

Теория и методология исторической информатики

Бородкин Л.И.

Приоритеты современной исторической информатики: технологии *e-Science*

Первое десятилетие XXI века принесло немало инноваций в информационную структуру науки и образования. Информационно-коммуникационные технологии (ИКТ) играют всё более активную роль и в формировании новой инфраструктуры современных исторических исследований, университетского образования.

В 2007 г. произошло «переформатирование» журнала «History and Computing», учрежденного в 1989 г. Международной ассоциацией «History & Computing» (АНС). Почти 20 лет работы редколлегии этого журнала создали ему репутацию ведущего периодического издания в области информатизации гуманитарного знания. Традиции этого издания будут продолжены в новом журнале – «International Journal of Humanities and Arts Computing» (*IJHAC*), который заявлен в статье редакторов, открывающей первый выпуск¹, в качестве преемника журнала «History and Computing». Новое название отражает тенденцию к усилению междисциплинарности современных исследований в области гуманитарных наук. Журнал представляет теперь сотрудничество трех научных организаций: International Association for History and Computing, Cultural Atlas Initiative, Digital Resources in the Humanities and Arts. Приоритетная тематика, которую будет поддерживать журнал, включает технологии визуализации, цифровые медиа, моделирование, ГИС в гуманитарных приложениях, информационную архитектуру², *e-Science*, Web-сервисы, опыт применения грид-технологий и другие направления современных разработок ИКТ. Приветствуются

также статьи и обзоры по электронным ресурсам и электронным публикациям (*e-Resources* и *e-Publications*).

Появление здесь ряда новых для гуманитарной аудитории терминов и понятий требует комментария: ведь они в ближайшие годы войдут в активный лексикон исторической информатики, а затем – и в общенаучный лексикон гуманитариев. Этот процесс уже идет, его инициировали известные голландские специалисты по исторической информатике, авторы программной работы о прошлом, настоящем и будущем этой междисциплинарной области – Онно Боонстра, Питер Доорн, Лен Брёре³.

Немного о новых направлениях ИКТ

Наиболее характерной чертой развития ИКТ в 2000-х гг. является формирование новых компонентов электронной среды взаимодействия. Этот процесс породил новые методологические подходы, программные и аппаратные решения, новые термины и понятия. Широкое распространение получили новые категории: *e-Science*, *e-Social Science*, *e-data*, *e-Research*, *e-Publications*, *e-Culture*, *e-Learning*, *e-Business*, *e-Banking*, *e-Medicine*, *e-Government* и т.д., связанные с развитием *распределенной инфраструктуры* информационных технологий и ресурсов. Наверное, излишним будет напоминание, что "e" означает здесь "*electronic*". Нас интересует прежде всего содержание термина *e-Science*, возникшего изначально в Великобритании, где крупные исследовательские проекты в этой области начались в 2001 г. Именно там было дано первое определение *e-Science*, получившее в дальнейшем широкое распространение: «научно-технологическая область, в которой всё возрастающую роль играют распределенное глобальное взаимодействие посредством сети Интернет, с использованием очень больших коллекций данных, компьютерных ресурсов терауровня и высококачественной визуализации, доступных индивидуальному пользователю»⁴. Надо сказать, что в русском языке термин *e-Science* существует пока преимущественно в англоязычном варианте⁵.

Иногда возникает вопрос: в чем отличие концепции *e-Science* от хорошо известной нам глобальной сети Интернет? Ведь пользователь Интернета имеет доступ со своего компьютера к разнообразным ресурсам глобальной сети, размещенным на сайтах всех стран мира; при этом набор веб-сервисов непрерывно расширяется. Дело в том, что кроме «обычной» информации, размещенной в Интернете, специалисты, работающие в сфере науки и образования, нуждаются в доступе к

крупномасштабным информационным массивам, базам данных, имеющим объемы памяти, измеряемые терабайтами. Работа с такими массивами требует вычислительных мощностей с производительностью уровня терафлоп. Обычный пользователь Интернета работать в этом режиме не сможет. Сегодня потребность в таких крупномасштабных ресурсах и компьютерных мощностях испытывают в основном ученые-естественники, но в ближайшем будущем к этому рубежу подойдут и разработчики проектов в различных областях социально-гуманитарного знания⁶. Задача *e-Science*, таким образом, – создание организационных и технологических структур, разработка соответствующего программного обеспечения для функционирования новой информационной среды с распределенными ресурсами (информационными и вычислительными), обеспечивающих доступ к ним индивидуальных пользователей, исследовательских групп, лабораторий и институтов (эти сообщества называют *виртуальными организациями, collaboratories*).

Основное русло реализации задач *e-Science* прокладывают *grid-технологии*. Эта концепция (нередко ее называют Grid Computing – распределенные сети, или "решетки" вычислительных ресурсов) соответствует одному из ведущих и перспективных направлений развития ИКТ. В последние годы начата разработка средств поддержки Grid на уровне стандартов и продуктов. Ситуацию можно сравнить с состоянием сети WWW в конце 1990-х гг., при этом специалисты предсказывают, что Grid произведет такую же революцию в области обработки данных, какую Интернет произвел в сфере компьютерных коммуникаций. Так, корпорация *Oracle* развивает свои продукты в направлении более широкого применения Grid-технологий (символично, что в обозначении очередной, десятой версии Oracle появилась латинская буква *g*). Следует отметить, что в платформенных решениях Oracle и других ведущих поставщиков ИКТ уже давно наметилась ориентация на распределенные вычисления на основе кластерных систем (в том числе на платформе Intel Itanium и Linux)⁷. Grid-технологии используют новую архитектуру программного обеспечения, предназначенную для объединения большого количества недорогих модулей памяти и серверов, создающих распределенный виртуальный ресурс, транспарентный и легко доступный для пользователя (независимо от места нахождения его компьютера). Одна из наиболее сильных сторон Grid-технологий – новые возможности эффективного распространения информации и обучение с доступом к обширным ресурсам⁸. Благодаря всё более широкому применению современных ИКТ в образо-

вании, доступ к грид-инфраструктуре получают школы и вузы, исследовательские группы и широкий круг пользователей различных приложений, что приведёт к качественным изменениям в учебном процессе, научных исследованиях, и, в конечном итоге, в разных видах экономической деятельности. При этом специфика термина *e-Social Sciences* заключается не только в предметной области используемых распределенных информационных ресурсов, но и в акценте на взаимодействие ученых-обществоведов со специалистами по ИКТ⁹.

В этой связи представляет интерес деятельность Национального Центра по развитию *e*-социальных наук (NCeSS)¹⁰, основанного пять лет назад Британским Советом по экономическим и социальным исследованиям (ESRC). Задача этого Центра – анализ особенностей использования грид-технологий в социальных науках, разработка средств их поддержки, подготовка рекомендаций по стратегии развития *e*-социальных наук. NCeSS предоставляет также консультационные услуги, обучение и доступ к онлайн ресурсам – для помощи исследователям из различных областей социального знания в их адаптации к возможностям и реалиям *e*-социальных наук. NCeSS включает координационный хаб (центр) в Манчестерском университете, семь исследовательских групп и 12 небольших исследовательских проектов, ориентированных как на количественные, так и на качественные подходы тех или иных социальных наук. Так, один из этих проектов (Collaboratory for Quantitative *e*-Social Science – *CQeSS*), действующий в Ланкастерском университете, имеет целью содействие квантитативным исследованиям в различных областях социальных наук, включая развитие исследовательского аналитического инструментария, необходимого для обработки больших массивов количественных данных и доступного в режиме «распределенного использования». Речь идет о таких грид-разработках, как SABRE, GROWL, SAKAI¹¹.

Что касается гуманитарных наук, то перспективы использования грид-технологий связываются здесь во многом с расширяющимися возможностями онлайн-доступа к огромным массивам оцифрованных книг, журналов, изобразительных материалов и т.д. Эффективность использования этих ресурсов резко возрастает, когда исследователь может связывать не только различные тексты, но и тексты с изобразительными источниками, историческими картами, записями интервью, видеоматериалами и т.д., размещенными на различных сайтах. Сегодня в Великобритании осуществляется поддержка четырех проектов по развитию виртуальной среды исследований в области гуманитарных наук, девяти регулярных научных семинаров и семи

исследовательских проектов в конкретных областях гуманитарного знания (археология, музейное дело, классические исследования, история Византии)¹². В этих проектах используются различные технологии *e-Science*, включая обработку изображений, текста древних манускриптов, интеграцию национальных баз данных по материалам археологических экспедиций, компьютерное моделирование средневековых битв, использование 3-D сканирования для анализа поверхности объектов музейного хранения, ГИС. Существенно, что работа над этими проектами идет в тесном сотрудничестве гуманитариев и специалистов из области ИКТ. Это междисциплинарное взаимодействие необходимо еще и потому, что массивы данных, анализируемые гуманитариями (в частности, историками), имеют более сложную, «размытую» ("*fuzzy*") структуру в сравнении с данными, полученными в естественнонаучных исследованиях.

Подобные проекты ведутся не только в Великобритании¹³. П. Доорн и П. Алькховен отмечают заметную роль голландской программы САТСН, развивающей возможности доступа к оцифрованным объектам культурного наследия в рамках активного сотрудничества между ИТ-специалистами, институтами культурного наследия и учеными-гуманитариями. Участники десяти проектов этой программы разрабатывают соответствующее программное обеспечение, проводят эксперименты, используя грид-технологии хранения данных в сотрудничестве с голландским *Big Grid Project*. Для усиления возможностей доступа к оцифрованным данным гуманитарного профиля Королевская Академия наук и искусств Голландии (KNAW) и Голландский научный фонд (NWO) создали новый институт – DANS (Data Archiving and Networked Services). DANS функционирует как сетевая организация, с центром, ответственным за организацию инфраструктуры данных¹⁴. В Германии развивается т.н. D-Grid инициатива, в рамках которой создается TextGrid программа, ориентированная на совместную деятельность сетевого сообщества ученых, проводящих редактирование, аннотирование, анализ и публикацию текстов. Дальнейшая цель программы – разработка технологий семантических сетей (Semantic Grid)¹⁵.

На европейском уровне подобные институты и программы работают вместе под эгидой DARIAH (Digital Research Infrastructure for the Arts and Humanities). Цель этого Центра – обеспечить координацию усилий по сохранению культурного наследия в Европе и обеспечению доступа к материалам гуманитарных исследований¹⁶.

Эти разработки дают основания для введения термина *e-Humanities* (менее принятого пока, чем *e-Science*)¹⁷. В работе Й. Кирца рассмотрены десять направлений разработок, развитие которых приведет к повышению роли *e-Humanities* в развитии гуманитарных наук¹⁸:

1) создание больших коллекций оцифрованных материалов (таких, например, как корпус источников);

2) смена акцентов в образовательных программах, ориентированных на овладение навыками работы со стандартными пакетами: предлагается изучать компьютерные методы и технологии, связанные с решением исследовательских задач, характерных для соответствующей области гуманитарного знания;

3) helpdesks: их роль на гуманитарных факультетах должна измениться, они должны содержать больше возможностей для работы студента-гуманитария в новой исследовательской информационной среде, обеспечивать использование специализированного программного обеспечения;

4) установление более тесных связей гуманитарной информатики (*humanities computing*) с «базовой» информатикой;

5) обеспечение более широкого доступа к массивам электронных данных; они должны быть приложенными к соответствующей публикации; архивы машиночитаемых данных должны принимать их на хранение; необходимо разработать стандарты метаданных для упорядочения электронных библиотек, архивов и других ресурсов;

6) создание библиотек программного обеспечения, прошедшего апробацию в исследовательских проектах; они должны быть хорошо документированными;

7) развитие веб-сервисов, которые повысят возможности использования данных и ресурсов в различных исследованиях;

8) разработка схем метаданных становится критическим аспектом развития *e-humanities* в ближайшие годы;

9) разработка основанных на онтологиях авторских программных средств (а также программного обеспечения, созданного в коллективных проектах) – важная составляющая процесса; структура метаданных при этом должна быть согласована с соответствующей структурой, принятой в архивах машиночитаемых данных;

10) развитие удобного интерфейса и средств визуализации – для повышения эффективности исследований мультимедийных данных, демографических и социально-экономических рядов, материалов археологических экспедиций и т.д.

Приоритеты современной исторической информатики...

Как отмечают П. Доорн и П. Алькховен, наиболее высокие оценки экспертов получили пп. 1, 4 и 2, что подчеркивает приоритетность создания коллекций оцифрованных материалов, трансформации образовательных программ в области ИКТ-приложений и более тесной связи «отраслевой» гуманитарной информатики с «базовой» информатикой¹⁹. Проф. Ш. Росс, директор Центра гуманитарной информатики Института информации Университета Глазго также рассматривает в качестве важнейшей задачи развития *e-Humanities* в ближайшем десятилетии – обеспечение хранения оцифрованных данных и онлайн-доступа к ним²⁰.

* * *

Одним из наиболее известных проектов в области грид-технологий является проект EGEE – *Enabling Grids for E-sciencE*, направленный на создание международной грид-инфраструктуры²¹. Проект выполняется консорциумом из 70 институтов в 27 странах, объединенных в региональные гриды. В результате создается самый крупный в мире Грид с суммарной вычислительной мощностью 20000 ЦПУ. Европейское научно-исследовательское сообщество получает в распоряжение общий рынок компьютеринга, обеспечивающий круглосуточный доступ к крупнейшим вычислительным ресурсам. Доступ не будет зависеть от места расположения потребителей и будет основываться на научных сетях Geant и NRNs. Российские участники образовали региональный консорциум РДИГ (Российский Грид для интенсивных операций с данными – Russian Data Intensive GRID, RDIG)²². Наиболее известное приложение проекта – это обработка данных на Грид в глобально распределенном режиме с Большого адронного коллайдера (*Large Hadron Collider – LHC*), который построен в ЦЕРНе (Женева). Эксперименты на коллайдере LHC порождают потоки данных с экстремальными характеристиками: скорость 100 Мбайт/сек, общий объем – несколько петабайтов в год. По прогнозам специалистов, в полном варианте концепция Grid Computing будет реализована в рамках архитектуры Open Grid Services Architecture, которая фактически должна привести к созданию динамических Web-сервисов.

Между тем веб-сервисы в последние годы получили новый импульс к развитию, который привел к формированию концепции Web 2.0. Появление термина *Web 2.0* принято связывать со статьёй Тима О'Рейли «What Is Web 2.0» от 30 сентября 2005 года²³, впервые опубликованной на русском языке в журнале «Компьютерра» (№№37 (609) и 38 (610) от 14 и 19 октября 2005 года, соответственно) и затем выложенной под заголовком «Что такое Веб 2.0» на веб-сайте «Ком-

пьютерра online»²⁴. В этой статье Т. О'Рейли связал появление большого числа сайтов, объединённых некоторыми общими принципами, с общей тенденцией развития интернет-сообщества и назвал это явление Web 2.0 в противовес «старому» Web 1.0. Несмотря на то, что значение этого термина до сих пор является предметом многочисленных споров, те исследователи, которые признают существование Web 2.0, подчеркивают фактор широкого участия "обычных" пользователей глобальной сети в формировании и развитии коллективных Интернет-ресурсов и выделяют несколько основных аспектов и технологий этого явления: Wiki – технология коллективного создания гипертекста; Flickr – социальный фотосервис; YouTube – социальный видеосервис; Google Maps – карты и спутниковые снимки Земли; Blogs – сетевые дневники; Del.icio.us – он-лайн закладки; Netvides – сетевой десктоп и др.

Новые тенденции в деятельности национальных ветвей АНС

Заметное место в тематике зарубежных ассоциаций АНС в последние годы занимают вопросы разработки научно-образовательных ресурсов, использования новых ИКТ в обучении студентов-историков (включая технологии дистанционного обучения, Веб 2.0 и др.). Так, этой тематике была посвящена очередная ежегодная конференция американской ветви АНС, которая прошла в этот раз в виртуальном режиме (апрель 2008 г.). Конференция имела характерное название: *Web 2.0 / History 2.0: Making History Together*. В центре внимания участников были следующие вопросы:

- что включает «исторический сегмент» Web 2.0?
- в чем его отличия от соответствующего сегмента Web 1.0?
- что заставляет нас выходить за пределы возможностей Web 1.0?
- каковы приложения Web 2.0 в исторических исследованиях и образовании?
- каковы «плюсы» и «минусы» Web 2.0 в этих приложениях?

Рассмотрение данных вопросов потребовало обращения к конкретным аспектам сетевого взаимодействия, указанным организаторами конференции: программы «Open source», используемые для интерпретации истории; тэги в исторических приложениях; технологии Wiki, блоги и RSS в исторических приложениях; право собственности на данные; базы данных и исторический анализ; картографирование

истории (документация, основанная на возможностях Google); создание звуковых и видео-передач (подкастов), цифровое видео (например, YouTube), виртуальные миры в исторических приложениях.

Новые тенденции развития исторической информатики обсуждались также в Лондоне, где в ноябре 2008 г. состоялся круглый стол, организованный Британской ассоциацией «History and Computing» (АНС–УК) и посвященный обсуждению рукописи книги «*Historical Methods for the 21st Century*», которая выйдет в свет весной 2009 г. (как в «бумажной» форме, так и в он-лайне). Книга, содержащая обзор и характеристику методического инструментария историка XXI века, написана благодаря совместной инициативе АНС–УК и Национальной архивной службы Великобритании, Института исторических исследований Лондонского университета, а также Британского Академического Центра высшего образования в области истории, классических исследований и археологии. Главная цель книги, как указывают авторы, – преодолеть разрыв между публикациями, посвященными общей характеристике применения ИКТ в истории, и изданиями, ориентированными на те или иные продвинутые компоненты ИКТ, конкретные методы и специальные технологии. Другая задача авторов – сблизить направления использования ИКТ в исторических исследованиях, с одной стороны, и в историческом образовании, – с другой.

Представляет интерес содержание этой книги, которая, судя по всему, станет заметной вехой в развитии исторической информатики. Работа содержит следующие разделы: роль теории; категории исторического анализа; поиск в электронных каталогах Национальных архивов; использование технологий оцифровки; организация данных; использование коллекций данных; технологии Web 2.0; просопография; технологии изучения истории народонаселения; ГИС; исследования локальной истории; аудио-архивы; видео-архивы; text mining; data mining (интеллектуальный поиск и анализ данных); статистический анализ.

Представленная тематика активно обсуждалась на ежегодной британской конференции АНС–УК в ноябре 2007 г., где акцент был сделан на использовании ИКТ в образовании студентов-историков всех трех уровней (бакалавриат, магистратура, PhD). На шести секциях этой конференции обсуждались следующие вопросы: 1) компьютерные технологии в обучении аспирантов-историков; 2) веб-ресурсы национальных архивов: поиск в он-лайновых каталогах; 3) креативное использование ИКТ в обучении историков – бакалавров и магистров; 4) критика источников: «цифровая дипломатика» в учебном плане

в учебном плане историков; 5) аккумуляция и дифференциация профессиональных знаний и навыков при переходе с уровня бакалавра на уровень магистра: роль ИКТ; 6) современное содержание специализации по профилю «History and Computing» в британских университетах.

Одна из задач, стоящих перед специалистами по исторической информатике, работающими в университетах, – включить рассмотрение отмеченных выше тенденций в учебном плане специализации по профилю «Историческая информатика».

Примечания

1. Редакторы *ИНАС* – Дэвид Боденхамер (США) и Пол Элл (Великобритания); редакционный совет журнала в основном не изменился – в него входят ученые из Австралии, Великобритании, Германии, Голландии, Норвегии, России, США, Тайваня.
2. Информационная архитектура – совокупность методов и приёмов структурирования информации (знаний или данных). Чаще всего термин «информационная архитектура» применяется в веб-разработке; в этом контексте информационная архитектура связывается прежде всего с принципами систематизации информации и навигации по ней с целью помочь пользователям более успешно находить и обрабатывать нужные им данные.
3. Boonstra O., Breure L. and Doorn P. Past, Present and Future of Historical Information Science. Amsterdam, NIWI-KNAW. 2004. См. об этой книге: Бородкин Л.И. Историческая информатика в точке бифуркации: движение к Historical Information Science // Круг идей: алгоритмы и технологии исторической информатики. Москва-Барнаул, 2005; Гарскова И.М. Историческая информатика и квантитативная история: преемственность и взаимодействие // Анализ и моделирование социально-исторических процессов. М.: КомКнига, 2006.
4. См., например: <http://www.escience.am>, <http://www.nesc.ac.uk>
5. В отличие, например, от термина e-Government, который получил распространение в русском языке как «электронное правительство».
6. Уже сейчас историками созданы огромные базы данных, например, в области исторической демографии. Эти базы содержат миллионы (а в отдельных случаях – десятки миллионов) записей, основанных на первичных материалах переписей населения.
7. См., например: <http://www.interface.ru/home.asp?artId=10375>
8. Здесь возникает известная аналогия с единой энергетической системой, включающей множество производителей энергоресурсов, поставляющих электроэнергию в общую сеть, к которой подключены миллионы потребителей («пользователей»), получающих доступ к сети, когда они в этом

Приоритеты современной исторической информатики...

нуждаются (и не интересующихся, откуда конкретно пришла к ним энергия).

9. О подходах к созданию структур новой информационной среды в российской науке см.: Паринов С.И. E-Science – онлайнное будущее науки. Приложение к журналу "Информационные технологии". №9. 2007.
10. <http://www.ncess.ac.uk/>
11. <http://www.ncess.ac.uk/research/quantitative/>
12. Грантовая поддержка этих проектов проводится усилиями Arts and Humanities Research Council (AHRC) и UK e-Science Core Programme.
13. Alkhoven P., Doorn P. New Research Perspectives for the Humanities // International Journal of Humanities and Arts Computing. Vol. 1, No. 1, 2007. P. 35–47.
14. Ibid. P. 45. Отметим, что директором-организатором DANS является д-р П. Доорн, Президент АНС.
15. Ibid.
16. Ibid.
17. Ibid. P. 35.
18. Kitz J. E-based Humanities and E-Humanities on a SURF platform. (A report commissioned by SURF-DARE), KRA publishing research, 1 June 2004, Amsterdam.
19. Alkhoven P., Doorn P. Op. cit. P. 38–39.
20. Ibid. P. 42.
21. Проект финансировался Европейской комиссией в рамках 6-й Рамочной программы IST FP6. Подробнее о проекте см.: <http://www.eu-egee.org/>
22. <http://www.egee-rdig.ru>
23. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
24. <http://www.computerra.ru/think/234100/>