

Технологии электронного текста как основа информатизации исторической науки и образования

Д. А. Гагарина

Пермский государственный университет, Пермь, Россия

Текст, будучи преобладающей формой существования исторических источников и практически единственной формой исторических публикаций, является особым объектом для историка. Информатизация исторической науки и исторического образования в последние десятилетия выявила и апробировала различные направления, связанные с использованием как традиционных технологий и программных продуктов, так и специализированных. Однако, на наш взгляд, именно информационные технологии работы с текстом являются во многих смыслах основой информатизации в области истории. Существует несколько факторов, позволяющих утверждать это.

Очевидно, электронные тексты являются наиболее массовыми и распространенными среди прочих электронных исторических ресурсов. Свидетельством тому являются количественные показатели оцифровки исторических источников и публикаций, создания электронных коллекций и библиотек и т.п., что в свою очередь предоставляет новые возможности доступа историков-исследователей к самым разным источникам. Однако, несмотря на важность обозначенных фактов, количество опубликованных источников само по себе не может рассматриваться как основной фактор эффективности использования электронных исторических ресурсов.

Текст является для исторической науки и образования и средством (например, учебники), и предметом (например, исторические источники) изучения, и в этом также состоит его особая роль, несвойственная другим формам представления информации. В том случае, если источник имеется в электронном виде, может происходить интеграция средства обучения, объекта и предмета исследования в рамках одного конкретного текста. Так, за счет использования языков разметки или гиперссылок в текст исторического источника может быть заложен алгоритм его изучения или модель его интерпретации, в том числе вариативные. С использованием элементов

разметки (HTML, XML, TEI) текст приобретает качественно новые возможности и значение: традиционное понятие текста расширяется за счет введения в него структуры, нелинейности, установления перекрестных связей, гипертекст позволяет реализовать изучение механизма образования ассоциативных связей, задать алгоритм его прочтения, что особенно важно при изучении студентами исторических источников. Процесс такой разметки текста является элементом моделирования информации, представленной в тексте, он важен как с точки зрения изучения рассматриваемого текста, так и с точки зрения формирования соответствующего умения.

Вместе с тем, еще более важным, на наш взгляд, является следующее: электронный текст, в отличие от многих других ИКТ, используемых историками, – «прямой наследник» традиционного (печатного) текста. (Так, столь востребованные исторической информатикой технологии, как, например, базы данных и мультимедиа гораздо в большей степени «самодостаточны» по отношению к их прародителям.) При этом электронная форма, сохраняя смысл и предназначение текста как такового, допускает качественно новые возможности работы, позволяет историку использовать исследовательские инструменты, не доступные ранее, не отрицая при этом традиционные технологии работы с текстом «классического» источниковедения.

Здесь же можно отметить, что в случае с созданием и использованием электронного текста, как правило, не происходит подмены исследовательского компонента технологией (что часто наблюдается с базами данных, когда сам факт создания базы может выдаваться за научный результат). Иначе говоря, электронный текст позволяет плавно, поэтапно, прозрачно сформировать предметную информационную компетентность в области истории.

Таким образом, формирование умений работы с электронным текстом становится одной из важнейших задач исторического образования в области ИКТ. Освоение текстовых технологий студентами-историками включает традиционные составляющие: создание, хранение, визуализация

(представление) и анализ текста, которые реализуются в рамках курсов информатики и исторической информатики.

Изучение обозначенных составляющих должно явно и неявно сопровождаться выявлением, во-первых, отличий электронного текста от традиционного печатного, во-вторых, особенностей работы именно с историческим текстом.

Что касается отличий электронного текста от печатного, то сами отличия и методика их рассмотрения могут являться темой отдельного издания. Здесь же ограничимся следующим примером. Рассматривая текст как объект исследования и пытаясь моделировать его, можно вводить разные уровни формализации. На низшем уровне текст – последовательность символов некоторого алфавита. Такой уровень «понятен» компьютеру – именно посимвольно мы вводим текст с помощью клавиатуры, но уже и эта модель может быть полезна с точки зрения истории. Далее можно рассматривать текст как последовательность слов, потом учитывать в модели словоформы, словосочетания, предложения, абзацы, параграфы, главы и т.д. Наконец, на некотором уровне в модели может каким-то образом учитываться смысл текста (например, через представление его семантическими схемами и т.п.), а также скрытая в тексте информация. Но, читая текст, историк-исследователь в первую очередь воспринимает его с точки зрения смысла – в сознании человека текст трудно формализуем. Электронный текст обуславливает этим одно из первых противоречий при работе с ними студента (или исследователя). Разрешение этого противоречия способствует как лучшему пониманию сущности текста вообще (независимо от формы представления), так и подготовке к использованию технологий компьютерной лингвистики, например, контент-анализа.

Наиболее дидактически емкий метод выявления особенностей работы с историческим текстом по сравнению с другими – это непосредственная работа с историческими источниками как в рамках, например, проектной деятельности, так и через вовлечение студентов в реализацию научных проектов. Так, в лаборатории исторической информатики Пермского государственного университета реализуются проекты создания

полнотекстовых информационных систем¹, обеспечивающих весь цикл работ с электронным текстом: от оцифровки источников до их представления в информационной системе и последующего анализа. Участие студентов в работе на разных этапах реализации этих и других проектов позволяет на практике познакомить их с традиционными проблемами сохранности источников и необходимости качественной оцифровки, с особенностями работы с древними и средневековыми текстовыми источниками, с выбором форматов для хранения и представления источников и др.

В заключение хотелось бы подчеркнуть, что вышесказанное ни в коем случае не принижает значения других видов ИКТ для исторической науки и образования, а лишь свидетельствует об особой роли текстовых технологий в процессе информатизации как исторической области в целом, так и в формировании предметной информационной компетентности ее представителей.

Digital text technologies as the basis for wider adoption of ICT in history scholarship and education

Dinara A. Gagarina

Perm State University, Perm, Russia

Text is traditionally the basic form of historical information. This paper discusses why electronic text will lead the way to the wider adoption of information and communication technologies in history scholarship and education.

¹ «Российские парламентарии начала XX века», грант РГНФ № 03-01-12012в, 2003-2004 гг.; «Журналы земских собраний как источник изучения истории местного самоуправления в России (II половина XIX – начало XX века). Информационная система», грант РФФИ № 04-07-90283-в, 2004-2006 гг.; «Источниковедческий анализ газеты «Пермские губернские ведомости» (XIX – начало XX вв.) на основе применения новых компьютерных технологий», грант РГНФ № 05-01-82102 а/У, 2004-2006 гг.; «Газета «Пермские губернские ведомости» 1838 – 1844 гг.: проблемы источниковедческого исследования и сохранения историко-культурного памятника», Грант РГНФ № 08-01-82101 а/У, 2008-2009 гг.; «Стенографические отчеты Государственной Думы, 1906–1917», грант РГНФ № 08-01-12108в., 2008-2010 гг.