

УДК 101.1:316

ОСОБЕННОСТИ ЦИФРОВЫХ ГУМАНИТАРНЫХ НАУК

Погорский Э.К. - кандидат философских наук, стажер, Университетский колледж Лондона (Великобритания).

Публикация подготовлена в рамках проекта «Потенциал молодежных интернет-сообществ: формирование информационно-коммуникативной культуры личности», при поддержке РГНФ (проект 12-33-01293a2).

Аннотация. Статья подготовлена по материалам летней школы Оксфордского университета и семинаров Эдинбургского университета по изучению цифровых гуманитарных наук, а также анализа российских работ в этой области. В работе представлены основополагающие принципы исследований в Digital Humanities, дано описание инструментов и техник, упрощающих исследователю работу с данными, включая такие аспекты как открытые данные, большие данные, сетевая модель данных, базы данных, визуализация, краудсорсинг.

Ключевые слова: цифровые гуманитарные науки, открытые данные, сетевые данные, большие данные, краудсорсинг, визуализация, базы данных.

FEATURES OF DIGITAL HUMANITIES

Pogorskiy E. K.

The paper was prepared within the framework of the project "The Potential of Youth Online Communities: The Formation of Information and Communication Culture of the Person" with support from the Russian Foundation for the Humanities (grant No. 12-33-01293a2).

Abstract. This paper was prepared as a summary of case studies from Digital Humanities Oxford Summer School, research seminars at the University of Edinburgh, and also from analysis of Russian theoretical works in Digital Humanities. This paper outlines basic principles of research in Digital Humanities,

tools and techniques of working with data, including such aspects as open data, network data, big data, databases, visualisation, and crowdsourcing.

Keywords: digital humanities, open data, network data, big data, crowdsourcing, visualisation, databases.

Цифровые гуманитарные науки представляют собой взгляд на традиционно гуманитарные области сквозь призму цифровых технологий. Это направление исследований носит трансдисциплинарный характер и базируется на методологии традиционных гуманитарных наук, включая историю, философию, лингвистику, литературу, искусство, археологию, культуру, музыку, социальные науки, а также методологию и логику исследований естественного профиля, таких как информатика, статистика, компьютерный анализ, машинное обучение. С распространением интернета, увеличением компьютерных мощностей, развитием электронной коммерции, «интернета вещей» и других областей человеческой жизни, сопровождающих «информационное общество» или «общество знания», повышается спрос на исследования в данном направлении, а также расширяются области применения цифровых гуманитарных наук, что, по всей видимости, приведет в ближайшие десятилетия к увеличению числа исследований и объемов средств, выделяемых на развитие Digital Humanities.

Анализ литературы российских и зарубежных авторов показывает схожесть взглядов на содержание и направление развития цифровых гуманитарных наук. Так, Е. Ю. Журавлева отмечает, что работа специалистов в области «цифровых гуманитарных наук» в основном направлена на оцифровку информации, работу с потоками данных, разработку цифровых инструментов (Журавлева, 2011). Тезис о существенной роли информатизации (Луков, Погорский, 2011) в слиянии гуманитарных и естественных наук, а также преодолению их отчуждения друг от друга, выдвигается также Г. В. Можяевой (Можяева, 2013). Кроме того, Журавлева выделяет три способа создания моделей гуманитарной науки в цифровой среде: первый способ заключается в непосредственном переносе практик гуманитарных наук в цифровую среду;

второй способ раскрывается через содержание дисциплин «гуманитарные вычисления» и «гуманитарная информатика»; а третий заключается в формировании системы новых исследовательских практик и технологий (Журавлева, 2011). Важно также обратить внимание на недостатки и опасности цифровых гуманитарных наук, выделяемые профессором Кельнского университета Манфредом Таллером, среди которых: 1) акцент на инфраструктуру для Digital Humanities отбрасывает на периферию обеспечение необходимыми данными и публикацию исследовательского инструментария; 2) IT может поддерживать гуманитарные науки во многих формах и национальных традициях, а не только в узкой области исследования (например, только анализ текстов); 3) надвигающаяся «мобильная революция» несет свои вызовы для новой трансформации гуманитарных наук; 4) под влиянием Digital Humanities происходит не только освоение существующих технологий, но и их развитие, а также появление новых (Таллер, 2012).

Область практического применения Digital Humanities связана с кодированием и декодированием текстовых источников, конструированием географических информационных систем, лексикометрией, оцифровкой культурного, научного и технического наследия, картографией Интернета, добычей данных, архивами аудиозаписей, цифровым и мультимедийным искусством (Дакос, 2011). Стоит отметить, что этот список не окончательный и с развитием технических средств в науке постоянно появляются области исследований, попадающие в поле применения цифровых гуманитарных наук. Область цифровых гуманитарных наук также имеет и свои события, крупнейшими из которых являются: Digital Humanities Summer Institute (Victoria, Canada); Digital Humanities Oxford Summer School (UK), DH Leipzig (Germany), Humanities Intensive Learning and Teaching (College Park, Maryland, USA), SH Switzerland (Bern), SH EDIROM (Paderborn, Germany).

Переходя от описания источников возникновения Digital Humanities к практике выполнения проектов в области цифровых гуманитарных наук, заметим, что еще на этапе планирования проекта следует обратить внимание на

наличие необходимых ресурсов, финансирования, материальной базы, места проведения работ, а также принять в расчет наличие необходимого количества исполнителей, доступность экспертов по теме проекта, возможна ли эффективная коммуникация между исследователями из разных научных областей, приоритеты университета и стэйкхолдеров, наличие контактов с другими исследовательскими группами, сроки проведения, запланированные мероприятия и события по теме проекта. Кроме того, разработка и планирование цифровых проектов зачастую включает проектирование пользовательского опыта (user experience design), включающего контентную стратегию, визуальный дизайн, интерактивный дизайн, информационную архитектуру, функциональность. Управление проектами в Digital Humanities включает в себя достаточно общие принципы, характерные для управления любым проектом и уже неоднократно описанные ранее, а также достаточно специализированные области, которые еще недостаточно раскрыты в русскоязычной научной литературе. В настоящей статье внимание уделяется следующим специализированным областям: открытые данные, большие данные, базы данных, сетевая модель данных, визуализация данных и краудсорсинг (совместная работа над проектом, привлечение общественности к работе над проектом).

Стоит отметить, что некоторые из этих аспектов уже хорошо освоены российской наукой, но часть из них еще не сильно развита или только набирает обороты, — например, принцип открытых данных, которому достаточно большое внимание уделяется в Великобритании. Также заслуживает внимания тот факт, что при предоставлении финансирования проектов в области Digital Humanities от западных фондов, к доказательству и детальности обоснования необходимости применения таких ресурсов, как программное и техническое обеспечение предъявляются весьма высокие требования, сравнимые с требованиями к концепции проекта. Также заявки на гранты в Великобритании зачастую включают два нюанса, которые обычно не учитывают российские фонды, но которые также влияют на планирование и ход выполнения проекта,

— это стратегия управления рисками, раскрывающая потенциальные угрозы выполнения проекта и допустимые изменения вследствие возможных рисков, и коммуникационная стратегия, обозначающая список стэйкхолдеров и способы коммуникаций с ними.

Открытые данные. Открытые исследовательские данные (Open data), — пожалуй, один из главных трендов последних лет в научном мире. Междисциплинарный обмен данными, размещение массивов информации в различных форматах, создание научных сообществ не вокруг научных школ, а вокруг данных — такие цели преследует концепция открытых данных. Эта концепция включает в себя инициативы, способствующие поддержке публикации данных в свободном доступе, и поддерживаемые университетами, научными фондами, правительствами (datashare.is.ed.ac.uk, data.gov.uk, data.gov). Одним из примеров реализации открытых данных может выступать ТЕИ (Text encoding initiative, www.tei-c.org) — международная инициатива, объединяющая вопросы разработки и представления текста в электронной форме, ставящая своей целью разработку декодирующих методов, которые делают текст «читаемым» и пригодным для машинной обработки в гуманитарных, лингвистических и социальных науках.

В качестве иллюстрации внедрения концепции открытых данных можно привести в пример ведущие университеты мира, в которых реализуются мероприятия по продвижению инициатив Open data, проводятся семинары и инструктаж по созданию и подготовке к открытому доступу исследовательских материалов, развертывается соответствующая инфраструктура в виде предоставления ученым и преподавателям, задействованным в научных исследованиях, дискового пространства и серверов для размещения данных. Выполняется сбор информации со всех этапов исследований, включая промежуточные данные, результатов полевых исследований и другие материалы, которые могут служить основой для нового знания.

Большие данные. Большие данные (big data) сегодня являются одним из главных локомотивов развития супер-компьютеров и компьютерных

мощностей. В июне 2013 г. правительство Великобритании опубликовало *«Информационную экономическую стратегию»*, в которой Большие данные занимают одно из пяти главных направлений развития цифровой экономики, наряду с *концепцией умных городов (smart cities)*, *облачными вычислениями (cloud computing)*, *Интернетом вещей (Internet of things)* и *электронной коммерцией (e-commerce)*. Помимо прочего, данный документ расставляет приоритеты при выделении финансирования на научные проекты, способствуя вовлечению представителей научного сообщества в создание инструментов анализа больших данных.

С появлением инструментов для анализа неструктурированной информации перед обществом возникает вопрос о создании открытых, продуктивных, креативных организаций и государственных систем управления. Алекс Пентланд считает, что этот вопрос относится к разряду социальной физики (Pentland, 2014) и очень важен для общества именно сейчас в силу нарастающей глобальной конкуренции, изменений окружающей среды и общего недоверия людей к власти. Социальная физика отвечает на этот вопрос с помощью повсеместно распространяющихся больших данных — знанию, которое делает доступным анализ всех сторон человеческой жизни. С использованием этого типа данных в прогнозировании и моделировании общество может изменить свою социальную систему в гораздо более организованную и динамичную сторону.

Следующие интернет-проекты появившиеся в последние несколько лет, иллюстрируют перспективность направления работы с большими данными: *Zoouniverse.org*, являющийся проектом «гражданской науки» и привлекающий интернет-пользователей к решению исследовательских задач, в середине лета 2014 г. уже привлек более 1,1 млн участников; *Ushahidi* — некоммерческий интернет-сервис, предоставляющий платформу для создания информационных массивов, визуализации данных и создания интерактивных карт. Широко используется как платформа для гражданской журналистики и предоставляет возможность отображения данных на мультимедийных картах в

целях мониторинга выборов, состояния окружающей среды, распространения эпидемий и других процессов; GitHub — предоставляет программную среду для совместной веб-разработки проектов с открытым исходным кодом, поддерживающих научную активность; Myexperiment — сервис по проектированию плана разработки и этапов реализации исследовательских проектов, объединяющий ряд британских университетов и порядка десяти тысяч исследователей по всему миру.

Базы данных. Для описания данного пункта обратимся к электронной библиотеке NathiTrust.org. Проект NathiTrust.org, основанный в 2008 г. на базе университетов Калифорнии, Индианы и Иллинойса, объединяет электронные копии более трех миллионов исследовательских записей (книги, отчеты, публикации и другие документы), доступных для чтения и полнотекстового поиска, и аккумулированных из 60 научных библиотек США. При проектировании инструментов для эффективного использования фондов и ресурсов этой электронной библиотеки организаторы сталкиваются с множеством проблем управления таким массивом данных. И для NathiTrust, как и для любого другого цифрового проекта, оперирующего данными, один из главных вопросов, требующий решения в первую очередь, — вопрос организации баз данных, включающий в себя следующие вызовы: качество данных, организация структуры данных, недостатки описания, преодоление ограничений авторских прав для научных целей, корректное визуальное отображение, создание и поддержка сообщества энтузиастов вокруг проекта.

Переходя к более детальному рассмотрению вопроса об организации баз данных в цифровых гуманитарных проектах, следует отметить, что от исследователя зависит выбор структурирования данных: к примеру, информация может быть организована в виде динамических таблиц (где записи содержат единственное значение, а также доступна численная аналитика с визуализацией в виде графиков и диаграмм), реляционных баз данных (записи в которых содержат множество значений), документо-ориентированных или частично-структурированных баз данных, базы данных формата XML

(категоризирующие элементы текста) или RDF (где записи организованы по принципу отношений между объектами и основаны на принципах семантической сети, например, «Ивану 35 лет» или «Иван знает Сергея»). Также при проектировании архитектуры баз данных в рамках реализации международных проектов следует учитывать возможные проблемы с отображением дат (последовательность день/месяц/год, месяц/день/год или год/месяц/день), историческими датами (8 мая и 9 мая), единообразием словаря в отношении фактов и оценочных суждений (возможно, вероятно, может быть), а также языковые нюансы (американский и британский английский). Кроме того, упорядочить структуру баз данных для исследователей помогают различные инструменты, — например, сервис projectblacklight.org, используемый в электронной библиотеке NathlTrust.

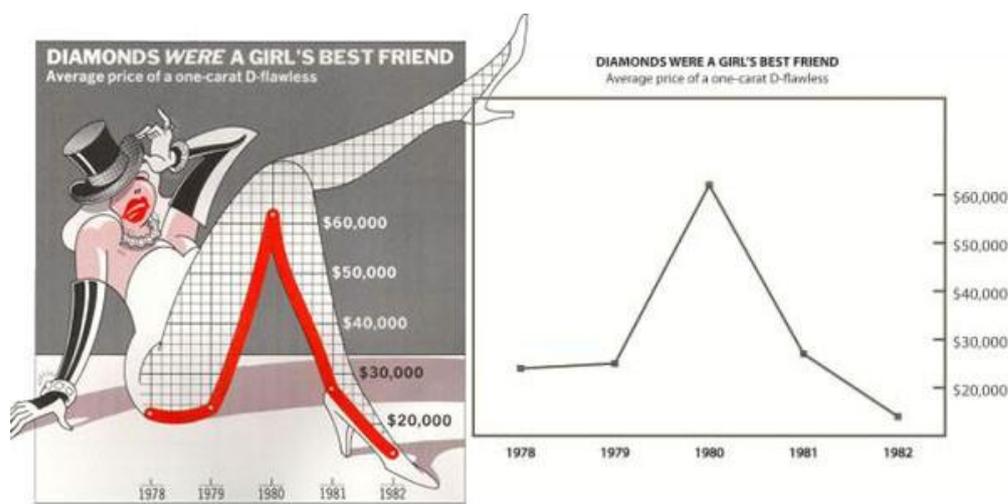
Сетевая модель данных. Для научного мира сетевая модель организации знания отнюдь не нова, и сегодня с развитием коммуникаций она продолжает сохраняться в виде перекликающихся ссылок в научных трудах, научных обменах, совместных проектах, международных конференциях и научных событиях. Развитие трансдисциплинарных исследований, в свою очередь, ставит перед учеными задачу сопоставления различных источников, структур, форматов и объёмов данных. Все эти особенности требуют сопоставления и взаимодействия между исследовательскими материалами таким образом, что они были организованы по сетевому принципу, повторяя организацию научного мира в целом. И если раньше трансдисциплинарность была характерна для масштабных исследовательских проектов, то в цифровых гуманитарных науках трансдисциплинарность является основой и доступна для локальных проектов, реализуемыми, в том числе, отдельными учеными.

Для иллюстрации сетевой модели данных в цифровых гуманитарных науках, представим проект в области Digital Humanities, основывающийся на анализе переписки между людьми одной эпохи из разных городов. Такой проект будет включать различные типы данных и инструменты для работы с ними: хронологические данные являются источником для различных графиков

или анимации последовательности изменений; географические данные могут быть использованы в картографическом представлении; просопографические данные представляются в сетевой визуализации; текстовые материалы и тематический контент подходят для транскрипций и аннотаций, анализа текста, его структуры, упоминания тех или иных объектов, эмоциональности.

Визуализация данных. Визуализация данных позволяет создать наглядное суммарное представление об объекте путем соединения различных типов данных, например, времени и характеристик. Также в цифровых гуманитарных науках визуализация облегчает запоминание, позволяет быстро оценить данные, способствует распространению и популяризации знания.

Процесс создания визуализации включает в себя работу с данными, их фильтрацию и обобщение, расшифровку, перевод в графический вид и отображение (на дисплее, бумаге или каком-либо другом носителе). Например, визуализация текстов в поэзии, отражающая интонацию произведения, позитивную или негативную составляющую. Также заслуживают внимания наблюдения об изменении одного и того же произведения одного автора в разных переводах, то есть насколько в зависимости от точки зрения автора перевода, его картины мира (Погорский, 2012), произведение может приобретать более позитивную или негативную окраску по сравнению с оригиналом. Другой иллюстрацией применения визуализации данных является научный спор, длящийся с 1984 г. и получивший продолжение в публикациях 2010–2011 гг. (Bateman et al., 2010; Few, 2011) и последующих лет — какой из двух графиков, представленных ниже (Holmes, 1984), является лучшим.



Краудсорсинг. Вовлечение интернет-пользователей к участию в научной деятельности и развитие проектов гражданской науки, как показывают результаты проекта Zooniverse.org, является достаточно эффективным и перспективным направлением. Примерами успешных краудсорсинговых проектов в искусстве и гуманитарных науках также могут являться разработки Оксфордского университета “The Great War Archives” и “What's the Score at the Bodleian?”, проект Университетского Колледжа Лондона “Transcribe Bentham”, проект Би-Би-Си “Your Paintings”.

Краудсорсинг научных проектов в области цифровых гуманитарных наук имеет следующие преимущества: возможность генерации дополнительных материалов по теме проекта, анализ большого количества данных, открытость к дискуссиям, образование сообщества вокруг проекта, создание или расширение существующих контактов, возможности для новых исследовательских вопросов. Кроме того, краудсорсинг имеет и другие позитивные стороны, такие как: увеличение количества визитов в музеи и галереи, повышение популярности различных видов искусства и культурного наследия. Сюда же относятся и позитивные эффекты, оказываемые на интернет-пользователей, работающих над краудсорсинговыми проектами, в виде улучшения языковых навыков, повышения самооценки во время безработицы, самоорганизации, борьбы с вредными привычками и отстранением от драматичных жизненных событий как результат погружения в работу (Eccles, 2014).

При привлечении интернет-волонтеров к исследовательскому проекту следует обратить внимание на следующие аспекты: первое и наиболее важное — ясно поставленный вопрос об исследовательской проблеме, далее — четкое понимание возможности добавить ценность проекту и внести свою пользу, наличие сообщества (блоги, форумы, вики) (Dunn, Hedges, 2013). В дополнение можно отметить, что четкое описание задачи, а также наличие детализации способов их выполнения являются ключом к мотивации волонтеров. Важно не отклоняться в задачах далеко от темы проекта, так как зачастую сама область исследования является мотивирующей. Также для повышения мотивации следует продвигать общение и создание горизонтальных связей между волонтерами, обратная связь и благодарность за проделанную работу значительно способствует воодушевлению, так как большинство участников проекта не имеют личной мотивации (Dunn, Hedges, 2012).

Литература

1. Дакос, П. М. (2011) Манифест Digital Humanities [Электронный ресурс] // ThatCamp Paris. URL:<http://tcp.hypotheses.org/501> [архивировано в [WebCite](#)] (дата обращения: 18.10.2014).
2. Журавлева, Е. Ю. (2011) Современные модели развития гуманитарных наук в цифровой среде // Вопросы философии. № 5. С. 91–98.
3. Луков, Вал. А., Погорский, Э. К. (2011) [Информатика и новая информационная реальность: философский взгляд](#) // Знание. Понимание. Умение. № 2. С. 289–293.
4. Таллер, М. (2012) [Дискуссии вокруг Digital Humanities](#) // Историческая информатика. № 1. С. 5–13.
5. Можяева, Г. В. (2013) [Гуманитарные науки в эпоху цифровых технологий: от отраслевой информатики к Digital Humanities](#) // Открытое и дистанционное образование. №. 3(51). С. 10–16.
6. Погорский, Э. К. (2012) [Картина мира](#) // Знание. Понимание. Умение. № 4. С. 322–323.

7. Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A. M., McDine, D., Brooks, C. (2010) Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts // ACM Conference on Human Factors in Computing Systems (CHI 2010). Atlanta, GA, USA. P. 2573–2582.
8. Dunn, S., Hedges, M. (2012) Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research. London : Arts and Humanities Research Council. 56 p.
9. Dunn, S., Hedges, M. (2013) Crowd-sourcing as a Component of Humanities Research Infrastructures // International Journal of Humanities and Arts Computing. Vol. 7. No. 1. P. 147–169.
10. Eccles, K. (2014) Crowdsourcing Projects in the Arts and Humanities: Impact of Tagging on Behaviour : A Paper Presented at ‘Digital Humanities at Oxford Summer School’, University of Oxford, July 2014.
11. Few, S. (2011) The Chartjunk Debate : A Close Examination of Recent Findings [Электронный ресурс] // Visual Business Intelligence Newsletter.
URL:http://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf[архивировано в [WebCite](#)] (дата обращения: 18.10.2014).
12. Holmes, N. (1984) Designer’s Guide to Creating Charts and Diagrams. New York : Watson-Guptill Publications. 192 p.
13. Pentland, A. (2014) Social Physics: How Good Ideas Spread : The Lessons From a New Science. New York : The Penguin Press. x, 300 p.
14. _____
15. REFERENCES
16. Dakos, P. M. (2011) Manifest Digital Humanities [The Digital Humanities Manifesto]. ThatCamp Paris[online] Available at: <http://tcp.hypotheses.org/501> [archived in [WebCite](#)] (accessed 18.10.2014). (In Russ.).

17. Zhuravleva, E. Yu. (2011) Sovremennyye modeli razvitiia gumanitarnykh nauk v tsifrovoi srede [Contemporary Patterns of the Humanities' Development in the Digital Environment]. Voprosy filosofii, no. 5, pp. 91–98. (In Russ.).
18. Lukov, Val. A. and Pogorskiy, E. K. (2011) Informatika i novaia informatsionnaia real'nost': filosofskii vzgliad [[Informatics and New Information Reality: A Philosophical View](#)]. Znanie. Ponimanie. Umenie, no. 2, pp. 289–293. (In Russ.).
19. Thaller, M. (2012) Diskussii vokrug Digital Humanities [Controversies around the Digital Humanities]. Istoricheskaia informatika, no. 1, pp. 5–13. (In Russ.).
20. Mozhaeva, G. V. (2013) Gumanitarnye nauki v epokhu tsifrovyykh tekhnologii: ot otraslevoi informatiki k Digital Humanities [[The Humanities in the Digital Age: From Branch Informatics to Digital Humanities](#)]. Otkrytoe i distantsionnoe obrazovanie, no. 3(51), pp. 10–16. (In Russ.).
21. Pogorskiy, E. K. (2012) Kartina mira [[World View](#)]. Znanie. Ponimanie. Umenie, no. 4, pp. 322–323.
22. Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A. M., McDine, D. and Brooks, C. (2010) Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In: ACM Conference on Human Factors in Computing Systems (CHI 2010). Atlanta, GA, USA. Pp. 2573–2582.
23. Dunn, S. and Hedges, M. (2012) Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research. London, Arts and Humanities Research Council. 56 p.
24. Dunn, S. and Hedges, M. (2013) Crowd-sourcing as a Component of Humanities Research Infrastructures. International Journal of Humanities and Arts Computing, vol. 7, no. 1, pp. 147–169.

25. Eccles, K. (2014) Crowdsourcing Projects in the Arts and Humanities: Impact of Tagging on Behaviour : A Paper Presented at 'Digital Humanities at Oxford Summer School', University of Oxford, July 2014.
26. Few, S. (2011) The Chartjunk Debate : A Close Examination of Recent Findings. Visual Business Intelligence Newsletter [online] Available at:http://www.perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf[archived in [WebCite](#)] (accessed 18.10.2014).
27. Holmes, N. (1984) Designer's Guide to Creating Charts and Diagrams. New York, Watson-Guptill Publications. 192 p.
28. Pentland, A. (2014) Social Physics : How Good Ideas Spread : The Lessons from a New Science. New York, The Penguin Press. x, 300 p.
29. Библиограф. описание: Погорский Э. К. Особенности цифровых гуманитарных наук [Электронный ресурс] // Информационный гуманитарный портал «Знание. Понимание. Умение». 2014. № 5 (сентябрь — октябрь). URL: http://www.zpu-journal.ru/e-zpu/2014/5/Pogorskiy_Digital-Humanities/[архивировано в [WebCite](#)] (дата обращения: дд.мм.гггг).