

ДИСКУССИИ И ОБСУЖДЕНИЯ

DISCUSSION AND DEBATE

20 ЛЕТ КОМПЬЮТЕРНОМУ ИСТОЧНИКОВЕДЕНИЮ: PER ASPERA AD ASTRA?

20 YEARS OF COMPUTERIZED SOURCE STUDIES: PER ASPERA AD ASTRA?

Варфоломеев А. Г., Иванов А. С.

Компьютерное источниковедение. Семантическое связывание информации в репрезентации и критике исторических источников. Петрозаводск : ПетрГУ, 2013. 202 с. ISBN 978-5-8021-1573-2

Многообразие и безграничность взаимосвязей, присущих явлениям объективного мира, обуславливают то, что в исторических источниках, несмотря на избирательное отражение ими действительности, содержится в сущности неисчерпаемый объем скрытой информации, характеризующий взаимосвязи явлений действительности.

И. Д. Ковальченко

Термин «компьютерное источниковедение» появился в 1993 г., споры о его сути, применимости, полноте и точности не утихают вот уже 20 лет, а сам термин все еще часто используется в кавычках¹. Вместе с тем терминологический спор сегодня удачным образом перешел в методическое русло, позволив разделить теоретические дебаты и практические решения. Книга А. Г. Варфоломеева и А. С. Иванова — замечательный пример целена-

правленного движения к решению задачи источниковедческого инструментария историков на современном этапе развития компьютерной техники и информационных технологий. Авторы широко известны как междисциплинарный и творческий дуэт, смело подступающий к самым сложным и важным задачам источниковедческих исследований. Монография «Компьютерное источниковедение» — важная веха на этом пути, который авторы прошли в ногу со спорами о заглавном термине, применяя новейшие компьютерные технологии для репрезентации и изучения средневековых письменных источников по истории России².

Книга посвящена нескольким крайне важным для современной практики историка-исследователя вопросам: во-первых, повышению информа-

¹ Впервые термин прозвучал в 1992 г. на конференции в Ужгороде, а опубликован был в 1993 г. в работе: Моисеенко Т., Свищев М. Изучение аграрной истории России последних десятилетий: перспективы «компьютерного источниковедения» // История и компьютер: новые информационные технологии в исторических исследованиях и образовании / под ред. Л. Бородкина и В. Леверманна. Gottingen, 1993. С. 152. Первоначально «компьютерное источниковедение» отождествлялось с исторической информатикой. Подробнее о термине и актуальных проблемах источниковедения в контексте развития исторической информатики см.: Гарскова И. М. Источниковедческие проблемы исторической информатики // Российская история. 2010. № 3. С. 155–156.

² Одна из ключевых работ авторов уже отпраздновала 10-летие: Иванов А. С., Варфоломеев А. Г. Технология XML как инструмент компьютерного источниковедения (на примере формулярного анализа документов приказного делопроизводства) // Круг идей: алгоритмы и технологии исторической информатики. М., 2005. С. 241–281.

ционной отдачи источников, во-вторых, возможностям комплексного подхода к историческим источникам, в-третьих, проблемам семантической разметки источников и, в-четвертых, актуальным возможностям связывания информации исторических источников³.

«В наши дни компьютерное источниковедение все еще находится в процессе становления, а в среде специалистов существуют разногласия в определении его предмета» (с. 6). При этом авторы подчеркивают, что «можно говорить лишь о *направлении* в рамках исторического источниковедения», которое «возможно лишь в относительно кратковременной перспективе — пока компьютерные технологии являются чем-то «новым» и «необычным» для значительной части историков и источниковедов» (с. 8). Если посмотреть на эволюцию определения, то оно не сильно изменилось за 20 лет. Так, В. С. Тяжельникова в 1995 г. давала следующее определение: «компьютерное источниковедение — это новое научное направление, возникшее под влиянием теории информации на стыке исторической информатики и источниковедения. Оно изучает как традиционные проблемы источниковедения средствами новых компьютерных технологий, так и источниковедческие проблемы, порождаемые теорией информации и развитием информационных технологий»⁴. При этом авторы монографии скорее проявили скромность, подчеркивая кратковременный эффект от компьютерного источниковедения, потому что суть их исследования, показывающего потенциальные возможности компьютерных технологий в репрезентации, обработке, изучении и интерпретации информации слабоструктурированных уникальных средневековых источников по истории России, не просто новаторская, но еще и отрывающая широкие перспективы для будущих изысканий.

Авторы определяют центральную задачу источниковедения в традициях школы источниковедения И. Д. Ковальченко как повышение инфор-

мативной отдачи источников⁵. При этом главным вкладом теории информации в источниковедческую практику, с точки зрения авторов, является переход от «свидетельств исторических источников» к введению в научный арсенал понятия «информация исторических источников», а вместе с этим и новаторский путь к извлечению скрытой (структурной) информации. Важная задача монографии «Компьютерное источниковедение» — сделать следующий шаг, перейти от постулатов к методике решения источниковедческих задач.

Первым этапом в компьютерном источниковедении можно назвать формирование комплекса источников, обогащая новыми технологиями традицию комплексного источниковедческого подхода В. Л. Черепнина, С. О. Шмидта, В. Л. Янина. По мнению авторов, «естественному» комплексу источников присущи общность происхождения, спонтанность формирования комплекса, иерархичность его структуры и наличие устойчивых связей между источниками (с. 24). Данный подход, действительно, позволяет взглянуть на источники системно, однако он не лишен и некоторых методологических проблем. Главная проблема признается самими авторами: реконструкция источникового комплекса — процедура достаточно гипотетичная, особенно когда речь идет о неопределенном количестве утраченных со временем источников. Пользуясь достижениями археографии, авторы рассматривают сильные и слабые стороны тематического, систематического и регионального подходов в реконструкции комплексов источников. Комплексный подход весьма сложен, прежде всего в силу физической рассредоточенности источников, но такой подход позволяет перейти от «иллюстративного» к «сплошному» изучению исторических источников⁶.

Бесспорным достижением книги является то, что теоретические постулаты проверяются на практике, в частности, подробно освещаются принци-

³ Данные вопросы становятся особенно актуальными в рамках дискуссии о новых цифровых исследовательских практиках историков. См., напр.: Mussel J. Doing and making: history as digital practice // *History in the Digital Age* / Ed. by Toni Weller. Routledge, 2013. P. 79–94.

⁴ Тяжельникова В. С. Компьютерное источниковедение: к постановке проблемы // *Круг идей: развитие исторической информатики*. М., 1995. С. 272. Похожее определение представлено в работе: Владимирова В. Н., Цыб С. В. Источниковедение в век компьютера (вместо предисловия) // *Источник, метод, компьютер*. Барнаул, 1996. С. 3–8: «компьютерное источниковедение» — это источниковедческая работа (и ее результаты), ориентированная на определенный набор компьютерных технологий».

⁵ См., напр.: Ковальченко И. Д. Методы исторического исследования. 2-е изд. М., 2003. С. 120: «Главная из прикладных задач источниковедения состоит во все более явственно обнаруживающейся потребности повышения информативной отдачи источников». Подробнее об этом авторы уже писали в докладе: Иванов А. С., Варфоломеев А. Г. Идеи И. Д. Ковальченко об информационной ценности и неисчерпаемости источника в современном компьютерном источниковедении // *Идеи академика И. Д. Ковальченко в XXI веке*. М., 2009. С. 154–164.

⁶ Интересно отметить, что данные размышления авторов вписываются в весьма пылкую дискуссию вокруг книги Франко Моретти (Moretti) “Distant reading” (L.; N. Y., 2013), в которой практика «пристального чтения» (close reading) противопоставляется «отдаленному чтению» (distant reading), состоящему в количественном и систематическом анализе большого корпуса текстов (комплекса документов).

пы построения двух источниковых комплексов: «Moscovitica — Ruthenica» (XIII–XV вв.) и Динабургского комплекса (XVII в.). На примере анализа этих комплексов становится очевиден «парадокс достаточного основания», когда именно исследователь может (и может ли вообще?) утверждать, что источниковый комплекс сформирован полностью. Авторы находят весьма изящный ответ на этот вопрос, заявляя, что «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» (с. 48).

Полезным для исследователей и источниковедов станет и алгоритм применения инструментария компьютерного источниковедения, который разделяется на несколько этапов: 1) перевод текстов источников и метаданных о них в электронную форму; 2) разметка текстов и метаданных; 3) связывание информации в источник-ориентированных базах данных или семантических сетях; 4) применение аналитического инструментария компьютерного источниковедения; 5) генерация новых знаний на основе автоматического вывода при творческом участии исследователя (с. 54).

При этом невозможно полностью согласиться с мыслью авторов, что источниковеды при выборе способа цифрового представления письменного источника оказываются перед дилеммой — изображение или текст (с. 55). В работах последних лет скорее заметна тенденция характеризовать переход от изображения к тексту (ср. классическое «установление текста») как последовательные этапы оцифровки⁷. Например, Я. Грегори отмечает, что по сложившимся традициям оцифровки сначала создается отсканированная (факсимильная) копия, а затем на её основе устанавливается текст и проводится кодирование (часто самые затратные по времени и средствам исследовательские процедуры)⁸. Очевидно, что для полноцен-

ной разметки источника исследователь должен работать с текстом, однако не менее важно избежать частой подмены источника его установленным текстом. В своем исследовании авторы обосновывают пользу для задач источниковедения древовидной организации информации (на примере XML-дерева, реализованной в стандарте TEI), при этом подчеркивая необходимость сохранения метаданных об источниках, равно как и сведений об их палеографических особенностях. Как показано на конкретных примерах (в частности, электронной публикации Отписки С. Л. Стешнева царю Алексею Михайловичу 1656 г.), технология XML является удобным средством для дипломатически и палеографически точного воспроизведения текстов. Более того, данная технология при условии тщательно проведенной многоуровневой разметки текстов позволяет проводить (и частично автоматизировать) формулярный анализ комплекса документов⁹.

Авторы справедливо выступают против убеждения, что работа с XML-документом лишь воспроизводит основные этапы традиционной источниковедческой критики. Они указывают на то, что XML-подход позволяет проводить комплексное изучение неограниченного комплекса источников (что, заметим, никак не отменяет большого и кропотливого труда по созданию таких комплексов), любые операции могут быть гибко подлажены под задачи конкретного исследования, а XML-документ как динамическую базу данных можно использовать для самых разных исследовательских задач. Однако авторы напоминают, что важно не увлекаться чисто технологической стороной XML-технологии, а помнить о задаче извлечения информации и ее интерпретации (с. 112–113).

Особое внимание в книге уделено источниковедческим возможностям семантических сетей. Важным этапом комплексного подхода является переход от XML-описания к смысловому связыванию источников, в чем удобным подспорьем могут стать семантические сети, реализуемые в различных «исследовательских средах» (VRE, virtual research environment). По опыту создания информационной системы «Источник» авторы формулируют принципы, необходимые для систем организации работы сетевых сообществ исследователей письменных источников: во-первых, это сплошная

⁷ Особенности цифровых вызовов в истории рассматриваются, например, в статье: Бородкин Л. Digital history: применение цифровых медиа в сохранении историко-культурного наследия? // Историческая информатика. 2012. Т. 1, № 1. С. 14–21.

⁸ Gregory Ian (2014) Challenges and opportunities for digital history. *Front. Digit. Humanit.* 1:1. doi: 10.3389/fdigh.2014.00001 Грегори идет в своих рассуждениях ещё дальше, утверждая, что современное поколение историков — первое, которое сталкивается со сложными и неполными цифровыми (или оцифрованными) источниками, кишущими ошибками.

⁹ Сегодня возможности применения XML, особенно к средневековым текстам, являются одной из важных тенденций исторической информатики, см., напр.: Burkard B., Vogeler G., Gruner S. Informatics for Historians: Tools for Medieval Document XML Markup, and their Impact on the History-Sciences // *Journal of Universal Computer Science*. Vol 14. № 2 (2007). P. 193–210.

репрезентация комплекса источников, во-вторых, модульность системы, позволяющая варьировать набор необходимых инструментов в зависимости от потребностей исследователя, в-третьих, возможность многоцелевого использования информации (от простого поиска до сложного исследования), в-четвертых, возможность накопления и представления знаний (методов и результатов критики источников) и, наконец, в-пятых, открытость системы (с. 121–126)¹⁰.

Следует учитывать, что информационные системы создают значительный объем задокументированных знаний, равно как и порождают неявные знания, часто сопряженные с накопленным опытом экспертов. Экспертное знание, таким образом, становится драйвером создания сетевых сообществ исследователей (*community of practice*). Так решается важный практический вопрос об инфраструктурной основе для развития и совершенствования инструментария современного компьютерного источниковедения.

Первостепенной становится задача создания сетевыми сообществами исследователей полноценных и всесторонних семантических публикаций источниковых комплексов с инструментами их аналитической и синтетической критики. По сути, семантическая публикация предполагает работу с тремя смысловыми слоями: 1) термины и концепты, являющиеся основой разметки и желательные связанные с контролируруемыми тезаурусами и онтологиями; 2) метainформация о тексте, выраженная в форме триплетов «субъект — предикат — объект»; 3) информация о ходе исследования, задачах, гипотезах, логике дискуссий (с. 128–129).

Получается, что семантические публикации повышают качество поиска и связывания информации, сближая формальный алгоритмический поиск с исследовательским эвристическим поиском. Тем более, что семантические публикации как базы знаний могут стать основой для процедур автоматического вывода и выдвижения новых гипотез. Например, на основе семантических публикаций возможно представить структуру исторических источников в виде графов, позволяющих оценить сложность и подобие изучаемых структур. Для записи правил и вывода знаний о семантической структуре текстов исторических источников авторы рекомендуют стандарт RuleML.

В монографии приводятся также примеры построения семантических сетей на осно-

ве Semantic Media Wiki (SMW). К сожалению, о возможности использования XML-документов в SMW сказано лишь как об исследовательской перспективе (например, на основе XML2Wiki). Но даже с помощью упрощенных транскрипций и переводов средневековых текстов удалось выявить принципиальные ограничения вики-технологии — в рамках SMW невозможно создавать «вложенные» и «пересекающиеся» текстовые фрагменты, соединять вместе структурную и семантическую разметки, использовать стандарты TEI или CEI, привязывать текст к изображению источника.

Особый интерес вызывает опыт использования контролируемого естественного языка Attempto Controlled English (ACE), позволяющего тексты исторического источника перевести в формулу логики предикатов первого порядка, над которыми можно производить логический вывод, получая новые знания. Вероятную критику со стороны источниковедов в практике использования ACE вызовет необходимость перевода текста оригинала, однако авторы считают процедуру перевода техническим подспорьем, позволяющим решить задачу многоаспектного семантического связывания информации из нескольких источников, причем ACE позволяет включить даже больший объем источниковой информации, чем SMW, TEI или CEI, с возможностью при накоплении достаточного количества информации даже автоматического вывода.

Авторами монографии разработан очень интересный пример модели исторической семантической сети (HSN — Historical Semantic Network) для поддержки исследований по региональной истории — latgalesdati.lv — Латгальского края. Принципиальное отличие такой семантической сети состоит в том, что она является не только хранилищем ретроспективной информации, но и исследовательским инструментом. В основу модели HSN положены «исторические объекты», т. е. прежде всего исторические источники, а также лица, организации, топонимы и т. д. На основе исторических источников происходит реконструирование исторических объектов с их упорядочением и связыванием. На основе связей объектов формируется информация об исторических событиях. Концепция HSN позволяет проводить различные операции: установление идентичности объектов, поиск похожих событий, обнаружение и оценка противоречивости информации, обобщение информации из различных источников.

Сегодняшние «цифровые вызовы» исторической профессии требуют системного ответа. В условиях формирования пусть и аморфного представ-

¹⁰ Данные принципы созвучны таким международным инициативам, как «Скрипто» (см.: *Scripto — community transcription tool*. URL: <http://scripto.org/documentation/about/>).

ления о «цифровых гуманитарных науках»¹¹ все больший спрос у ученых возникает на программные инструменты и реально применимые прототипы для источниковедческих, археографических, дипломатических, палеографических исследований. В «Компьютерном источниковедении» А. Г. Варфоломеев и А. С. Иванов представляют широкий спектр современных компьютерных технологий, применимых для широкого круга источниковедческих задач. Россыпь технологий и подходов, уже апробированных на конкретно-историческом ма-

териале, может быть воспринята и применена в источниковедческой практике. Компьютерное источниковедение (уже без кавычек) ждёт новых исследователей, которые вновь и вновь будут ставить вопрос о том, что же такое компьютерное источниковедение. И есть надежда, что ответы на этот ключевой вопрос исторической профессии в наш цифровой век будут инновационными и творческими.

*А. Ю. Володин
(МГУ имени М. В. Ломоносова)*

¹¹ Таллер М. Дискуссии вокруг Digital Humanities // Историческая информатика. 2012. № 1. С. 5–13; Володин А. Ю. Digital Humanities (цифровые гуманитарные науки): в поисках самоопределения // Вестник Пермского университета. Серия «История». 2014. № 3. С. 5–12.
