка показать, как достижению этой цели могут способствовать ключевые этапы по репрезентации и обработке источниковой и внеисточниковой информации в компьютерном источниковедении: создание информационных моделей структуры и содержания исторических источников на основе многоуровневой разметки, системное агрегирование и семантическое связывание источниковой информации и метаданных, репрезентация источников в семантических сетях, верификация источниковой информации и реконструкция аспектов исторической реальности в исторической семантической сети.

Концептуальные подходы к системному использованию современных компьютерных технологий в источниковедении, рассмотренные в монографии, могут быть в полной мере реализованы в источниковедческой практике лишь в том случае, если в семантических сетях будут представлены соответствующим образом размеченные большие комплексы исторических источников. Едва ли эту задачу могут решить в рамках локальных проектов небольшие коллективы специалистов. Поэтому перспективы развития компьютерного источниковедения как направления общего исторического источниковедения связаны не столько с совершенствованием компьютерных технологий как таковых, сколько с возможностями развития инфраструктуры источниковедческих исследований, прежде всего, – с появлением сетевых сообществ исследователей исторических источников.

Библиография

1. *Акты Московского государства, изданные Императорской Академией Наук. Том II: Разрядный приказ. Московский стол. 1635–1659* / Под ред. Н. А. Попова. Санкт-Петербург, 1894. хiv, 773 с.

2. *Акты Московского государства, изданные Императорской Академией Наук. Том III: Разрядный приказ. Московский стол. 1660–1664* / Под ред. Д. Я. Самоквасова. Санкт-Петербург, 1901. ххх, 674 с.

3. Аршинский Л. В. Содержательный и формальный выводы в логиках с векторной семантикой. *Автоматика и телемеханика*. 2007. № 1. С 153–162.

4. Бабалык М. Г., Варфоломеев А. Г. О применении XML-технологий в текстологическом анализе списков древнерусского апокрифа «Беседа трех святителей». В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 35. *Материалы XI конференции АИК. Москва, 13–15 декабря 2008 г.* Москва; Барнаул: Азбука, 2008. С. 51–52.

5. Баранов В. А. Полнотекстовая коллекция славянских евангелий проекта «Манускрипт» и специализированные инструменты разметки: модуль фрагментирования. *Вестник Пермского университета. История*. 2011. № 2 (16). С.40–47.

6. Баранов В. А., Вотинцев А. А., Гнутиков Р. М., Миронов А. Н., Романенко В. А. Специализированный текстовый редактор «Манускрипт» системы обработки древних рукописей. В кн. *Информационный бюллетень Ассоциации «История и компьютер»*. № 31, сентябрь 2003. Москва, 2003. С.159–165.

7. Баранов В. А., Дубовцев С. В. Электронное критическое издание средневекового славянского текста: модель данных и визуализация лингвистических единиц. *Интеллектуальные системы в производстве*. 2010. № 1. С. 280–287.

8. Беленький И. Л. Источниковедение историческое. В кн.: *Источниковедение: Проблемные лекции. Учебно-методический модуль* / Отв. ред. О. М. Медушевская; РГГУ. ИАИ. Кафедра источниковедения и вспомогательных исторических дисциплин. Москва: Издательство Ипполитова, 2005. С. 100–112. (Серия «Я иду на занятия...»).

9. Бобиньская Ц. Пробелы в источниках. Методологический анализ. *Вопросы истории*. 1965. № 6. С. 76–86.

10. Бородкин Л. И. Историческая информатика: Этапы развития. *Новая и новейшая история*. 1997. № 1. С. 3–22.

11. Бородкин Л. И. Квантитативная история в системе координат модернизма и постмодернизма. *Новая и новейшая история*. 1998. № 5. С. 3–16.

12. Бородкин Л. И. Клиометрика: pro et contra (виртуальный диалог). В кн.: *Экономическая история. Обзорение* / Центр экономической истории при Историческом факультете МГУ им. М. В. Ломоносова; под ред. Л. И. Бородкина. Вып. 7. Москва, 2001. С. 114–132.

13. Бородкин Л. И. Математические методы классификации древних текстов. В кн.: *Методы количественного анализа текстов нарративных источников: Сб. статей* / АН СССР, Институт истории СССР; отв. ред. Б. М. Клосс. Москва: Институт истории СССР, 1983. С. 8–30.

14. Бородкин Л. И. Методология моделирования в зарубежной клиометрике: новые тенденции. В кн.: *Проблемы исторического познания: Материалы международной конференции. Москва, 19–21 мая 1996 г.* / Институт всеобщей истории РАН; отв. ред. Г. Н. Севостьянов. Москва: Наука, 1999. С. 130–140.

15. Бородкин Л. И. Приоритеты современной исторической информатики: технологии e-Science. В кн.: *Круг идей: Междисциплинарные подходы в исторической информатике: Труды X конференции Ассоциации «История и компьютер»* / Ред. Л. И. Бородкин, И. М. Гарскова. Москва, 2008. С. 5–15.

16. Бородкин Л. И., Милов Л. В. О некоторых аспектах автоматизации текстологического исследования (Закон Судный людям). В кн.: *Математические методы в историко-экономических и историко-культурных исследованиях: [Сб. статей]* / АН СССР, Отделение истории; отв. ред. И. Д. Ковальченко. Москва: Наука, 1977. С. 230–279.

17. Вагин В. Н., Головина Е. Ю., Загорянская А. А., Фомина М. В. *Достоверный и правдоподобный вывод в интеллектуальных системах*. Москва: Физматлит, 2004. 703 с.

18. Валк С. Н. Археографическая «легенда». В кн.: *Археографический ежегодник за 1971 год* / Отв. ред. С. О. Шмидт. Москва: Наука, 1972. С. 5–16.

19. Владимиров В. Н., Цыб С. В. Источниковедение в век компьютера (вместо предисловия). В кн.: *Источник. Метод. Компьютер: Сборник научных трудов* / Отв. ред. В. Н. Владимиров, С. В. Цыб. Барнаул: Издательство Алтайского государственного университета, 1996. С. 3–7.

20. Варфоломеев А. Г. Алгоритмическое и программное обеспечение работы сетевых сообществ исследователей текстов. В кн.: *Компьютерные науки и технологии. Часть 1: Сборник трудов первой Международной научно-технической конференции*. Белгород: ГиК, 2009. С. 30–33.

21. Варфоломеев А. Г., Бабалык М. Г., Пигин А. В. О проекте виртуальной среды для исследования списков «Беседы трех святителей». В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XII Всероссийской научной конференции “RCDL’2010” (Казань, 13–17 октября 2010 г.)*. Казань, 2010. С. 551–556.

22. Варфоломеев А. Г., Иванов А. С. Виртуальные среды для научных исследований в арсенале историка: тенденции и перспективы. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 36, октябрь 2010 г. Специальный выпуск: *Материалы XII конференции АИК «Информационные ресурсы, технологии и модели реконструкции исторических процессов и явлений»*. Москва, 22–24 октября. Москва: Издательство Московского университета, 2010. С. 5–7.

23. Варфоломеев А. Г., Иванов А. С. Модели для адекватной репрезентации неполной и/или противоречивой исторической информации в исторической семантической сети: К постановке проблемы. В кн.: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XIX starptautisko zinātnisko lasījumu materiāli. Vēsture XIII = Proceedings of the 19th International Scientific Readings of the Faculty of Humanities. History XIII / Daugavpils Universitāte; atb. red. I. Saleniece*. Daugavpils: Saule, 2010. 210.–216. lpp.

24. Варфоломеев А. Г., Иванов А. С. Модели структуры и содержания исторических источников. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 37. Специальный

выпуск: *Труды международной конференции «Компьютерные технологии и математические методы в исторических исследованиях»* (Петрозаводск, 11–16 июля 2011 г.). Петрозаводск, 2011. С. 25–31.

25. Варфоломеев А. Г., Иванов А. С. Принципы электронных публикаций комплексов исторических документов со средствами палеографического, текстологического и дипломатического анализа. В кн.: *Современные информационные технологии и письменное наследие: от древних текстов к электронным библиотекам. El'Manuscript-08: Материалы Международной научной конференции. (Казань, 26–30 августа 2008 г.)*. Казань: Издательство Казанского университета, 2008. С. 60–63.

26. Варфоломеев А. Г., Иванов А. С. Семантические публикации информации исторических источников на основе технологии Wiki. В кн.: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XX starptautisko zinātnisko lasījumu materiāli. Vēsture XIV = Proceedings of the 20th International Scientific Readings of the Faculty of Humanities. History XIV / Daugavpils Universitāte; atb. red. I. Saleniece. Daugavpils: Saule, 2011. 339.–347. lpp.*

27. Варфоломеев А. Г., Иванов А. С. Семантические публикации комплексов исторических источников. В кн.: *Информационные технологии и письменное наследие. El'Manuscript-10: Материалы международной научной конференции (Уфа, 28–31 октября 2010 г.) / Отв. ред. В. А. Баранов. Уфа; Ижевск: Вагант, 2010. С. 42–46.*

28. Варфоломеев А. Г., Иванов А. С. Технология XML как инструмент компьютерного источниковедения: (на примере изучения комплекса средневековых документов по истории Динабурга). В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. №32, апрель 2004: *Специальный выпуск: Материалы IX конференции Ассоциации «История и компьютер»*. Подмосковье, 22–25 апреля 2004 г. Москва; Томск: Издательство Томского университета, 2004. С. 137–139.

29. Варфоломеев А. Г., Иванов А. С. Технология XML: Современная реализация источникориентированного подхода в работе с комплексами исторических документов. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. Специальный выпуск, №34, май 2006: *Материалы X конференции Ассоциации «История и компьютер»*. Москва, 12–14 мая 2006 г. Москва; Тамбов: Издательство Тамбовского университета, 2006. С. 61–62.

30. Варфоломеев А. Г., Иванов А. С. XML-технологии в анализе информации полнотекстовых баз данных. В кн.: *Vēsture: Avoti un cilvēki. Starptautiskās konferences «XV Zinātniskie lasījumi» materiāli. Vēsture IX = Proceedings of the 15th International Scientific Readings. History IX / Daugavpils Universitāte; atb. red. H. Soms. Daugavpils: Saule, 2006. 164.–174. lpp.*

31. Варфоломеев А., Иванов А., Сомс Х. Проект семантической сети для поддержки исследований по региональной истории. В кн.: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XVIII starptautisko zinātnisko lasījumu materiāli. Vēsture XII / Daugavpils Universitāte; atb. red. H. Soms. Daugavpils: Saule, 2009. 206.–211. lpp.*

32. Варфоломеев А. Г., Иванов А. С., Сомс Г. Семантическая сеть как модель представления знаний по региональной истории. В кн.: *Информационный бюллетень Ассоциации «История и компьютер». № 35, декабрь 2008. Специальный выпуск: Материалы XI конференции АИК «Инновационные подходы в исторических исследованиях: информационные технологии, модели и методы». Москва, 13–15 декабря 2008 г. Москва; Барнаул: Азбука, 2008. С. 158–160.*

33. Варфоломеев А. Г., Каргинова Н. В., Кравцов И. В., Москвин Н. Д. Применение RuleML для представления и вывода знаний о семантической структуре фольклорных текстов, полученных на основе их теоретико-графовых моделей. В кн.: *Одиннадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2008, 28 сентября – 3 октября 2008 г., г. Дубна, Россия): Труды конференции. Т. 2. Москва: ЛЕНАНД, 2008. С. 183–191.*

34. Варфоломеев А. Г., Кравцов И. В. Аналитические Web-публикации исторических документов. В кн.: *Научный сервис в сети Интернет: многоядерный компьютерный мир. 15 лет РФФИ: Труды Всероссийской научной конференции (24–29 сентября 2007 г., г. Новороссийск). Москва: Издательство МГУ, 2007. С. 389–390.*

35. Варфоломеев А. Г., Кравцов И. В. Использование технологии XML для публикации методики и результатов исследования текстов исторических источников. В кн.: *Информационный бюллетень Ассоциации «История и компьютер». Специальный выпуск, № 34, май 2006: Материалы X конференции Ассоциации «История и компьютер». Москва, 12–14 мая 2006 г. Москва; Тамбов: Издательство Тамбовского университета, 2006. С. 63–64.*

36. Варфоломеев А. Г., Кравцов И. В. Приобретение и представление знаний в сетевом сообществе исследователей текстов. В кн.: *Вторая конференция «Системный анализ и информационные технологии» САИТ-2007, Обнинск, 10–14 сентября 2007 г.: Тр. конф.: в 2 т.* Т. 1. Москва, 2007. С. 104–106.

37. Варфоломеев А. Г., Кравцов И. В., Филатов В. О. SVG-визуализация в цифровых библиотеках рукописных документов. В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Девятой Всероссийской научной конференции RCDL'2007. (Переславль-Залесский, Россия, 14–18 октября 2007 г.)*. Переславль-Залесский: Издательство «Университет города Переславля», 2007. С. 230–235.

38. Варшавчик М. А. Вопросы логики исторического исследования и исторический источник. *Вопросы истории*. 1968. № 10. С. 76–89.

39. Волков С. С. *Лексика русских челобитных XVII века: Форумляр, традиционные этикетные и стилевые средства*. Ленинград: Издательство Ленинградского университета, 1974. 164 с.

40. Гагарина Д. А., Кирьянов И. К., Корниенко С. И. Историко-ориентированные информационные системы: опыт реализации «Пермских» проектов. *Вестник Пермского университета. История*. 2011. Вып. 2 (16). С. 35–39.

41. Гагарина Д. А., Корниенко С. И. Рукописные и старопечатные кириллические книги в Интернете. *Вестник Пермского университета. История*. 2009. Вып. 3 (10). С. 112–118.

42. Гарскова И. М. *Базы и банки данных в исторических исследованиях*. Москва: МГУ; Göttingen: Max-Planck-Institut für Geschichte, [1994]. 215 с. (Серия «Новые учебные пособия по историческим дисциплинам»).

43. Гарскова И. М. Источниковедческие проблемы исторической информатики. *Российская история*. 2010. № 3. С. 151–161.

44. Гарскова И. М. Основные направления развития исторической информатики в конце XX – начале XXI века. *Вестник Московского государственного университета*. 2010. № 6. С. 75–103.

45. Григорьева И. В. *Источниковедение новой и новейшей истории стран Европы и Америки: Учебник для студентов вузов, обу-*

чающихся по специальности «История». Москва: Высшая школа, 1984. 335 с.

46. Демидова Н. Ф. *Служилая бюрократия в России XVII в. и ее роль в формировании абсолютизма*. Москва: Наука, 1987. 228 с.

47. Дзярновіч А. Гісторыя архіўнага комплексу Moscowitica–Ruthenica Рыжскага магістрата. В кн.: *Віцебска-Рыжскія акты XIII–XVII ст.* Вып. I: *Дакументы гаспадарча-гандлёвыя, XV–XVI ст.* Мінск: АТНЕНАЕUM, 2005. С. 10–19.

48. Добрушкин Е. М. Археографическая теория и источниковедческая практика (к вопросу о взаимосвязи). В кн.: *Россия в IX–XX веках: Проблемы истории, историографии и источниковедения. Сборник статей и тезисов докладов Вторых чтений, посвященных памяти А. А. Зимина* / Институт российской истории РАН; РГГУ. Москва: Русский мир, 1999. С. 129–132.

49. *Дополнения к Тому III-му Дворцовых Разрядов, издаваемых по высочайшему повелению II-м Отделением Собственной Его Императорского Величества канцелярии*. Санкт-Петербург, 1854. 484 столб.

50. Дройзен И. Г. *Историка* / Пер. с немецкого Г. И. Федоровой. Санкт-Петербург: Владимир Даль, 2004. 583 с. (Серия «Historica»).

51. Елизаров А. И., Есипова В. А., Калайда В. Т., Климкин В. М. Кластеризация филиграней по численным характеристикам изображений. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 31, сентябрь 2003. Москва, 2003. С. 223–234.

52. Енш Г. «Polonica» в Рижском городском государственном архиве. В кн.: *Ученые записки Института славяноведения АН СССР*. Т. II. Москва; Ленинград: Издательство АН СССР, 1950. С. 311–325.

53. Ерошкин Н. П. *История государственных учреждений до-революционной России*. Изд. 2-е, исправленное и дополненное. Москва: Высшая школа, 1968. 368 с.

54. Зимин А. А. *Методика издания древнерусских актов* / Отв. ред. Н. В. Устюгов. Москва: МГИАИ, 1959. 64 с.

55. Зиновьева Е. И. *Записные кабальные книги Московского государства XVI–XVII веков: Структура. Лексика. Фразеология*. Санкт-Петербург: Издательство СПбГУ, 2000. 236 с.

56. Иванов А. С. Археография и актовое источниковедение в эпоху компьютерных технологий. В кн.: *Проблемы дипломатики, кодикологии и актовой археографии: Материалы XXIV Международной научной конференции. Москва 2–3 февраля 2012 г.* / Отв. ред. Ю. Э. Шустова. Москва: РГГУ, 2012. С. 321–323.

57. Иванов А. С. База данных как динамическая модель истории региона. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*. № 30: *Специальный выпуск: Материалы VIII конференции Ассоциации «История и компьютер» в Санкт-Петербурге, 26–29 июня 2002 г.* Москва, 2002. С. 170–172.

58. Иванов А. С. Источниковедческие и археографические аспекты реконструкции исторических комплексов источников (отдел «Moscovitica–Ruthenica» в бывшем архиве Рижского магистрата). В кн.: *Сословия, институты и государственная власть в России. (Средние века и раннее Новое время): Сборник статей памяти академика Л. В. Черепнина* / Институт всеобщей истории Российской академии наук; отв. ред. В. Л. Янин, В. Д. Назаров. Москва: Языки славянских культур, 2010. С. 97–105. (Серия «Studia Philologica»).

59. Иванов А. С. Источнико-ориентированные базы данных в изучении локальной истории. В кн.: *XI Zinātnisko lasījumu materiāli: Vēstures sekcija. 5. krājums* / Daugavpils Pedagoģiskā universitāte. Daugavpils: Saule, 2001. 37.–42. lpp.

60. Иванов А. К вопросу о повышении информативной отдачи документов приказного делопроизводства в исторических исследованиях. В кн.: *Проблемы источниковедения и историографии: Сборник научных трудов* / Отв. ред. В. И. Гальцов. Калининград: Издательство Калининградского государственного университета, 2001. С. 22–27.

61. Иванов А. С. Комплекс источников по истории Великого Новгорода в Латвийском государственном историческом архиве (XII–XVI вв.). В кн.: *Новгородика-2010. Вечевой Новгород: Материалы международной научно-практической конференции 20–22 сентября 2010 г.* Часть 2. Великий Новгород, 2011. С. 89–103.

62. Иванов А. С. Комплексное источниковедение региональной истории. В кн.: *Научные труды Рижского института мировой экономики*. Вып. 2. Рига: Варты, 1998. С. 71–75.

63. Иванов А. С. Компьютерное источниковедение. В кн.: *Информационный бюллетень Ассоциации «История и компьютер»*.

№ 37. *Специальный выпуск: Труды международной конференции «Компьютерные технологии и математические методы в исторических исследованиях» (Петрозаводск, 11–16 июля 2011 г.)*. Петрозаводск, 2011. С. 56–62.

64. Иванов А. С. Методологические проблемы создания базы данных для региональных исследований: (Пример Латгальского края). В кн.: *Новые информационные ресурсы и технологии в исторических исследованиях и образовании: Сборник тезисов докладов и сообщений Всероссийской конференции. Подмосковье, «Боровое», 6–9 апреля 2000 г.* Москва, 2000. С. 71–72.

65. Иванов А. С. «Moscovitica–Ruthenica» в Латвийском государственном историческом архиве: история формирования комплекса, состав и введение в научный оборот. *Древняя Русь: Вопросы медиевистики*. 2004. № 3 (17). С. 47–54; № 4 (18). С. 94–106.

66. Иванов А. С. Проблемы введения в научный оборот обширных комплексов источников: коллекция «Moscovitica–Ruthenica» в Латвийском государственном историческом архиве. В кн.: *XXI век: Актуальные проблемы исторической науки. Материалы международной научной конференции, посвященной 70-летию Исторического факультета БГУ. Минск, 15–16 апреля 2004 г.* / Белорусский государственный университет, Ассоциация «История и компьютер», Белорусская психоисторическая ассоциация; отв. ред. В. Н. Сидорцов. Минск: Белорусский государственный университет, 2004. С. 100–101.

67. Иванов, А. С. Работа с XML-документом как воспроизведение основных этапов источниковедческой критики: новые технологии и возможность коррекции традиционных подходов. В кн.: *Информационный бюллетень Ассоциации «История и компьютер». Специальный выпуск. № 34, май 2006: Материалы X конференции Ассоциации «История и компьютер». Москва, 12–14 мая 2006 г.* Москва; Тамбов: Издательство Тамбовского университета, 2006. С. 66–67.

68. Иванов А. С., Варфоломеев А. Г. Идеи И. Д. Ковальченко об информационной ценности и неисчерпаемости источника в современном компьютерном источниковедении. В кн.: *Идеи академика И. Д. Ковальченко в XXI веке. Материалы IV научных чтений памяти академика И. Д. Ковальченко. Москва, МГУ им. М. В. Ломоносова. 10 декабря 2008 г.* / МГУ им. М. В. Ломоносова; Отделение

историко-филологических наук РАН. Москва: Издательство Московского университета, 2009. С. 154–164.

69. Иванов А. С., Варфоломеев А. Г. Использование технологии XML для введения в научный оборот комплекса документов *Moscovitica–Ruthenica*. В кн.: *RCDL'2004. Электронные библиотеки: перспективные методы и технологии, электронные коллекции = Digital Libraries: Advanced Methods and Technologies, Digital Collections. Труды Шестой Всероссийской научной конференции. Пушино, Россия. 29 сентября – 1 октября 2004 г.* Пушино, 2004. С. 285–289.

70. Иванов А. С., Варфоломеев А. Г. О перспективах и пределах применения методов компьютерного источниковедения в медиевистике (на примере комплексного изучения коллекции «*Moscovitica–Ruthenica*»). *Древняя Русь: Вопросы медиевистики*. 2005. № 3 (21). С. 34–35.

71. Иванов А. С., Варфоломеев А. Г. Проект издания комплекса документов «*Moscovitica–Ruthenica*». В кн.: *Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам: Материалы международной научной конференции (Ижевск, 13–17 июля 2006 г.)* / Отв. ред. В. А. Баранов. Ижевск: Издательство Ижевского государственного технического университета, 2006. С. 65–70.

72. Иванов А. С., Варфоломеев А. Г. Публикация и анализ рукописных исторических документов с помощью технологии XML. *Humanitāro Zinātņu Vēstnesis / Daugavpils Universitāte*. 2009, Nr. 16. 69.–84. lpp.

73. Иванов А. С., Варфоломеев А. Г. Репрезентация содержания исторических источников в семантических сетях: Attempto Controlled English (ACE). В кн.: *Информационные технологии и письменное наследие. El'Manuscript–2012: Материалы IV международной научной конференции. Петрозаводск, 3–8 сентября 2012 года* / Отв. ред. В. А. Баранов, А. Г. Варфоломеев. Петрозаводск; Ижевск: Петрозаводский государственный университет; Ижевский государственный технический университет им. М. Т. Калашникова, 2012. С. 112–120.

74. Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения (на примере формулярного анализа документов приказного делопроизводства). В кн.: *Круг идей: Алгоритмы и технологии исторической информатики. Труды*

IX конференции Ассоциации «История и компьютер» / Ред. Л. И. Бородкин, В. Н. Владимиров. Москва; Барнаул: Издательство Алтайского университета, 2005. С. 241–281.

75. Иванов А. С., Кузнецов А. М. Динабург в документах времен царя Алексея Михайловича. В кн.: *Динабург, Двинск, Даугавпилс в истории, культуре, литературе*. Даугавпилс: Центр русской культуры (Дом Каллистратова), 1999. С. 9–46. (Серия «Двинская старина». Вып. 1).

76. Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов (1656–1666)* = Ivanovs, A., Kuzņecovs, A. *Dinaburga Krievijas Valsts seno aktu arhīva dokumentos (1656–1666)*. Часть 1. Daugavpils: Saule, 2002. 199 lpp.

77. Иванов А. С., Кузнецов А. М. *Динабург в документах Российского государственного архива древних актов (1656–1666)* = Ivanovs, A., Kuzņecovs, A. *Dinaburga Krievijas Valsts seno aktu arhīva dokumentos (1656–1666)*. Часть 2. Daugavpils: Saule, 2002. 198 lpp.

78. Иванов А. С., Кузнецов А. М. Динабург в начальный период Первой Северной войны (1655–1656). В кн.: *Славянские чтения*. II. Даугавпилс; Резекне: Издательство Латгальского культурного центра, 2002. С. 222–236.

79. Иванов А. С., Кузнецов А. М. Комплексный подход к изучению документов приказного делопроизводства. *Древняя Русь: Вопросы медиевистики*. 2003. № 4 (14). С. 30–31.

80. Иванов А., Кузнецов А. *Смоленско-рижские акты: XIII в. – первая половина XIV в.: Документы комплекса Moscovitica–Ruthenica об отношениях Смоленска и Риги* = Ivanovs, A., Kuzņecovs, A. *Smolenskās–Rīgas aktis: 13.gs.–14.gs. pirmā puse: Kompleksa Moscovitica–Ruthenica dokumenti par Smolenskās un Rīgas attiecībām*. [Rīga]: Latvijas Valsts vēstures arhīvs; Latvijas Arhīvistu biedrība; Daugavpils Universitātes Latgales pētniecības institūts, 2009. 768 с.

81. Илизаров Б. С. О понятии ценности документального исторического источника. *История СССР*. 1986. № 5. С. 104–111.

82. Ильин Н., Киселев С., Рябышкин В., Танков С. Технологии извлечения знаний из текста. *Открытые системы. СУБД*. 2006. № 6. <http://www.osp.ru/os/2006/06/2700556/> (09.12.2012).

83. Илюшенко М. П., Кузнецова Т. В., Лившиц Я. З. *Документоведение. Документ и системы документации: Учебное пособие*. Москва: МГИАИ, 1977. 84 с.

84. *Информационные технологии для историков: Учебное пособие к практикуму по курсу «Информатика и математика»* / Отв. ред. Л. И. Бородкин. Москва: МГУ, 2006. 236 с.

85. *Историческая информатика: Учебное пособие* / Е. Б. Белова, Л. И. Бородкин, И. М. Гарскова, Т. Ф. Измestьева, В. В. Лазарев; под ред. Л. И. Бородкина и И. М. Гарсковой. [Москва]: Мосгорархив, 1996. 400 с.

86. *История делопроизводства в СССР: Учебное пособие* / Ред. Я. З. Лившиц, В. А. Цикулин. Москва: МГИАИ, 1974. 169 с.

87. *Источниковедение: Теория. История. Метод. Источники российской истории: Учебное пособие для гуманитарных специальностей* / И. Н. Данилевский, В. В. Кабанов, О. М. Медушевская, М. Ф. Румянцева. Москва: РГГУ, 1998. 702 с.

88. Каргинова Н. В., Кравцов И. В., Москин Н. Д., Варфоломеев А. Г. Проект электронной библиотеки методик и результатов исследований текстовых коллекций для системы «Источник». В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Десятой Всероссийской научной конференции «RCDL'2008» (Дубна, Россия, 7–11 октября 2008 г.)*. Дубна: ОИЯИ, 2008. С. 239–245.

89. Каштанов С. М. *Актовая археография*. Москва: Наука, 1998. 318 с.

90. Каштанов С. М. Дипломатика как специальная историческая дисциплина. *Вопросы истории*. 1965. № 1. С. 39–44.

91. Каштанов С. М. Дипломатический состав древнерусского акта. В кн.: *Вспомогательные исторические дисциплины*. II. Ленинград: Наука, 1969. С. 143–159.

92. Каштанов С. М. *Очерки русской дипломатики* / Отв. ред. А. А. Зимин. Москва: Наука, 1970. 502 с.

93. Каштанов С. М. *Русская дипломатика*. Москва: Высшая школа, 1988. 231 с. (Серия «Библиотека историка»).

94. Кирьянов И. К., Корниенко С. И. Стенографические отчеты Государственной Думы начала XX века: от традиционного к компь-

ютерному источниковедению. *Вестник Челябинского государственного университета*. 2009. № 6 (144). (История, вып. 30). С. 144–148.

95. Ковальченко И. Д. Исторический источник в свете учения об информации. (К постановке проблемы). *История СССР*. 1982. № 3. С. 129–148.

96. Ковальченко И. Д. *Методы исторического исследования*. Москва: Наука, 1987. 439 с.

97. Ковальченко И. Д. О моделировании исторических явлений и процессов. *Вопросы истории*. 1978. № 8. С. 72–93.

98. Ковальченко И. Д. О применении математико-статистических методов в исторических исследованиях. В кн.: *Источниковедение: Теоретические и методологические проблемы* / Отв. ред. С. О. Шмидт. Москва: Наука, 1969. С. 115–133.

99. Ковальченко И. Д. Об опыте математико-статистической обработки выборочных данных о крестьянском хозяйстве России XIX в. *Вестник Московского университета*. Серия 8. История. 1966. № 1. С. 76–89.

100. Ковальченко И. Д. Применение количественных методов и ЭВМ в исторических исследованиях. *Вопросы истории*. 1984. № 9. С. 61–73.

101. Ковальченко И. Д., Милов Л. В. *Всероссийский аграрный рынок. XVIII – начало XX в.: Опыт количественного анализа*. Москва: Наука, 1974. 413 с.

102. Ковальченко И. Д., Селунская Н. Б., Литваков Б. М. *Социально-экономический строй помещичьего хозяйства Европейской России в эпоху капитализма: Источники и методы изучения*. Москва: Наука, 1982. 264 с.

103. Козлов В. П. *Основы теоретической и прикладной археографии*. 2-е изд., переработанное и дополненное. Москва: Российская политическая энциклопедия (РОССПЭН), 2008. 224 с.

104. Козлов В. П. *Российская археография конца XVIII – первой четверти XIX века*. Москва: РГГУ, 1999. 416 с.

105. Копосов Н. Е. *Как думают историки*. Москва: Новое литературное обозрение, 2001. 326 с. (Научное приложение, вып. XXXI).

106. Корнева И. И. Основные принципы передачи текстов документов при публикации и некоторые вопросы воспроизведения

источников ранних периодов (XIV–XVIII вв.). В кн.: *Труды МГИАИ*. Том. VIII. Москва: МГИАИ, 1957. С. 43–65.

107. Королев Г. И. *Медиевистическая археография за рубежом: Труды XVI–XVIII веков*. Москва: РГГУ, 2001. 198 с.

108. Королев Г. И. *Медиевистическая археография за рубежом: Труды XIX – начала XX в.* Москва: РГГУ, 2003. 218 с.

109. Кортава Т. В. *Московский приказный язык XVII века как особый тип письменного языка*. Москва: Издательство Московского университета, 1998. 109 с.

110. Котков С. И. О совместном издании древнерусских скорописных памятников лингвистами и историками. В кн.: *Лингвистическое источниковедение*. Москва: Издательство АН СССР, 1963. С. 5–23.

111. Кравцов И. В., Варфоломеев А. Г. Принципы организации информационного пространства сетевого сообщества исследователей рукописных текстов. В кн.: *Информационное общество. Интеллектуальная обработка информации. Информационные технологии. Материалы 7-ой международной конференции НТИ-2007 (Москва, 24–26 октября 2007 г.)*. Москва, 2007. С. 383–386.

112. Кравцов И. В., Филатов В. О. Информационная система для работы с коллекциями рукописных исторических документов. *Информационные технологии моделирования и управления*. 2007. № 2 (36). С. 188–196.

113. Крейн Д., Паскарелло Э., Джеймс Д. *Аях в действии*. Москва: Вильямс, 2006. 640 с.

114. Кроче Б. *Теория и история историографии* / Пер. с итальянского И. М. Заславской. Москва: Языки русской культуры, 1998. 191 с. (Серия «Studia Historica»).

115. Кузнецов А. М. Почему Даугавпилс назывался Борисоглебовым. *Русская речь*. 1994. № 1. С. 83–86.

116. Куперштох В. Л., Трофимов В. А. Алгоритм анализа структуры матрицы связи. *Автоматика и телемеханика*. 1975. № 11. С. 170–180.

117. Ланглуа Ш.-В., Сеньобос Ш. *Введение в изучение истории* / Пер. с французского А. Серебряковой. Санкт-Петербург: Издание О. Н. Поповой, 1899. 275, ii с. (Серия «Образовательная библиотека»).

118. Лаппо-Данилевский А. С. *Очерк русской дипломатики частных актов. Лекции, читанные слушателям «Архивных Курсов» при Петроградском Археологическом Институте в 1918 году.* Петроград, 1920. 189 с.

119. Лаппо-Данилевский А. С. *Очерк русской дипломатики частных актов.* Санкт-Петербург: Северная звезда, 2007. 284 с.

120. Лихачев Д. С. (при участии А. А. Алексеева и А. Г. Боброва). *Текстология: На материале русской литературы X–XVII веков.* Изд. 3-е., переработанное и дополненное. Санкт-Петербург: Алетейя, 2001. 759 с. (Серия «Славянская библиотека. Bibliotheca slavica»).

121. Медушевская О. М. *Современное зарубежное источниковедение.* Москва: Высшая школа, 1983. 143 с.

122. Митяев К. Г. *История и организация делопроизводства в СССР: Учебное пособие* / Под ред. А. В. Чернова. Москва: МГИАИ, 1959. 359 с.

123. Москин Н. Д. Методы анализа теоретико-графовых моделей текстов. *Вестник Поморского университета. Серия «Естественные и точные науки».* № 3. 2006. С. 105–108.

124. Москин Н. Д. *Теоретико-графовые модели структуры фольклорных текстов, алгоритмы поиска закономерностей и их программная реализация.* Автореф. дисс. на соиск. уч. степени к.т.н. Петрозаводск, 2006.

125. Напиерский К. Э., сост. *Грамоты, касающиеся до сношений северо-западной России с Ригою и ганзейскими городами в XII, XIII и XIV веке. Найдены в Рижском архиве К.Э. Напиерским и изданы Археографическою комиссиею.* Санкт-Петербург: Императорская Академия Наук, 1857. [24] с., [11] л. илл.

126. Новиков А. И. *Семантика текста и ее формализация.* Москва: Наука, 1983. 215 с.

127. Орешников А. С. О применении формализации к описанию деловых текстов. *Источниковедение и история русского языка.* Москва, 1964. С. 150–156.

128. *От Нестора до Фонвизина: Новые методы определения авторства* / Под ред. Л. В. Милова. Москва: Прогресс, 1994. 445 с.

129. *Отбор исторических источников для публикации в различных типах и видах изданий документов. (Методические рекомендации)* / ГАУ при СМ СССР; ВНИИДАД. Москва, 1986. 105 с.

130. Паринов С. И. О тенденциях и ориентирах развития открытых электронных библиотек (из опыта создания и эксплуатации системы Соционет). *Электронные библиотеки*. 2004. Т. 7. № 6. С.1–14.

131. Пигалева С. В., Корниенко С. И., Гагарина Д. А., Горбачева Н. Г. Газета «Пермские губернские ведомости» (1838–1844 гг.): традиционный источниковедческий анализ и информационные технологии. *Вестник Пермского университета. История*. 2008. Вып. 7 (23). С. 72–76.

132. *Правила издания исторических документов в СССР*. 2-е, переработанное и дополненное изд. Москва: ГАУ при СМ СССР, 1990. 187 с.

133. *Правила лингвистического издания памятников древнерусской письменности* / Институт русского языка АН СССР. Москва: Издательство АН СССР, 1961. 64 с.

134. Пронштейн А. П. Дипломатика как научная дисциплина. *История СССР*. 1971. № 2. С. 178–180.

135. Пронштейн А. П. *Методика исторического исследования*. Ростов-на-Дону: Издательство Ростовского университета, 1971. 467 с.

136. Пронштейн А. П., Данилевский И. Н. *Вопросы теории и методики исторического исследования*. Москва: Высшая школа, 1986. 208 с.

137. Ратмэн Д. Б. «Новая социальная история» в США. *Новая и новейшая история*. 1990. № 2. С. 66–80.

138. Риккерт Г. *Границы естественнонаучного образования понятий. Логическое введение в исторические науки*. Санкт-Петербург: Наука, 1997. 532 с. (Серия «Слово о сущем»).

139. Рогачевский А. Л. *Кульмская грамота – памятник права Пруссии XIII в.* Санкт-Петербург: Издательство Санкт-Петербургского университета, 2002. 368 с.

140. Скороходько Э. Ф. *Семантические сети и автоматическая обработка текста*. Киев: Наукова думка, 1983. 218 с.

141. Соколов А. К., Тяжелникова В. С. Массовые источники и компьютеризация исторических исследований. В кн.: *Источниковедение новейшей истории России: Теория, методология, практика*. Москва: Высшая школа, 2004. С. 446–496.

142. Старостин Е. В. *Зарубежное архивоведение: Проблемы истории, теории и методологии* / РГГУ, ИАИ. Москва: Русский мир, 1997. 332 с.

143. Тартаковский А. Г. Социальные функции источников как методологическая проблема источниковедения. *История СССР*. 1983. № 3. С. 112–130.

144. Тихомиров М. Н. Приказное делопроизводство в XVII веке: [Курс лекций]. В кн.: Тихомиров М. Н. *Российское государство XV–XVII веков*. Москва: Наука, 1973. С. 348–383.

145. Тяжельникова В. С. Компьютерное источниковедение: К постановке проблемы. В кн.: *Круг идей: Развитие исторической информатики: Труды II конференции АИК* / Отв. ред. Л. И. Бородин, В. С. Тяжельникова. Москва: Издательство Московского городского объединения архивов, 1995. С. 270–284.

146. Уайт Х. *Метаистория: Историческое воображение в Европе XIX века* / Пер. с англ. Екатеринбург: Издательство Уральского университета, 2002. 527 с. (Серия «Studia humanitatis»).

147. Филатов В. О., Кравцов И. В. Технологии создания информационной системы для работы с полнотекстовыми базами данных исторических документов. В кн.: *Современные информационные технологии и письменное наследие: от древних рукописей к электронным текстам: Материалы международной научной конференции (Ижевск, 13–17 июля 2006 г.)* / Отв. ред. В. А. Баранов. Ижевск: Издательство Ижевского государственного технического университета, 2006. С. 168–173.

148. Филатов В. О., Кравцов И. В., Варфоломеев А. Г. Информационная система для работы с полнотекстовыми базами данных исторических документов на основе технологии XML. В кн.: *Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды Восьмой Всероссийской научной конференции (RCDL'2006). Суздаль, 17–19 октября 2006 г.* Ярославль: Ярославский государственный университет им. П. Г. Демидова, 2006. С. 337–344.

149. Холзнер С. *XSLT*. Санкт-Петербург: Питер, 2002. 544 с. (Серия «Библиотека программиста»).

150. Хорошкевич А. Л., сост. *Полоцкие грамоты XIII– начала XVII вв.* I вып. / Институт истории СССР АН СССР; отв. ред. А. А. Зимин. Москва, 1977. 227 с.

151. Хорошкевич А. Л., сост. *Полоцкие грамоты XIII–начала XVI вв.* II вып. / Институт истории СССР АН СССР; отв. ред. А. А. Зимин. Москва, 1978. 219 с.

152. *Центральный государственный архив Латвийской ССР: Краткий справочник.* Часть I: (1220–1918) / Отв. составитель Н. Н. Рыжов. Рига: Авотс, 1980. 172 с.

153. Черепнин Л. В. *Русские феодальные архивы XIV–XV веков* / АН СССР, Институт истории. Часть первая. Москва; Ленинград: Издательство АН СССР, 1948. 470 с.

154. Черепнин Л. В. *Русские феодальные архивы XIV–XV веков* / АН СССР, Институт истории. Часть вторая. Москва: Издательство АН СССР, 1951. 428 с.

155. Чернов А. В. *Вооруженные силы Русского государства в XV–XVII в.в.: С образования централизованного государства до реформы при Петре I.* Москва: Воениздат, 1954. 223 с.

156. Чирков С. В. *Археография в творчестве русских ученых конца XIX – начала XX века* / Отв. ред. С. О. Шмидт. Москва: Знак, 2005. 320 с. (Серия «Studia historica»).

157. Шафер С. *HTML, XHTML и CSS. Библия пользователя.* 5-е изд. Москва: Диалектика, 2010. 656 с.

158. Шевель А. Технология GRID. *Открытые системы.* СУБД. 2001. № 2. С. 36–39. <http://www.osp.ru/os/2001/02/179932/> (04.12.2012).

159. Шепелев Л. Е. Проблемы источниковедческого изучения делопроизводственных документов государственных учреждений XIX – начала XX вв. В кн.: *Источниковедение отечественной истории: Сб. статей.* 1975. Москва: Наука, 1976. С. 248–271.

160. Шмидт С. О. Современные проблемы источниковедения. В кн.: Шмидт С. О. *Путь историка: Избранные труды по источниковедению и историографии.* Москва: РГГУ, 1997. С. 21–63.

161. Юмашева Ю. Ю. Утопия или утро завтрашнего дня? (Новые средства работы с электронными источниками). В кн.: *Новые информационные ресурсы и технологии в исторических исследованиях и образовании: Сборник тезисов докладов и сообщений Всероссийской конференции. Подмосковье, «Боровое», 6–9 апреля 2000 г.* Москва, 2000. С. 259–262.

162. Янин В. Л. *Очерки комплексного источниковедения: Средневековый Новгород.* Москва: Высшая школа, 1977. 240 с.

163. Abukhanfusa, K., ed. *The Principle of Provenance: First Stockholm Conference on Archival Theory and the Principle of Provenance, 2–3 September, 1993*. Stockholm: Swedish National Archives, 1994. 207 p.

164. Ahonen, E., Hyvönen, E. Publishing Historical Texts on the Semantic Web – A Case Study. In: *Proceedings of the Third IEEE International Conference on Semantic Computing (ICSC2009)*. Berkeley (CA), 2009. Pp. 167–173.

165. Anderson, S. E-science for the Arts and Humanities. A Discussion Paper. *AHRB E-Research Expert Seminar, 28th April 2004*. <http://www.ahds.ac.uk/e-science/e-science-discussion-2004.pdf> (04.12.2012).

166. Ardichvilli, A., Page, V., Wentling, T. Motivation and Barriers to Participation in Virtual Knowledge-sharing Communities of Practice. In: *Journal of Knowledge Management*. 2003. Vol. 7, no. 1. Pp. 64–77. doi:10.1108/13673270310463626.

167. Aschenbrenner, A., Blanke, T., Dunn, S., Kerzel, M., Rapp, A., Zielinski, A. Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur. *Bibliothek: Forschung und Praxis*. 2007. Bd. 31, N. 1. S. 11–21. http://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2007/Jg_31-Nr_1/Jg_31-Nr_1_Aufsaeetze/Jg_31-2007-Nr_1-S_11-21.pdf (05.12.2012).

168. Aschenbrenner, A., Gietz, P., Küster, M. W., Ludwig, Ch., Neuroth, H. TextGrid – a Modular Platform for Collaborative Textual Editing. In: *Proceedings of the International Workshop on Digital Library Goes e-Science (DLSci06)*. Alicante, 2006. Pp. 27–36. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/ecdl06_textgrid.pdf (06.08.2012).

169. Auer, S., Dietzold, S., Riechert, T. OntoWiki – A Tool for Social, Semantic Collaboration. In: Cruz, I. et al., eds. *The Semantic Web – ISWC 2006: 5th International Semantic Web Conference. ISWC 2006. Athens, GA, USA, November 5–9, 2006. Proceedings*. Berlin; Heidelberg: Springer-Verlag, 2006. Pp. 736–749. (Lecture Notes in Computer Science, 4273).

170. Aydelotte, W. O. Quantification in History. In: Rowney, D. K., Graham, J. Q., eds. *Quantitative History: Selected Readings in the Quan-*

titative Analysis of Historical Data. Homewood (Ill.): The Dorsey Press, 1969. Pp. 3–22.

171. Baruzzo, A., Casoto, P., Challapalli, P., Dattolo, A., Pudota, N., Tasso, C. Toward Semantic Digital Libraries: Exploiting Web2.0 and Semantic Services in Cultural Heritage. *Journal of Digital Information*. 2009. Vol. 10, no. 6. <http://journals.tdl.org/jodi/article/viewArticle/688/576> (07.08.2012).

172. Belnap, N. A Useful Four-Valued Logic. In: Dunn, J. M., Epstein, G., eds. *Modern Uses of Multiple-Valued Logic: Invited Papers from the Fifth International Symposium on Multiple-Valued Logic held at Indiana University, Bloomington, Indiana, May 13-16, 1975*. Dordrecht; Boston: Reidel, 1977. Pp. 8–37. (Series “Episteme”, vol. 2).

173. Benninghoven, F., hrsg. *Liv-, est- un kurländische Urkundenregister bis zum Jahre 1300*. Hamburg, 1959. 216 S.

174. Berners-Lee, T., Hendler, J., Lassila, O. The Semantic Web. *Scientific American*. 2001. Vol. 284, no. 5. Pp. 34–43.

175. Bes, L., Frankot, E., Brand, H., eds. *Baltic Connections: Archival Guide to the Maritime Relations of the Countries around the Baltic Sea (Including the Netherlands), 1450–1800*. Vol. II: *Latvia, Lithuania, the Netherlands*. Leiden; Boston: Brill, 2007. xxv, 787–1603 pp. (The Northern World, vol. 36/II).

176. Blanke, T., Hedges, M., Palmer, R. Restful Services for the e-Humanities – Web Services that Work for the e-Humanities Ecosystem. In: *DEST’09. 3rd IEEE International Conference on Digital Ecosystems and Technologies*. [S.l.]: IEEE, 2009. Pp. 637–642.

177. Bolvig, A., Lindley, Ph., eds. *History and Images: Towards a New Iconology*. Turnhout: Brepols, 2003. xxx, 430 p. («Medieval Texts and Cultures of Northern Europe», 5).

178. Boonstra, O., Breure, L., Doorn, P. *Past, Present and Future of Historical Information Science*. Amsterdam: NIWI – KNAW, 2004. 130 p.

179. Bulterman, D. C. A., Rutledge, L. W. *SMIL 3.0. Flexible Multimedia for the Web, Mobile Devices and Daisy Talking Books*. 2nd ed. Berlin: Springer, 2009. xxviii, 508 p. (X.media.publishing).

180. Bunke, H. Graph Matching: Theoretical Foundations, Algorithms, and Applications. *International Conference on Vision Interface, Montreal, Quebec, Canada, 2000*. Pp. 82–88. <http://www.ai.rug.nl/ki2/literature/graphmatch-bunke.pdf> (05.12.2012).

181. Burkard, B., Gruner, S., Vogeler, G. Informatics for Historians. Tools for Medieval Document XML Markup and Their Impact on History-Science. *Journal for Universal Computer Science*. 2008. Vol. 14, no. 2. Pp. 193–210.

182. Burke, P. *Sociology and History*. London; Boston: G. Allen & Unwin, 1980. 116 p.

183. Chevalier, P., Granjon, L., Leclercq, É., Millereux, A., Savonnet, M., Sapin, C. Base de Données annotées et Wiki pour la constitution du corpus numérique CARE. *Hortus Artium Medievalium*. 2012. Vol. 18, no. 1. Pp. 27–35.

184. Ciravegna, F., Greengrass, M., Hitchcock, T., Chapman, S., McLaughlin, J., Bhagdev, R. Finding Needles in Haystacks: Data-mining in Distributed Historical Datasets. In: Greengrass, M., Hughes, L., eds. *The Virtual Representation of the Past*. [S.l.]: Ashgate, 2008, reprinted in 2010. Pp. 65–79 (Digital Research in the Arts and Humanities).

185. Ciula, A., Spence, P., Vieira, J. M., Poupeau, G. Expressing Complex Associations in Medieval Historical Documents: The Henry III Fine Rolls Project. In: *Digital Humanities 2007: The 19th Joint International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing*. University of Illinois, Urbana-Champaign June 4 – June 8, 2007. *Conference Abstracts*. Urbana-Champaign: University of Illinois, 2007. Pp. 34–37.

186. Clark, P., Harrison, P., Jenkins, T., Thompson, J., Wojcik, R. Acquiring and Using World Knowledge Using a Restricted Subset of English. In: *The 18th International FLAIRS Conference (FLAIRS'05)*. 2005. <http://www.cs.utexas.edu/users/pclark/papers/flairs.pdf> (07.08.2012).

187. de Coi, J. L., Fuchs, N. E., Kaljurand, K., Kuhn, T. Controlled English for Reasoning on the Semantic Web. In: Bry, F., Małuszyński, J., eds. *Semantic Techniques for the Web: The Reverse Perspective*. Berlin; Heidelberg: Springer-Verlag, 2009. Pp. 276–308. (Lecture Notes in Computer Science, 5500).

188. Coombs, J., Renear, A., DeRose, S. Markup Systems and the Future of Scholarly Text Processing. *Communications of the ACM*. November 1987. Vol. 30, issue 11. Pp. 933–947.

189. Cossard, P. K. Electronic Medievalia: Global Warming for Humanities Computing? Strategic Changes in the Economic Forecast. *The Heroic Age. A Journal of Early Medieval Northwestern Europe*. Oc-

tober 2006, Issue 9. <http://www.mun.ca/mst/heroicage/issues/9/em.html> (05.12.2012).

190. Cox, A. M. What are Communities of Practice? A Comparative Review of Four Seminal Works. In: *Journal of Information Science*. 2005. Vol. 31, no. 6. Pp. 527–540. doi:10.1177/0165551505057016.

191. Doerr, M. The CIDOC CRM. An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*. 2003. Vol. 24, no. 3. Pp. 75–92.

192. Dong, Y., Li, M. HyO-XTM: A Set of Hyper-Graph Operations on XML Topic Map toward Knowledge Management. *Future Generation Computer Systems*. 2004. Vol. 20, no. 1 (January). Pp. 81–100.

193. Duchein, M. Theoretical Principles and Practical Problems of Respect des Fonds in Archival Science. *Archivaria*. Summer 1983. Vol. 16. Pp. 64–82.

194. Eisenberg, J. D. *SVG Essentials*. [S.l.]: O'Reilly Media, 2002. 364 p.

195. Fellbaum, C., ed. *WordNet: An Electronic Lexical Database*. Cambridge (MA): MIT Press, 1998. 422 p.

196. Fuchs, N. E., ed. *Controlled Natural Language: Workshop on Controlled Natural Language, CNL 2009. Marettimo Island, Italy, June 2009. Revised Papers*. Berlin; Heidelberg: Springer-Verlag, 2010. x, 291 p. (Lecture Notes in Computer Science, 5972).

197. Gietz, P. et al. TextGrid and e-Humanities. In: *Second IEEE International Conference on e-Science and Grid Computing (e-Science'06). December 4–6, 2006, Amsterdam, Netherlands*. 2006. <http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/TextGrid-Amsterdam-2006-final.pdf> (05.12.2012).

198. Goetz, H.-W. *Proseminar Geschichte: Mittelalter*. 3., überarbeitete Auflage. Stuttgart: Verlag Eugen Ulmer, 2006. 399 S. (UTB 1719).

199. Grigoriev, N. XSL Formatting Objects Tutorial. <http://www.renderx.com/tutorial.html> (19.11.2012); <http://www.renderx.com/files/XSLFormattingObjectsTutorial.pdf> (09.12.2012).

200. Hähnle, R. *Automated Deduction in Multiple-Valued Logics*. Oxford: Clarendon Press; New York: Oxford University Press, 1993. ix, 172 p. (International Series of Monographs on Computer Science, 10).

201. Hansen, A. M. Text Encoding of Manuscripts: Danish Prayer Books from the 16th Century. *Le Médiéviste et l'ordinateur. Histoire médiévale, informatique et nouvelles technologies*. 2002, no. 41. http://lemo.irht.cnrs.fr/41/mo41_09.htm (05.12.2012).

202. Helbig, H. *Knowledge Representation and the Semantics of Natural Language*. Berlin; Heidelberg: Springer-Verlag, 2006. xviii, 647 p. (Cognitive Technologies).

203. Herwijnen, E. van. *Practical SGML*. 2nd ed. Boston: Kluwer Academic Publishers, 1994. xx, 288 p.

204. Himmelfarb, G. *The New History and the Old*. Cambridge (MA): Harvard University Press (Belknap Press), 1987. 209 p.

205. Hockey, S. The History of Humanities Computing. In: Schreibman, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities*. Malden (MA); Oxford; Carlton: Blackwell Publishing, 2004. Pp. 3–19.

206. Hollaus, F., Gau, M., Miklas, H., Sablatnig, R. Modern Technologies for Manuscript Research. In: *Информационные технологии и письменное наследие. El'Manuscript–2012: Материалы IV международной научной конференции. Петрозаводск, 3–8 сентября 2012 года* / отв. ред. В. А. Баранов, А. Г. Варфоломеев. Петрозаводск; Ижевск: Петрозаводский государственный университет; Ижевский государственный технический университет им. М. Т. Калашникова, 2012. С. 57–64.

207. Hughes, L. Conclusion: Virtual Representations of the Past – New Research Methods, Tools and Communities of Practice. In: Greengrass, M., Hughes, L., eds. *The Virtual Representation of the Past*. [S.l.]: Ashgate, 2008, reprinted in 2010. Pp. 191–201 (Digital Research in the Arts and Humanities).

208. Hyvönen, E., Mäkela, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., Junnila, M., Kettula, S. MuseumFinland – Finnish Museums on the Semantic Web. *Journal of Web Semantics*. 2005. Vol. 3, no. 2. Pp. 224–241.

209. Ide, N., Woolner, D. Historical Ontologies. In: Ahmad, K., Brewster, Chr., Stevenson, M. eds. *Words and Intelligence*. Vol. II: *Essays in Honor of Yorick Wilks*. Dordrecht: Springer, 2007. Pp. 137–152.

210. Ivanovs, A. Complex Approach to the Research of the 1949 Deportation (the Case of Latvia): Collection of Oral History and Documen-

tary Records 'The Voices of the Deported, 25 March 1949'. *Lituanistica*. 2011. T. 57, Nr. 1 (83). Pp. 106–111.

211. Ivanovs, A. Dokumenti par Rīgas, Gotlandes un Hanzas attiecībām ar Lielo Novgorodu Latvijas Valsts vēstures arhīvā (12.–16. gs.). *Latvijas Arhīvi*. 2010, Nr. 2. 7.–34. lpp.

212. Ivanovs, A. Dokumentu kompleksa apzināšana, rekonstrukcija un izpēte: arheogrāfiskie un avotpētnieciskie aspekti. *Humanitāro Zinātņu Vēstnesis* / Daugavpils Universitāte. 2006, Nr. 10. 76.–88. lpp.

213. Ivanovs, A. Kompleksa «Moscowitica–Ruthenica» ieviešana zinātnes aprītē: arheogrāfisks apskats.” *Latvijas Arhīvi*. 2004. Nr. 2. 47.–85. lpp.

214. Ivanovs, A. Reģionālisma problēma Latgales vēstures kompleksajā izpētē. (Problēmas nostādne). Grām.: *IX Zinātnisko lasījumu materiāli. Vēstures sekcija*. 3. krājums / Daugavpils Pedagoģiskā universitāte; atb. red. V. Šalda. Daugavpils: Saule, 2000. 9.–15. lpp.

215. Ivanovs, A. Vēstures avotu kompleksa rekonstrukcijas problēma: kolekcija Moscowitica–Ruthenica. Grām.: *Vēsture: Avoti un cilvēki. Humanitārās fakultātes XIV starptautisko zinātnisko lasījumu materiāli. Vēsture VIII* / Daugavpils Universitāte; atb. red. V. Šalda. Daugavpils: Saule, 2004. 49.–57. lpp.

216. Ivanovs, A. Zespól dokumentów «Moscowitica–Ruthenica» w Łotewskim Państwowym Archiwum Historycznym. *Zapiski Historyczne (Poświęcone historii Pomorza i krajów Bałtyckich)*. Toruń, 2006. Tom LXXI. Zeszyt 2-3. S. 125–143.

217. Ivanovs, A., Varfolomeyev, A. Editing and Exploratory Analysis of Medieval Documents by Means of XML Technologies: The Case of the Documentary Source Complex Moscowitica–Ruthenica. In: *Humanities, Computers and Cultural Heritage: Proceeding of the Association for History and Computing*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005. Pp. 155–160.

218. Ivanovs, A., Varfolomeyev, A. Semantic Publications of Charter Corpora (The Case of a Diplomatic Edition of the Complex of Old Russian Charters «Moscowitica–Ruthenica») = Publications sémantiques de corpus des chartes (le cas d'une édition diplomatique du complexe des anciennes chartes russes «Moscowitica–Ruthenica»). In: *International Conference «Digital Diplomats: Tools for the Digital Diplomatist»*. Napoli, 29 settembre – 1 ottobre 2011. [Naples], 2011. Pp. 36–40.

219. Kamp, H., Reyle, U. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic Publishers, 1993. viii, 716 p. (Studies in Linguistics and Philosophy, Vol. 42).
220. Klir, J. R., Yuan, B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River (NJ): Prentice Hall PTR, 1995. xv, 574 p.
221. Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., Studer, R. Semantic Wikipedia. *Journal of Web Semantics*. 2007. Vol. 5, no. 4. Pp. 251–261.
222. Kuhn, T. How Controlled English can Improve Semantic Wikis. In: *Proceedings of the Fourth Workshop on Semantic Wikis: European Semantic Web Conference 2009*. [S.I.], 2009. Pp. 1–16. (SEUR Workshop Proceedings, vol. 464).
223. Küster, M. W., Ludwig, Ch., Aschenbrenner, A. TextGrid as a Digital Ecosystem. In: *IEEE DEST. 2007, 21–23 February, Cairns*. http://www.textgrid.de/fileadmin/TextGrid/veroeffentlichungen/DigitalEcosystem07_CameraReady-1.pdf (07.08.2012); doi: 10.1109/DEST.2007.372029.
224. *The Menota Handbook: Guidelines for the Electronic Encoding of Medieval Nordic Primary Sources. Version 1.0*. Bergen: Medieval Nordic Text Archive, 2003. www.hit.uib.no/menota/guidelines (05.12.2012).
225. Mirzaee, V., Iverson, L., Hamidzadeh, B. Computational Representation of Semantics in Historical Documents. In: *Humanities, Computers and Cultural Heritage: Proceeding of the Association for History and Computing*. Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005. Pp. 199–206.
226. Musser, J., O'Reilly, T. *Web 2.0 Principles and Best Practices*. [S.I.] O'Reilly Media, Inc., 2006. 101 p.
227. O'Hara, R., Robinson, P. Computer-Assisted Methods of Stemmatic Analysis. In: Blake, N., Robinson, P., eds. *The Canterbury Tales Project Occasional Papers*. Oxford: Office for Humanities and Communication, 1993. Vol. 1. Pp. 53–74. <http://rjohara.net/cv/1993-ctp/> (05.12.2012).
228. O'Reilly, T. What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. Sept. 2005. <http://oreilly.com/web2/archive/what-is-web-20.html> (09.12.2012)

229. Ortí, M. M. C., ed. *Vocabulaire International de la Diplomatie*. 2nd ed. València: Universitat de València, 1997. 308 p.

230. Panton, K., Matuszek, C., Lenat, D., Schneider, D., Witbrock, M., Siegel, N., Shepard, B. Common Sense Reasoning – From Cyc to Intelligent Assistant. In: Cai, Y., Abascal, J., eds. *Ambient Intelligence in Everyday Life*. Berlin; Heidelberg: Springer-Verlag, 2006. Pp. 1–31. (Lecture Notes in Computer Science, 3864; Lecture Notes in Artificial Intelligence).

231. Pasin, M., Motta, E., Keynes M. PhiloSURFical: Browse Wittgensteins Tractatus with the Semantic Web. In: A. Pichler, H. Hrachovec, eds. *Wittgenstein and the Philosophy of Information. Proceedings of the 30th International Ludwig Wittgenstein Symposium*. Vol. 1. [S.l.]: Ontos Verlag, 2008. Pp. 319–333.

232. Pease, A., Murray, W. An English to Logic Translator for Ontology-Based Knowledge Representation Languages. In: *Proceedings of the 2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering*. Beijing, 2003. Pp. 777–783.

233. Pechter, R. Conformance Standard for the Predictive Model Markup Language. In: *DMSSP'06. Proceedings of the 4th International Workshop on Data Mining Standards, Services and Platforms*. 2006. Pp. 6–13. doi: 10.1145/1289612.1289613.

234. Poupeau, G. Réflexions sur l'utilisation de la TEI pour coder les sources diplomatiques à partir de l'exemple du Cartulaire blanc de l'abbaye de Saint-Denis. *Le Médiéviste et l'ordinateur: Histoire médiévale, informatique et nouvelles technologies*. 2004, no. 43. <http://lemo.irht.cnrs.fr/43/43-12.htm> (05.12.2012).

235. Rabb T. K., Rotberg, R. I., eds. *The New History, the 1980-s and Beyond: Studies in Interdisciplinary History*. Princeton (NJ): Princeton University Press, 1982. 332 p.

236. Robertson, B. G. Visualizing an Historical Semantic Web with HEML. In: *Proceedings of the 15th International Conference on World Wide Web (Edinburgh, Scotland, May 23–26, 2006)*. WWW'06. New York (NY): ACM Press, 2006. Pp. 1051–1052. doi:10.1145/1135777.1136010.

237. Robinson, P. Current Issues in Making Digital Editions of Medieval Texts – or, Do Electronic Scholarly Editions Have a Future? *Digital Medievalist*, 1 (Spring 2005). <http://www.digitalmedievalist.org/journal/1.1/robinson/> (05.12.2012).

238. Roos, T., Heikkilä, T. E. Evaluating Methods for Computer-assisted Stemmatology Using Artificial Benchmark Data Sets. *Literary and Linguistic Computing*. 2009. Vol. 24, no. 4. Pp. 417–433.

239. Rosner, M., Fuchs, N. E., eds. *CNL 2010: Second Workshop on Controlled Natural Languages. Pre-Proceedings of the Second Workshop on Controlled Natural Languages. Marettimo Island, Sicily, Italy, September 13–15, 2010*. <http://ceur-ws.org/Vol-622/> (07.08.2012).

240. Schmidt, D. The Role of Markup in the Digital Humanities. *Historical Social Research = Historische Sozialforschung*. 2012. Vol. 37, no. 3 (A Special Issue: Controversies around the Digital Humanities). Pp. 125–146.

241. Schreibman, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities*. Malden (MA); Oxford; Carlton: Blackwell Publishing, 2004. xxvii, 611 p.

242. Schwitter, R. Processing Coordinated Structures in PENG Light. In: Wang, D., Reynolds, M., eds. *AI 2011: Advances in Artificial Intelligence. 24th Australasian Joint Conference. Perth, Australia, December 5–8, 2011. Proceedings*. Berlin; Heidelberg: Springer-Verlag, 2011. Pp. 658–667. (Lecture Notes in Computer Science, 7106).

243. Shadbolt, N., Berners-Lee, T. Hall, W. The Semantic Web Revisited. *IEEE Intelligent Systems*. 2006. Vol. 21, Issue 3. Pp. 96–101.

244. Shapiro, S. C. and the SNePS Implementation Group. *SNePS 2.7 User's Manual*. February 11, 2008. viii, 115 p. <http://www.cse.buffalo.edu/sneps/Manuals/manual27.pdf> (05.12.2012).

245. Shirrell, J. *XML: A Deeper Understanding*. [Online Book]. Chapter 2: SGML. <http://www.xmlbook.info/> (17.11.2012).

246. Shotton, D. Semantic Publishing. The Coming Revolution in Scientific Journal Publishing. *Learned Publishing*. 2009. Vol. 22, no. 2. Pp. 85–94.

247. Shotton, D., Portwin, K., Klyne, G., Miles, A. Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS Computational Biology*. 2009. Vol. 5, no. 4. e1000361. doi:10.1371/journal.pcbi.1000361. <http://dx.doi.org/10.1371/journal.pcbi.1000361> (05.12.2012).

248. Smith, M. K. Communities of Practice. In: *The Encyclopedia of Informal Education*. 2003, 2009. http://www.infed.org/biblio/communities_of_practice.htm (05.12.2012).

249. Souzis, A. Building a Semantic Wiki. *IEEE Intelligent Systems*. September–October 2005. Vol. 20, no. 5. Pp. 87–91.
250. Stone, L. *The Past and the Present*. Boston: Routledge & K. Paul, 1981. xii, 274 p.
251. Stone, L. The Revival of Narrative: Reflections on a New Old History. *Past and Present*. 1979. Vol. 85, no. 1. Pp. 3–24.
252. Thaller, M. Controversies around the Digital Humanities: An Agenda. *Historical Social Research = Historische Sozialforschung*. 2012. Vol. 37, no. 3 (A Special Issue: Controversies around the Digital Humanities). Pp. 7–23.
253. Thomas, W. G. Computing and the Historical Imagination. In: Schreibman, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities*. Malden (MA); Oxford; Carlton: Blackwell Publishing, 2004. Pp. 56–68.
254. Vanhoutte, E. An Introduction to the TEI and the TEI Consortium. *Literary and Linguistic Computing*. 2004. Vol. 19, Issue 1. Pp. 9–16.
255. Varfolomeyev, A., Ivanovs, A. Knowledge-Based Scholarly Environment Project for Regional Historical Studies. In: *Interactive Systems and Technologies: The Problem of Human-Computer Interaction*. Vol. III: *Collection of Scientific Papers*. Ulyanovsk: UISTU, 2009. Pp. 273–276.
256. Varfolomeyev, A., Ivanovs, A. Wiki Technologies for Semantic Publication of Old Russian Charters. In: *Digital Humanities 2012: Conference Abstracts. University of Hamburg, Germany, July 16–22, 2012* / Ed. by J. Ch. Meister. Hamburg: Hamburg University Press, 2012. Pp. 405–407.
257. Varfolomeyev, A., Soms, H., Ivanovs, A. Knowledge-Based Information Systems in Research of Regional History. In: *Digital Humanities 2008: The 20th Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities and The 1st Joint International Conference of the Association for Literary and Linguistic Computing, the Association for Computers and the Humanities, and the Society for Digital Humanities – Société pour l'étude des médias interactifs. University of Oulu, Finland, 24–29 June, 2008. Conference Abstracts*. [Oulu]: University of Oulu, 2008. Pp. 210–211.

258. Vocabulaire Internationale de Diplomatie (Terminologie der Diplomatie). <http://www.cei.lmu.de/VID/> (05.12.2012).

259. de Waard, A. From Proteins to Fairytales: Directions in Semantic Publishing. *IEEE Intelligent Systems*. 2010. Vol. 25, no. 2. Pp. 83–88.

260. Witte, R., Gitzinger, T., Kappler, T., Krestler, R. A Semantic Wiki Approach to Cultural Heritage Data Management. In: *Language Technology for Cultural Heritage Data (LaTeCH 2008), Workshop at LREC 2008, June 1st, 2008, Marrakech, Morocco*. 2008. <http://www.l3s.de/web/upload/documents/1/latech08.pdf> (05.12.2012).

261. Wittern, C., Ciula, A., Tuohy, C. The Making of TEI P5. *Literary and Linguistic Computing*. 2009. Vol. 24, no. 3. Pp. 281–296. doi: 10.1093/lc/fqp017.

Список иллюстраций

Глава 2

2.1. [1219–1228/9 гг.] Условия к договору Смоленска с Ригой и Готландом, или так называемый «договор неизвестного смоленского князя», в историографии известный также как список К. (Акт № 1). *Латвийский национальный архив – Латвийский государственный исторический архив (ЛНА-ЛГИА). Ф. 8. Оп. 3. Capsula A. № 16. Л. 1. Подлинник. Пергамен – с. 34.*

2.2. 1284 г. [мая 18]. Грамота о предоставлении чистого пути немецким и русским купцам, написанная от имени смоленского князя Федора Ростиславича архиепископу рижскому Иоганну I, магистру Ливонского ордена Виллекину и ратманам города Риги. (Акт № 4). *ЛНА-ЛГИА. Ф. 8. Оп. 3. Capsula A. № 72. Л. 1. Подлинник. Пергамен – с. 35.*

2.3. 1284 г. Грамота смоленского князя Федора Ростиславича по судному делу о немецком «колоколе» (гире). (Акт № 5). *ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 6. Л. 3. Подлинник. Пергамен – с. 36.*

2.4. [1284–1297 гг., очевидно, ближе к 1285–1287 гг.] Грамота архиепископа рижского Иоганна II смоленскому князю Федору Ростиславичу с обвинением жителей города Витебска в необоснованной жалобе на рижан. (Акт № 6). *ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3. Современный список. Пергамен – с. 37.*

2.5. [1471 г.] августа 30. Грамота полоцкого воеводы Олехны Судимонтовича рижскому магистрату о готовности литовских послов начать переговоры с ливонскими послами. *ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 7. Л. 2. Подлинник – с. 38.*

2.6. [1656 г. не ранее июля 19]. Отписка боярина и воеводы Семена Лукьяновича Стрешнева царю Алексею Михайловичу об осаде города Динабурга. (Документ № 1). *Российский государственный архив древних актов (РГАДА). Ф. 210. Московский стол. Столб. 270. Л. 674–675. Подлинник – с. 42.*

2.7. 1656 г. [не ранее августа 9]. Отписка борисоглебского воеводы Василия Григорьевича Фефилатьева царю Алексею Михайловичу о приезде курмышан к городу Борисоглебскому (Динабургу). (Документ № 7). *РГАДА. Ф. 210. Московский стол. Столб. 285. Л. 91а–91б. Подлинник – с. 43.*

2.8. 1663 г. [не ранее ноября 20]. Отписка бывшего борисоглебского воеводы Никиты Степановича Бухвостова царю Алексею Михайловичу о невозможности передать новому воеводе Ивану Никитичу Сумарокову денежную казну и хлебные запасы без расписки. (Документ № 46). *РГАДА. Ф. 210. Белгородский стол. Столб. 518. Л. 103. Подлинник* – с. 44.

2.9. 1664 г. [не ранее апреля 14]. Отписка Ивана Никитича Сумарокова из Опочки царю Алексею Михайловичу о бедственном положении гарнизона города Борисоглебова (Динабурга), о бегстве солдат и о действиях литовских войск. (Документ № 50). *РГАДА. Ф. 210. Московский стол. Столб. 365. Л. 270. Подлинник* – с. 45.

Глава 3

3.1. Принципиальная структура XML-документа – с. 59.

3.2. Вариант электронной публикации документа. 1656 г. [не ранее июля 19]. Отписка С. Л. Стрешнева царю Алексею Михайловичу об осаде города Динабурга (воспроизведено начало документа). *РГАДА. Ф. 210. Московский стол. Столб. 270. Л. 674–675. Подлинник* – с. 73.

3.3. Графическая форма древнерусской буквы «аз» – с. 74.

3.4. Палеографическое воспроизведение древнерусского текста. [1284–1297 гг.] Грамота архиепископа рижского смоленскому князю Федору Ростиславичу (акт № 6). *ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 18. № 8. Л. 3. Современный список* – с. 76.

3.5. Деление текста статьи-обращения из договорной грамоты Новгорода с князем Ярославом Ярославичем, 60-е гг. XIII в. в исследовании С. М. Каштанова (Каштанов С. М. *Русская дипломатика*. Москва, 1988. С. 178) – с. 80.

3.6. Сравнение формуляров трех отписок (визуализация) – с. 106.

3.7. Пример выделения областей на растровом изображении источника для SVG-связывания с XML-файлом. [1471 г.] августа 30. Грамота полоцкого воеводы Олехны Судимонтовича. *ЛНА-ЛГИА. Ф. 673. Оп. 4. К. 19. № 7. Л. 2* – с. 108.

3.8. Временная линейка. Комплекс приказных документов XVII века по истории Динабурга – с. 111.

Глава 4

4.1. Тенденции в развитии Всемирной паутины – с. 114.

4.2. Функциональная структура информационной системы – с. 122.

4.3. Представление субкомплекса смоленско-рижских актов в виде открытой семантической сети (фрагмент) – с. 140.

4.4. Генерирование семантической сети на основе комплекса документов по истории Динабурга середины XVII века – с. 142.

4.5. Текст отписки воеводы С. Л. Стрешнева на сайте histdocs.referata.com – с. 147.

4.6. Связи между «историческими объектами», описывающие исторический факт – с. 156.

4.7. Репрезентация исторического факта в виде гиперграфа – с. 157.

4.8. Моделирование степени истинности высказываний – с. 158.

4.9. Выражение степени истинности в векторной логике – с. 159.

Summary

Computer-based Records Studies: Semantic Linkage of Information in Representation and Criticism of Historical Records

The monograph deals with the problems of historical records studies in the age of e-science. In the focus of the authors' attention, there are the potentialities of computer-based technologies in traditional spheres of historical research – historical records editing, external and internal source criticism, historical diplomatics, and auxiliary historical disciplines. The monograph is based on a number of pilot projects that have been implemented by the authors since 2000; the results have been presented at different international conferences and in a number of published research papers. Although this work reveals potentialities of computer-based technologies, first and foremost, in the field of representation and criticism of medieval documentary records related to the history of Russia, however, the same technologies can be efficiently used in an in-depth research of vast complexes of modern history records.

The impact of computer-based technologies on the above-mentioned traditional and fairly conservative fields of historical research poses a number of theoretical and methodological problems concerning the interconnection of traditional historical source studies and computer-based source studies, as well as the applicability and advantages of computer-based technologies in publication (representation on the Web) and exploratory analysis of written historical records. These problems are discussed in the preface to the monograph. In the authors' opinion, it is quite doubtful that computer-based records studies can be viewed as a separate auxiliary historical discipline, since this field of historical studies, which employs specific research methods and technologies, has no subject and object of its own. Therefore, computer-based records studies are methodologically integrated into the traditional historical records studies.

The first chapter “Increasing of Informative Output of Historical Records – a Chief Aim of Historical Records Studies” focuses on the methodological problems of representation of the structure of information provided by historical records. In order to reveal the informative potential of historical records to the full, two information layers – namely, expressed (overt) information and implicit information – should be retrieved from historical records. In the traditional source criticism, some

basic approaches to the investigation of these information layers have been recently substantiated. However, the disclosure of implicit information is based mainly on subjective, rather biased interpretations of evidences presented in historical records. It goes without saying that computer-based records studies can offer quite appropriate tools that might contribute to a comprehensive reconstruction of information structure of historical sources. In its turn, this reconstruction provides a solid basis for processing of information extracted from historical records by means of modern technologies and computer-based methods of external and internal source criticism. However, a prerequisite for an exhaustive disclosure of informative potential of any historical record is the linkage of its information with corresponding information provided by other records that form a definite system/ complex of interconnected historical sources.

Therefore, the second chapter “Complex Approach in Computer-based Records Studies and the Tasks of Investigation of the Complexes of Historical Sources” substantiates a complex approach in the computer-based records studies. According to this approach, ‘natural’ (historical) complexes of interconnected documentary records should be studied simultaneously, as a whole. Such complexes represent integral parts of definite systems that documents form in the course of performing their initial functions. The main features of a natural complex of historical records are as follows: first, common origin (provenance) of the documents that belong to a definite complex. As a rule, the common origin of historical records means that they are connected either with a concrete institution, or with a system of institutions, as well as with the thread of life of a person. Hence, any definite complex of documents comes into existence spontaneously; and researchers can only reconstruct such historical complexes of records. Second, within a natural complex of documents, there are close historical interconnections between the components of the complex. On the one hand, the links between documents can be direct ones, e.g., one document refers to another document and vice versa. On the other hand, there can be implicit links, too. For instance, historical records of different origin, which provide information about one and the same fact, represent the single complex of historical records. It should be also noted that hierarchical disposition within a system is a characteristic feature of any historical complex of documents. Such hierarchy reveals relative significance of the documents in the period of their circulation.

The empirical basis of the monograph encompasses three complexes of the documentary historical records. The first complex is represented by the Old Russian charters that reflect the course of relations between Riga and Smolensk in the 13th century. The second one embraces the charters related to mutual relations between Polotsk and Riga in the 1460s – 1470s. Both complexes form constituent parts of the vast collection of medieval and early modern records ‘Moscowitica–Ruthenica’ kept in the Latvian State Historical Archives – the structural department of the Latvian National Archives (Riga, Latvia). This collection of documents provides researchers with the firsthand information about the relations of Old Russian and Byelorussian lands and towns (Smolensk, Novgorod, Pskov, Polotsk, etc.), as well as Lithuania (later – Poland-Lithuania) with Riga, Livonia, Gothland, Hanseatic League, and some German towns in the late 12th – early 17th centuries. The third complex of historical records represents the official papers related to the history of Dünaburg (present-day Daugavpils) under the rule of Tsar Alexey Mihailovich in the middle of the 17th century.

The third and the fourth chapter of the monograph demonstrate some possibilities of applying the latest methods and approaches to the complex research of historical records. It should be noted that these methods and approaches have emerged due to the development of Semantic Web technologies. It seems that the principal characteristic feature of these technologies is their ‘universality’, i.e. they can be applied in the complex studies of any historical records that represent different epochs. In this respect, close attention should be devoted to text markup technologies, since text markup provides opportunities to reveal both formal structure and semantics of documentary historical records. At the same time, text markup can be conducive to the semantic linkage of information provided by different historical records within definite complexes. Therefore, the third chapter “Markup of the Texts of Historical Records – Means of Representation, Processing, and Semantic Linkage of Records’ Information” deals with the problems of text markup that can reveal simultaneously the formal structure and the contents of the documents. In this chapter, special attention is paid to the markup schemes TEI (Text Encoding Initiative) and CEI (Charters Encoding Initiative) developed on the basis of XML. The authors of the monograph show the applicability of CEI and TEI to comprehensive representation of paleographic features of historical sources as well as meta-information related to historical records. However, in the focus of the authors’ attention there are ‘informative models’ of the structure and contents of historical records. These models

are represented by means of text markup technologies. A paramount importance is attached to the representation of information about persons, places, institutions, terms, events, documents (actually, references to documents), etc., which are directly or indirectly mentioned in the texts of historical records. These aspects of information can be adequately revealed using both CEI and TEI markup schemes supplemented with additional elements. This ‘semantic markup’ can be performed simultaneously with the structural markup of texts, which forms a basis for further pattern analysis of documentary historical records.

Nowadays, when semantic publications of historical documents have become quite topical, revealing the formal structure and semantics of documents is a prerequisite for the representation of the complexes of historical records in semantic networks, as well as for the production of qualitative semantic publications, which are based on a detailed markup of texts. The fourth chapter “Semantic Web Technologies and Possibilities of Direct Linkage of Historical Records’ Information” shows the advantages of semantic publications of historical records. Firstly, such publications provide historians with up-to-date tools for external and internal source criticism. Secondly, semantic editions provide better facilities for searching for information, because such publications draw nearer search algorithms used by humans and computers. Thirdly, since formalized knowledge can generate new knowledge, semantic publications can be used as knowledge bases in order to advance hypotheses for further research by means of automatic inference. In addition, such semantic publications can provide a solid basis for specific Web information systems that incorporate texts of historical records, research tools, and research results into knowledge-based systems, which are created for a network community.

In the semantic publication, marked up objects should be linked either with analogous objects revealed in other historical records, or with corresponding data, which are represented in different specialized – historical – ontologies. However, ontologies that have been recently created for the purposes of historical research (e.g., CIDOC CRM, Pearl Harbor Project, CultureSampo Project, etc.) cannot meet all the requirements of the semantic publications of documentary records, since they do not reflect the specificity of written historical sources. Therefore, the ontology specially designed for the semantic publication of these historical sources should be created. In contrast with the event-oriented approach accepted in the above-mentioned ontologies, the authors of the monograph propose a document-oriented approach to the representation

of historical data and the description of different historical objects – documentary records, persons, places, historical events, etc. This ontology reflects two kinds of semantic links, which are constituted in the semantic publication. First, there are the links between historically and thematically interconnected records. These interconnections emerged when the documents were drawn up, as well as during their circulation. Second, there are the links to other historical records, which do not belong to a definite complex of historical sources. These links create a definite semantic network that represents the knowledge about the charters. Actually, the semantic network, which is described in the monograph, provides a model for capturing extensive information related to the complex of historical records; therefore, it can form a solid basis for the production of a semantic publication of documentary records. However, the production of semantic publications based on ontologies, which are directly recorded using RDF or OWL, is time-consuming. It seems that the tools provided by semantic Wiki-systems (e.g. Semantic MediaWiki – SMW) can facilitate this process. Therefore, in the monograph, the advantages and opportunities afforded by Wiki-systems in the production of semantic publications of documentary records are examined.

Unfortunately, the semantic layers that capture diverse marked up objects – structural divisions of texts, historical persons, place-names, events, meta-information, etc. – do not represent the tenor of historical records exhaustively. Nowadays, the possibility to fully represent the contents of historical records is afforded by controlled natural languages, which are understood not only by a human reader, but also by a computer. The monograph shows that one of these languages – Attempto Controlled English (ACE) – seems to be a very promising tool for the purposes of computer-based records studies.

Actually, the representation of historical records on the Web and the semantic linkage of their information is a preparatory work for performing synthetic operations, i.e. reconstruction of the aspects of historical reality on the basis of evidences provided by historical records. Therefore, in perspective, historical information retrieved from records should be incorporated into a knowledge-based information system that can be described as a specialized semantic network (Historical Semantic Network – HSN) created for the purposes of synthetic operations. The model of HSN is substantiated in the monograph. The first feature of HSN is the principal role assigned to historical records. It means that connecting links between the objects of the network – the data related to

events, persons, historical monuments, institutions, geographical sites, and so on, and so forth – are constituted (actually, reconstructed) in accordance with the evidences of historical sources. The links between the objects of HSN provide definite visions of historical facts. The second characteristic feature of HSN is the fuzziness of almost all aspects of information in the semantic network. The source of this fuzziness is the uncertainty of expert evaluations and interpretations of historical data; moreover, evidences of historical records can be either fragmentary and contradictory, or doubtful and uncertain. Therefore, the level of validity of information is ‘measured’ by variables, which can be expressed by ordinary numbers (0–1). Actually, they can be expressed by vector quantities as pairs (true, false); the components – true and false – can assume values from 0 to 1. The fuzziness values defined for some objects can be propagated within a semantic network by means of logical reasoning. This ‘measurement’ of validity of data is performed by users of HSN; the results of measurement should be changeable and complementary, thus reflecting the cooperation of researchers within the frameworks of a network community.

All in all, this monograph reveals the basic trends in the application of modern computer-based technologies that can be conducive to representation on the Web and in-depth analysis of vast complexes of historical records, as well as to synthetic operations. On a higher technological level, computer-based source studies can perform the following tasks: in-depth analysis of internal structure and tenor of historical records; retrieval of implicit information from historical records; verification of evidences provided by historical records; semantic linkage and aggregation of data extracted from sources; filling gaps in source information and generation of new knowledge.

Об авторах

Алексей Геннадьевич Варфоломеев (1960). Канд. физ.-мат. наук (1994). Декан математического факультета, доцент кафедры теории вероятностей и анализа данных Петрозаводского государственного университета. Член Совета Ассоциации «История и компьютер». Сферы научных интересов: информационные технологии поддержки научных исследований, математические методы в анализе текстов и изображений исторических источников (80 научных публикаций).

Александр Степанович Иванов (1964). Dr.hist. (1993). Профессор (Даугавпилсский университет, Институт исследования Латгалии, Даугавпилс, Латвия), ведущий исследователь (Институт регионалистики – REGI, Резекненская высшая школа, Резекне, Латвия). Сферы научных интересов: источниковедение, археография, историческая дипломатика, компьютерные технологии в источниковедении и публикации исторических источников, историография (130 научных публикаций).

Научное издание

Варфоломеев Алексей Геннадьевич

Иванов Александр Степанович

КОМПЬЮТЕРНОЕ ИСТОЧНИКОВЕДЕНИЕ

**Семантическое связывание информации
в репрезентации и критике
исторических источников**

Редактор *В. Е. Бюлер*

Художественный редактор *В. А. Басимов*

Подписано в печать 20.03.13. Формат 60x84 1/16

Бумага офсетная. Уч.-изд. л. 11,1

Тираж 500 экз. Изд. № 62.

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Отпечатано в РЦНИТ Петрозаводского государственного университета
185910, г. Петрозаводск, пр. Ленина, 33

ISBN 978-5-8021-1573-2



9 785802 115732