

История гуманитарного компьютинга

Сьюзан Хоки

Перевод с английского
Валерии Гавриленко
по изданию: © *Hockey S.*
*The History of Humanities
Computing // A Companion to
Digital Humanities / S. Schreiber,
R. Siemens, J. Unsworth*
(eds). Oxford: Wiley-Blackwell,
2004. Публикуется с любезного
разрешения автора.

Эмерит-профессор библиотечного дела
и информационной науки департамента ин-
формационных исследований Университет-
ского колледжа Лондона (UCL). Адрес:
Gower Street, London WC1E 6BT, United
Kingdom. E-mail: s.hockey@ucl.ac.uk.

Ключевые слова: автоматическая обработ-
ка; методология; дигитализация; кодировка;
данные.

Автор прослеживает историю компьютери-
зации и дигитализации гуманитарного зна-
ния за период от 1950 до 1990 года. Она

Вступление

000000000142647560000000
400000008866465300000007
1000000012428282000000034
4600000059525470000000071
3900000054299600000000027
3200000091294000600000068
5100000002970001900000013
9000000083000035600000031
9400000084000224800000004
1100000090003887900000032
1800000000094772100000068
890000000446071200000023
590000000685396400000026
0000000084882709000000007
0000000512796180000000000
0320680769372341732265313
6215872551172683854374584
2645840776709164616209889
2369911885911949897093920
9934164721821464242591813
8499297080485525401415973
5295097268267858345538994
6511133226416356413631418
4749086794856867561877792
6244154088781893801569853
9227030559555683310407958
9277706417688992359344184

Изучение истории любой меж-
дисциплинарной академической
сферы ставит ряд базовых вопро-
сов. Каковы пределы этой сфе-
ры? Пересекается ли она с други-
ми смежными областями, которые,
может быть, повлияли на ее дея-
тельность? Каково ее воздействие
на другие, возможно, более тра-
диционные дисциплины? Отда-
ет ли должным образом отчет о ее
деятельности простое хронологи-
ческое изложение? Возможны ли
отступления от него, которые мог-
ли бы навести нас на прежде не-
известные возможности? Каждый
из этих вопросов мог бы стать ос-
нованием для отдельного эссе,
но в рамках наличного здесь про-

начинает с новаторской деятельности патера Роберто Буза для приведения трудов Фомы Аквинского в автоматически читабельную форму и заканчивает Инициативой по кодированию текстов (*TEI*). Статья охватывает огромное разнообразие методов и большой круг вопросов, характеризующийся явной склонностью к междисциплинарности. По самой своей природе дигитализация гуманитаристики была призвана объединить «две культуры» (Чарльз Сноу), чтобы принести строгость и систематические однозначные процедурные методологии, характерные для «точной» науки, для решения проблем в области наук гуманитарных, которые до сих пор, как правило, продвигались с изрядной долей интуитивности и случайности. Вершиной достижений за рассматриваемый период стал *TEI*, повлиявший на профессиональное сообщество в целом.

Дигитализация гуманитаристики может внести существенный вклад в сохранение культурного наследия в интернете, что важно не только для академических пользователей, но и для всех учащихся и широкой общественности в целом. Наши инструменты и методы должны помочь исследованию посредством оцифровки и кодирования процессов, где предстоит найти, как сократить расходы на создание данных без потери научной ценности или функциональности. На протяжении всей своей истории дигитализация гуманитарного знания показала здоровый аппетит для фантазии и инноваций, продолжая поддерживать высокие академические стандарты. Теперь, когда интернет является доминирующей чертой повседневной жизни, для дигитализации гуманитаристики сложились возможности, о которых прежде нельзя было и мечтать.

странства и контекста выбранный подход призван представить развитие применения компьютерных систем в гуманитарных науках в хронологическом порядке. В пределах этой задачи я уделяю внимание лишь ключевым моментам — серьезным интеллектуальным прорывам или тем случаям, когда проделанная в гуманитарном компьютеринге работа осваивалась, развивалась или существенно использовалась в других дисциплинах.

Здесь не место обсуждать определение того, что имеется в виду под гуманитарным компьютерингом. Достаточно сказать, что нас интересует применение компьютерной обработки данных в исследованиях и преподавании в рамках дисциплин, расплывчато определенных как «гуманитарные науки» (*humanities*), а в британском английском — «искусства» (*arts*). Программные приложения, работающие с текстовыми источниками, заняли ведущее место в развитии гуманитарного компьютеринга, о чем свидетельствуют ключевые публикации в этой сфере. Поэтому неизбежно, что данное эссе фокусируется на этой области. Здесь также не будут предприниматься попытки дать определение «междисциплинарности», но ясно, что по самой своей природе гуманитарный компьютеринг должен был охватить «обе

культуры», ввести свойственные точным наукам жесткие и систематичные однозначные процедурные методологии, чтобы обратиться к тем проблемам в гуманитарии, которые до сих пор решались инстинктивно.

Истоки: 1949-й — начало 1970-х

В отличие от многих других междисциплинарных экспериментов, и дата, и обстоятельства рождения гуманитарного компьютеринга тайны не представляют. В 1949 году итальянский иезуитский священник отец Роберто Буза начал то, что даже на сегодняшний день представляется задачей монументальной: создание *index verborum*, указателя всех слов корпуса текстов Фомы Аквинского и других средневековых философов, насчитывающего в целом до одиннадцати миллионов слов средневековой латыни. Буза решил, что какая-то машина могла бы ему помочь, и, услышав о компьютерах, отправился за поддержкой к Томасу Дж. Ватсону в IBM в США¹. Помощь должна была скоро подоспеть, и Буза взялся за работу. Для проекта все тексты были постепенно перенесены на перфокарты и была разработана программа конкордансов. Целью было создание печатных томов, первый из которых был опубликован в 1974 году².

Чисто механическая программа конкордансов, где слова располагаются по алфавиту согласно их графическим формам (последовательностям букв), могла бы дать результат за гораздо более короткое время, однако Буза это бы не удовлетворило. Он хотел создать «лемматизированный» индекс, где слова расположены по словарным гнездам, а не просто по их формам. Его команда попыталась написать программу, чтобы справиться с этой задачей, в результате чего лемматизация всех одиннадцати миллионов словоупотреблений была полуавтоматически осуществлена людьми, работающими со словоформами, с которыми одна программа не могла справиться. Буза задал для своей работы очень высокие стандарты. Его тома напечатаны изящно; ни на каком этапе работы он не пошел на уступки ради того, чтобы ее ускорить. Он продолжает оказывать глубокое влияние на гуманитарный компьютеринг своим видением и воображением.

1. *Busa R. The Annals of Humanities Computing: The Index Thomisticus// Computers and the Humanities. October 1980. Vol. 14. № 2.*
2. *Idem. Index Thomisticus: Sancti Thomæ Aquinatis operum indices et concordantiæ: In 49 vols. Stuttgart: Frommann-Holzboog, 1974–.*

ем, не уступающим многим специалистам нынешнего поколения, воспитанного на интернете. В 1992 году вышел CD-ROM с корпусом текстов Аквината, содержащим гипертекстовые функции (*cum hypertextibus*)³ и сопровождаемым инструкцией на латыни, английском и итальянском языках. Отец Буза стал первым лауреатом премии имени Буза, учрежденной в знак признания выдающихся достижений в применении информационных технологий в гуманитарных исследованиях. В 1998 году на своей лекции, приуроченной к вручению премии в Дебрецене, он размышлял о потенциале Всемирной паутины WWW производить мультимедийные научные данные, сопровождаемые изоциренным аналитическим инструментарием⁴.

К 1960-м годам преимущества работы с конкордансами стали очевидны и для других исследователей. Серия из четырех статей Долорес Бертон в журнале «Компьютеры и гуманитарные науки» (*Computers and the Humanities*) в 1981–1982 годах попыталась объединить их, начав с дискуссии 1950-х⁵. Некоторые исследователи были независимыми учеными, чьи интересы концентрировались на одной группе текстов или авторов. В Великобритании Рой Уисби построил ряд конкордансов к ранним средневерхненемецким текстам⁶. В США конкордансы Стефена Перриша к стихотворениям Мэтью Арнольда и Уильяма Батлера Йейтса предварили серию конкордансов, опубликованных издательством Корнельского университета⁷. В этот же период возникли вычислительные центры при некоторых крупных языковых академиях Европы — главным образом для помощи в составлении словарей. Среди примеров — *Trésor de la Langue*

3. Thomae Aquinatis Opera Omnia Cum Hypertextibus in CD-ROM/R. Busa (ed.). Milano, 1992.
4. Busa R. Picture a Man...//Literary and Linguistic Computing. 1999. Vol. 14. № 1.
5. Burton D. Automated Concordances and Word Indexes: The Fifties//Computers and the Humanities. 1981. Vol. 15. № 1; *Idem*. Automated Concordances and Word Indexes: The Early Sixties and the Early Centers//Computers and the Humanities. 1981. Vol. 15. № 2; *Idem*. Automated Concordances and Word Indexes: The Process, the Programs, and the Products//Computers and the Humanities. 1981. Vol. 15. № 3; *Idem*. Automated Concordances and Word Indexes: Machine Decisions and Editorial Revisions//Computers and the Humanities. 1982. Vol. 16. № 4.
6. Wisbey R. The Analysis of Middle High German Texts by Computer: Some Lexicographical Aspects//Transactions of the Philological Society. 1963. Vol. 62. № 1.
7. Parrish S. Problems in the Making of Computer Concordances//Studies in Bibliography. 1962. Vol. 15. № 1.

*Française*⁸, основанный в Нанси для создания архива французских художественных текстов, и Институт голландской лексикологии в Лейдене⁹.

Хотя большая часть работы в это время была сосредоточена на построении конкордансов как самоцели, применение новых инструментов начало жить собственной жизнью. Использование количественных подходов в изучении стиля и авторства предвосхищает компьютерную обработку данных. Например, еще в 1851 году британский математик и логик Август де Морган предложил количественное изучение словаря как способ исследования авторства посланий Павла¹⁰, а Томас Корвин Менденхолл в конце XIX века описал свою счетную машину, с помощью которой две женщины подсчитывали количество слов, состоящих из двух, трех и т. д. букв у Шекспира, Марло, Бэкона и многих других авторов, пытаясь установить, кто написал Шекспира¹¹. Однако появление компьютеров сделало возможным регистрировать частотности слов в гораздо больших количествах и точнее, чем это может человек. В 1963 году шотландский священник Эндрю Мортон опубликовал в одной британской газете статью, где утверждал, что по данным компьютера святой Павел написал только четыре послания. Мортон обосновывал свое утверждение подсчетом общеупотребительных слов в греческом тексте с учетом некоторой элементарной статистики. Он продолжил изучение разных греческих текстов, выпуская статьи и книги, посвященные исследованию частотностей общеупотребительных слов (обычно частиц), а также длины предложений, хотя ему можно возразить, что пунктуация, идентифицирующая предложения, была внесена в греческие тексты современными издателями¹².

8. *Gorcy G. L'informatique et la mise en œuvre du trésor de la langue française (TLF), dictionnaire de la langue du 19e et du 20e siècle (1789–1960)*// *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries* / A. Cappelli, A. Zampolli (eds). Pisa, 1984.
9. *De Tollenaere F. The Problem of the Context in Computer-aided Lexicography*// *The Computer and Literary Studies* / A. J. Aitken, R. W. Bailey, N. Hamilton-Smith (eds). Edinburgh: Edinburgh University Press, 1973.
10. *Lord R. Studies in the History of Probability and Statistics: viii. de Morgan and the Statistical Study of Literary Style*// *Biometrika*. 1958. Vol. 45. № 1–2. P. 282.
11. *Mendenhall T. C. A Mechanical Solution of a Literary Problem*// *The Popular Science Monthly*. 1901. Vol. 60. № 7.
12. *Morton A. The Authorship of the Pauline Epistles: A Scientific Solution*. Saskatoon, 1965; *Morton A., Winspear A. It's Greek to the Computer*. Montreal: Harvest House, 1971.

Считается, что в изучении спорного авторства компьютер был впервые применен Альваром Эллегардом к так называемым письмам Юниуса, анонима второй половины XVIII века. Его опубликованное в 1962 году исследование использовало компьютер не для подсчета слов, а для машинных расчетов, позволивших Эллегарду получить общую картину словаря исходя из подсчетов вручную¹³. Вероятно, наиболее авторитетным является основанное на применении компьютера исследование авторства, также выполненное в начале 1960-х годов. Это работа Мостеллера и Уоллеса, посвященная «Запискам Федералиста», сборнику в поддержку ратификации Конституции США, где исследователи пытались установить авторство двадцати спорных статей¹⁴. С таким объемом текстов, написанных обоими кандидатами на авторство на одну ту же тему, исследование имело дело с идеальной ситуацией для компаративной работы. Мостеллер и Уоллес были сначала заинтересованы в статистических методах, которые применяли, но они сумели продемонстрировать, что автором обсуждаемых записок был, очевидно, Медисон. Их выводы в целом приняты, причем до такой степени, что «Федералист» используется в качестве теста новых методов установления авторства¹⁵.

В это время много обсуждались границы технологии. Анализируемыми данными были либо тексты, либо числа. Они тщательно вносились от руки на перфокарты, при этом каждая карта содержала до восьмидесяти знаков или одну линию текста (только заглавными буквами), или на бумажные перфоленты, где строчные буквы, может, и были допустимы, но человек никоим образом не мог их прочитать. У отца Буза есть свои рассказы о вагонах перфокарт, транспортируемых по Италии из одного центра в другой. Вся обработка данных выполнялась поэтапно, так что пользователь никак не мог увидеть результаты раньше вывода на печать, когда работа была уже завершена. Представление данных на основе набора символов вскоре было признано

13. *Ellegerd A.* A Statistical Method for Determining Authorship: The Junius Letters 1769–1772. Gothenburg: Elanders Boktryckeri Aktiebolg, 1962.
14. *Mosteller F., Wallace D.* Inference and Disputed Authorship: The Federalist. Reading, MA: Addison-Wesley, 1964.
15. *Holmes D., Forsyth R.* The Federalist Revisited: New Directions in Authorship Attribution // *Literary and Linguistic Computing*. 1995. Vol. 10. № 2; *Tweedie F.J., Singh S., Holmes D.I.* Neural Network Applications in Stylometry: The Federalist Papers // *Computers and the Humanities*. 1996. Vol. 30. № 1.

существенной проблемой, которую начали решать только сейчас, с появлением *Unicode*, хотя и не для любых гуманитарных текстов. Были разработаны разные методы изображения прописных и строчных букв на перфокартах, чаще всего введением звездочки или похожего символа перед заглавной буквой. Акценты и другие нестандартные символы подвергались схожей процедуре, а нелатинские алфавиты были целиком транслитерированы.

Самые объемные массивы данных записывались на магнитную ленту, а это можно делать только сериально. Проматывание всей кассеты из конца в конец занимало около четырех минут, так что была написана программа для минимизации количества прокруток кассеты. Произвольный, как на дисках, доступ к данным был невозможен. Таким образом, их нужно было записывать сериями. Это было не так проблематично для текстовых данных, однако для исторических текстов это могло означать упрощение данных, которые представляли несколько аспектов одного объекта (формируя несколько таблиц в технологии реляционной базы данных), сведение их к одному линейному потоку. Этого оказалось достаточно, чтобы отпугнуть историков от участия в проектах, основанных на применении компьютера.

Проблемы репрезентации коснулись далеко не только отдельных символов. Программам построения конкордансов и поисковым программам нужно было идентифицировать цитаты по их расположению в тексте. Методы, используемые поисковыми системами для обычных документов, не подходили, так как обычно они предполагали структуры документов, подобные статьям в журналах, и не могли справиться со структурами, свойственными в поэзии, драме или в тех рукописях, в которых важно размещение по строкам. Предложены были разные методы определения структур документов, но наиболее изощренным оказался метод, используемый программой построения конкордансов *COSOA*¹⁶. Смоделированный по формату, созданному Полом Брэтли для Архива древних шотландских текстов¹⁷, *COSOA* дает возможность пользователю задавать спецификацию структуры документа, свойственной определенному набору документов. Он также позволяет размечать параллельные структуры,

16. *Russell D. COSOA — A Word Count and Concordance Generator for Atlas.* Chilton, UK: Atlas Computer Laboratory, 1967.

17. *Hamilton-Smith N. A Versatile Concordance Program for a Textual Archive//The Computer in Literary and Linguistic Research/R. Wisbey (ed.).* Cambridge, MA: Cambridge University Press, 1971.

делая возможным, например, кодирование системы цитирования для печатной версии параллельно с системой цитирования для рукописного источника текста. *SOSOА* не занимает много памяти, но, кажется, не очень читабелен для человека.

Другая широко применяемая схема для цитаций была более зависима от формата перфокарт. В этой схеме, часто называемой фиксированным форматом, каждая строка начиналась с закодированной последовательности символов с информацией о цитате. Каждая единица в пределах цитаты располагалась в специальных колонках параллельно строке, например: название — в колонках 1–3, номер стиха — в 5–6, а номер строки — в 7–9. Ввод этой информации ускорялся функциями на перфораторе, но в то же время она занимала больше места в компьютерном файле.

Наследие этих схем цитирования до сих пор может быть обнаружено в электронных текстах, созданных некоторое время тому назад. Программа *SOSOА*, в частности, имела решающее воздействие: другие схемы были выведены из нее. *SOSOА* не может легко обращаться с разметкой небольших признаков объектов внутри контента, таких как имена, даты и аббревиатуры, но ее способность работать с параллельными структурами обгоняет в этом почти все современные схемы разметки.

В этот же период впервые сложились возможности кооперации через гуманитарный компьютеринг для обмена идеями и совместного решения проблем. В 1964 году *IBM* организовал конференцию в Йорктаун-Хайтс. Последующая публикация «Обработка литературных данных: материалы конференции», изданная Джесс Бессинджер и Стефеном Перришем¹⁸, читается почти как текст 1980-х годов, если абстрагироваться от перфокарт. В докладах выносят на обсуждение сложные вопросы кодирования рукописных текстов и автоматической сортировки конкордансов, где и варианты написания, и недостаток лемматизации отмечены как серьезная помеха.

Насколько можно судить, конференция в Йорктаун-Хайтс была уникальным и изолированным событием. Только через шесть лет, в марте 1970 года, в Кембриджском университете Роем Уисби и Майклом Фаррингдоном была организована конференция, задуманная как первая в серии регулярных конференций по литературному и лингвистическому компьютерин-

18. *Bessinger J., Parrish S. Literary Data Processing Conference Proceedings. White Plains, NY: IBM, 1965.*

гу и предшественница того, что станет встречами Ассоциации по литературному и лингвистическому компьютерингу / Ассоциации по компьютерам и гуманитарным наукам (ALLC/ACH). Это было поистине международное событие с достойными представителями с обеих сторон Атлантики и из Австралии. Материалы, тщательно подготовленные Уисби¹⁹, установили стандарт для последующих публикаций. Ретроспективно внимание на себя обращает акцент на вводе, выводе и программировании, а также на лексикографии, текстовом редактировании, преподавании языков и стилистике. Даже в это время вполне осознавалась необходимость архивирования и поддержки электронных текстов.

Другой показатель нарождающейся отрасли знаний — основание нового журнала. «Компьютеры и гуманитарные науки» начал выходить под редакцией Йозефа Рабена в 1966 году. Рабен с особой энергией пестовал новый журнал, и на протяжении нескольких лет, по крайней мере до регулярных конференций и образующихся после них ассоциаций, журнал был главным средством распространения информации о гуманитарном компьютеринге. Рабен осознавал необходимость быть в курсе происходящего, и редакция журнала постепенно стала первым местом, куда обращались желающие запустить какой-либо проект. Другие неофициальные бюллетени и журналы также помогали отдельным сообществам, в частности *Calculi* по компьютерам и классической филологии, издававшийся Стефеном Уэйтом.

На 1960-е приходится и создание центров по использованию компьютеров в гуманитарных науках. В 1963 году Уисби основал Центр литературного и лингвистического компьютеринга в Кембридже в качестве поддержки своей работы с ранними средневерхненемецкими текстами. В Тюбингене Вильгельм Отт собрал группу, начавшую разработку серии программ по текстовому анализу, в частности, для создания критических изданий. Программные модули *TuStep* до сих пор в ходу, они задают очень высокие стандарты научной работы во всех ее фазах — от ввода данных и сличения текстов до производства всего комплекса печатных томов.

Для этого раннего периода характерно, что технология часто стесняла работу, если под технологией понимать наборы символов, инструменты ввода/вывода и медленный оборот систем пакетной обработки данных. Впрочем, исследователи находили,

19. The Computer in Literary and Linguistic Research / R. Wisbey (ed.). Cambridge, MA: Cambridge University Press, 1971.

пусть и неуклюжие, способы решения некоторых из этих проблем. Что более характерно, выявленные ими ключевые проблемы все еще остаются, например необходимость справляться с вариативным написанием, различными вариантами рукописей и лемматизацией.

Консолидация: 1970-е — середина 1980-х

Если охарактеризовать этот период одним словом, то им будет «консолидация». Все больше людей применяли методологии, разработанные в течение предшествующего периода. Создавалось больше электронных текстов, стартовало все больше проектов, использующих те же приложения. Информация о том, что возможно, постепенно распространилась по обычным каналам в академической среде, все больше людей имели дело с компьютерами в повседневной жизни, начиная задумываться о том, как использовать компьютеры для своих исследований и преподавания.

Распространению информации способствовали «Компьютеры и гуманитарные науки» и постоянные конференции. Симпозиум в Кембридже в 1970 году положил начало серии конференций, проводившихся в Великобритании один раз в два года и ставших главным событием в области гуманитарного компьютеринга. На встречах в Эдинбурге (1972), Кардиффе (1974), Оксфорде (1976), Бирмингеме (1978) и Кембридже (1980) были представлены очень высококачественные доклады. Ассоциация по литературному и лингвистическому компьютерингу (*ALLC*) образовалась в результате встречи в Королевском колледже Лондона в 1973 году. С самого начала она трижды в год выпускала свой бюллетень, а также начала организовывать ежегодные встречи с приглашенными специалистами и к 1986 году основала журнал «Литературный и лингвистический компьютеринг» (*Literary and Linguistic Computing*). К середине 1970-х в Северной Америке началась другая серия конференций под названием «Международная конференция по компьютерингу в гуманитаристике» (*ICSH*). Конференции проводились по нечетным годам, чтобы чередоваться со встречами в Великобритании. Британские конференции и ежегодные встречи *ALLC* начали постепенно объединяться. Они продолжали заниматься литературной и лингвистической обработкой данных с некоторым акцентом на «лингвистическом», предложив форум для растущего числа европейских исследователей в области, ставшей известной как корпусная лингвистика. *ICSH* привлекала доклады с более широким тема-

тическим спектром, к примеру о применении компьютеров в преподавании письма, о музыке, искусстве и археологии. Ассоциация по компьютерам и гуманитарным наукам (АСН) возникла из этой конференции и была основана в 1978 году.

Требования гуманитарного компьютеринга начали признаваться и в академических вычислительных центрах. Еще во времена обработки данных на универсальных ЭВМ нужно было регистрироваться, чтобы воспользоваться любыми вычислительными средствами. А такая регистрация позволяла персоналу академических вычислительных центров определять потребности пользователей и планировать установку стандартного программного обеспечения, которое могло бы использоваться разными людьми. Вторая версия программы построения конкордансов *СОСОА* в Британии была разработана именно с целью установки на разных мейнфреймах²⁰. Программа была распространена по разным вычислительным центрам в середине 1970-х, и многие из них назначали одного человека, ответственного за поддержку. Неудовлетворенность ее пользовательским интерфейсом вместе с прекращением поддержки со стороны лаборатории «Атлас», где программа была написана, привели к тому, что финансовые структуры спонсировали разработку новой программы в Оксфордском университете. Названная Оксфордской программой конкордансов (*ОСР*), она была готова к реализации в 1982 году и вызвала интерес во всем мире у пользователей разных стран²¹. В это же время появилось и другое пакетное или комплексное программное обеспечение, что значительно снизило стоимость проектов по части средств обеспечения программных разработок.

Потребность в том, чтобы избежать дублирования усилий, тоже привела к объединению в области текстового архивирования и хранения. С появлением пакетного программного обеспечения и в связи с уже неактуальным программированием в больших объемах, в любом проекте подготовка электронных текстов начала занимать большую часть времени. Главным стимулом к созданию Оксфордского текстового архива (*ОТА*) в 1976 году была потребность удостовериться в том, что законченный исследователем текст не пропадет. *ОТА* предпринял

20. *Berry-Rogghe G., Crawford T.* Developing a Machine-independent Concordance Program for a Variety of Languages//The Computer and Literary Studies.

21. *Hockey S., Marriott I.* The Oxford Concordance Project (ОСР). Part 1–3// ALLC Bulletin 7. 1979; *Idem.* The Oxford Concordance Project (ОСР). Part 4// ALLC Bulletin 8. 1980.

сохранение электронных текстов и, при условии разрешения депозитора, а также с соответствующими разрешениями относительно авторских прав, делал эти тексты доступными для любого желающего воспользоваться ими в академических целях. Это было началом цифровой библиотеки, хотя изначально никто ее так не называл, и ее персонал должен был еще выработать свои методы описания и документации материала²². Массив недокументированных данных указал на необходимость выработки общепринятых процедур описания электронных текстов.

Подход ОТА состоял в том, чтобы предложить сервис по сохранению любых внесенных данных. Ему это удавалось достаточно длительное время и с помощью очень небольшого бюджета, однако сервис не поддерживал создание некоторых текстов. Группы ученых в некоторых дисциплинах скооперировались для создания архива текстов с целью использования его в качестве источника исследований. Из них особого внимания достоин *Thesaurus Linguae Graecae (TLG)*, созданный в Калифорнийском университете в Ирвайне и руководимый Теодором Бруннером на протяжении многих лет. Бруннер собрал миллионы долларов на поддержку создания «базы данных» древнегреческих текстов, включавшей всех авторов от Гомера до приблизительно 600 г. н. э., — около 70 миллионов слов²³. Дополнительная коллекция классических латинских текстов была создана позднее Институтом гуманитарных наук Паккарда, что вместе с *TLG* дало классическим филологам такой исследовательский ресурс, которому в других дисциплинах долгие годы не было равных. Только у специалистов по древнеанглийской литературе был доступ к похожему полному, но меньшему корпусу древнеанглийских текстов, составленному для древнеанглийского словаря²⁴.

В этот период было основано еще больше центров по обработке данных в гуманитарных науках. Некоторые из них, как, например, Норвежский вычислительный центр для гуманитарных наук (сегодня *HIT*) в Бергене, при существенной поддержке го-

22. Proud J. The Oxford Text Archive. L.: British Library R&D Report, 1989.

23. Brunner T. Classics and the Computer: The History of a Relationship// Accessing Antiquity: The Computerization of Classical Studies/J. Solomon (ed.). Tucson; L.: University of Arizona Press, 1993.

24. Healey A. The Corpus of the Dictionary of Old English: Its Delimitation, Compilation and Application// Fifth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary (September 18–19, 1989, St. Catherine's College, Oxford). Waterloo, ON: UW Centre for the New OED, 1989.

сударства внедряли широкий спектр прикладных технологий и проектов. Другие, как Центр компьютерного анализа текстов (ССАТ) в Университете Пенсильвании, сосредоточивались более узко на интересах научных и университетских кругов, которые способствовали формированию этих центров. Подобные организации стали возникать по всему миру, а ученые этих институций в целом пользовались значительной поддержкой.

В этот период были введены университетские курсы, посвященные разным аспектам гуманитарного компьютеринга. Одни преподавались персоналом академических вычислительных центров: они концентрировались в основном на технических аспектах применения программного обеспечения. Другие предусматривали широкий обзор областей приложения. Если курсы читались учеными, то их тематика обычно соответствовала интересам авторов, которые затем привлекали к этим областям и своих студентов. Помимо этого, не прекращалась дискуссия, должны ли студенты изучать программирование. Кто-то считал, что оно заняло место латыни в качестве «ментальной дисциплины»²⁵. Другие полагали, что оно сложно и отнимает слишком много времени от основной работы в гуманитарных науках. Некоторое время язык обработки строк Снобол был в моде, так как для студентов-гуманитариев он был проще остальных языков программирования, среди которых главным был Фортран.

Были также подвижки в развитии инструментов обработки, в основном в связи с переходом от ленты к хранению на дисках. Файлы больше не нужно было искать в порядке их следования. Какое-то время в ходу были разные способы упорядочивания материала в базах данных, и некоторые из них — весьма эффективные для гуманитарных текстов²⁶, но постепенно возобладала реляционная модель. При вводе в эксплуатацию на мейнфреймах она продемонстрировала лучшую структуру, в которой могли работать историки и другие имеющие дело с материалом, взятым из источников (не с самими источниками). Однако реляционные методы все еще представляли трудности в репрезентации той информации, которую нужно было вместиать в таблицы. В 1970-е годы были изобретены по меньшей мере два аппаратных устройства для помощи в поиске. Одно было установлено на компьютере Дэ-

25. *Hockey S. Workshop on Teaching Computers and the Humanities Courses // Literary and Linguistic Computing. 1986. Vol. 1. № 4. P. 228–229.*

26. *Burnard L. Principles of Database Design // Information Technology in the Humanities / S. Rahtz (ed.). Chichester: Ellis Horwood, 1987.*

вида Пакарда *Ibucis*, созданном для работы с *TLG* и некоторыми другими материалами классической филологии²⁷. Другое, контентно-адресуемое хранилище данных (*CAFS*), работало на британских компьютерах *ICL*²⁸. Идея переноса обработки данных на hardware казалась очень заманчивой гуманитариям, имеющим дело с большими массивами данных, но она так и не прижилась в полной мере, возможно, потому, что ее опередило быстрое усовершенствование стандартного аппаратного оборудования.

В публикациях этого периода преобладают работы на основе лексических исследований, порожденных программами конкордансов. Результаты представляли интерес либо для каких-то видов стилистического анализа, либо для использования в лингвистике. При различных лексических подсчетах применялась все более сложная математика, из-за чего некоторые, более гуманитарного склада, участники процесса оставались не у дел. Кроме того, в отношении методологии было мало действительно нового или любопытного и, возможно, меньше, чем хотелось бы, критической оценки методологий. Важные тенденции этого периода относятся больше к системам обслуживания, разработанным благодаря наличию все большего числа средств распространения (конференции и журналы), а также благодаря признанию необходимости стандартного программного обеспечения, архивирования и хранения текстов. Распространение осуществлялось через бюллетени по гуманитарному компьютерингу и намного меньше — через гуманитарные издания широкого профиля. Похоже, мы все еще пребывали на той стадии, когда гуманитарные исследования, осуществляемые с помощью компьютера, не пользовались особым уважением и ученые предпочитали публиковаться там, где их работы имели больший шанс быть принятыми.

Новые тенденции: середина 1980-х — начало 1990-х годов

В этот период произошли существенные сдвиги в развитии гуманитарного компьютеринга. Некоторые можно отнести к двум новым техническим решениям — персональному компьютеру и электронной почте. Другие произошли в результате роста экс-

27. The Humanities Computing Yearbook 1989–1990: A Comprehensive Guide to Software and Other Resources / I. Lancashire (ed.). Oxford: Clarendon Press, 1991. P. 204–205.

28. Burnard L. CAFS: A New Solution to an Old Problem // Literary and Linguistic Computing. 1987. Vol. 2. № 2.

плуатации технологий и в связи с необходимостью сократить затраты усилий.

Сначала появилось несколько разных конкурирующих брендов персональных компьютеров. Одни были созданы для игр, другие служили автономными текстовыми редакторами и не могли быть использованы ни для чего другого, третьи предназначались для рынка образования, а не для общего пользования. Постепенно персональные компьютеры *IBM* и модели на основе архитектуры *IBM* начали доминировать. Компьютеры *Apple Macintosh* также широко применялись, особенно в области графики.

Персональный компьютер теперь — неотъемлемая часть научной жизни, однако во времена своего появления он был значительно дороже, чем сейчас, и его первыми покупателями были энтузиасты и интересующиеся. Первоначальным результатом оказалось то, что больше не требовалось регистрироваться в вычислительном центре, чтобы воспользоваться компьютером. Пользователи ПК могли делать что хотели и уже не были ограничены тем, что имелось в доступе прежде.

К концу 1980-х появились три основанные на *DOS* программы текстового анализа: *Word-Cruncher*, *TACT* и *MicrOOSP* — все с очень хорошей функциональностью. Владельцы ПК могли работать с ними на дому и, в случае с *Word-Cruncher* и *TACT*, получать мгновенные результаты поиска. *MicrOOSP* была разработана из мэйнфреймовой программы с применением метода пакетного индексирования, а не интерактивного поиска. Однако главным приложением персональных компьютеров было то, что применялось во всех других дисциплинах, а именно обработка текстовых данных. Это привлекло еще больше пользователей, которые очень мало знали о других приложениях и были склонны считать, что функции программ текстовой обработки — это все, что компьютеры могли для них сделать.

Apple Macintosh привлек пользователей-гуманитариев по двум причинам. Во-первых, у него задолго до *Windows* был графический интерфейс. Это означало, что он гораздо удобнее с точки зрения отображения нестандартных символов. Наконец-то можно было теперь увидеть на экране древнеанглийские, греческие, кириллические символы и почти любой другой алфавит и легко работать с содержащим эти символы текстом. Во-вторых, *Macintosh* появился с программой, с помощью которой можно было легко создавать примитивные гипертексты. *HyperCard* предоставил модель карточных файлов со связями

между ними. Он включал и простой инструмент программирования с возможностью — впервые для ученых-гуманитариев — легко писать компьютерные программы. Вскоре была признана польза гипертекста для преподавания и появились разные примеры этого. Один из них — рабочая станция «Беовульф», созданная Патриком Коннером²⁹. Он дает текст со ссылками на современную английскую версию и различные языковедческие и контекстные примечания. Также первая версия проекта «Персей» была выпущена для последнего пользователя *HyperCard*.

Создание сети контактов, по крайней мере для электронной почты, первоначально предназначалось для групп программистов и исследовательских институтов. К середине 1980-х годов большинство академических компьютерных сервисных центров разработали средства отправления и получения электронной почты вне международных границ. Электронными адресами жадно обменивались на конференции *ALLC* в Ницце 1985 года; началась эра мгновенной коммуникации. Вскоре электронная почта отправлялась группам пользователей и была установлена программа для электронных групп дискуссий *ListServ*. *Ansaxnet*, самая старая электронная группа дискуссий в гуманитарии, была разработана Коннером в 1986 году³⁰.

Весной 1987 года на конференции *ICSH* в Колумбии, Южная Каролина, встретились специалисты, в гуманитарном компьютеринге работающие в основном в сфере поддержки, и сошлись на том, что нужно найти способ поддержания связи на регулярной основе. Уиллард МакКарти, в то время работавший в Университете Торонто, взялся за рассмотрение того, что они могли бы предпринять. После конференции он обнаружил существование *ListServ*, и вскоре появился «Гуманист»³¹. Первое сообщение было отправлено 7 мая 1987 года. МакКарти взял на себя функцию редактора того, что сам он предпочитает называть «электронным семинаром», и с тех пор, за исключением пропуска в начале 1990-х, когда «Гуманист» издавался Университетом Брауна, продолжает выступать в этой роли.

«Гуманист» стал чем-то вроде модели электронных групп дискуссий. МакКарти соблюдает отличные стандарты редакци-

29. *Conner P.* The Beowulf Workstation: One Model of Computer-assisted Literary Pedagogy // *Literary and Linguistic Computing*. 1991. Vol. 6. № 1.

30. *Idem.* Networking in the Humanities: Lessons from Ansaxnet // *Computers and the Humanities*. 1992. Vol. 26. № 3.

31. *McCarty W.* Humanist: Lessons from a Global Electronic Seminar // *Computers and the Humanities*. 1992. Vol. 26. № 3.

рования, и уровень дискуссии также в целом высокий. Для тех из нас, кто находится в Европе, регулярный утренний рацион из трех–шести дайджестов «Гуманиста» — приятное начало дня. «Гуманист», став главным в поддержке и развитии сообщества, оказал существенный вклад в формирование гуманитарного компьютеринга. Его архивы, восходящие к 1987 году, являют собой неисчерпаемый источник сведений касательно тенденций и проблем этого периода. Он был также взят за образец основателями *Linguist List*, главного электронного форума по лингвистике.

В этот период обретает печатную форму единственная широкомасштабная попытка составления библиографии проектов, программ и публикаций. Вышли два тома «Ежегодника гуманитарного компьютеринга» (*Humanities Computing Yearbook, HCY*). Первый, под редакцией Яна Ланкашира и Уилларда МакКарти, вышел в 1988 году и насчитывал около 400 страниц. Второй том, за 1989–1990 годы, — почти 700 страниц и с гораздо лучшим указателем. На протяжении нескольких лет, пока издание не начало устаревать, «Ежегодник» был чрезвычайно ценным источником, исполняя ту роль, которую ранее взял на себя журнал «Компьютеры и гуманитарные науки», переставший выходить в начале 1970-х. Подготовка «Ежегодника» была поистине трудоемким предприятием, и следующие тома не появились. К началу 1990-х сложилось общее мнение, что в будущем база данных онлайн была бы более эффективным ресурсом. Хотя и предпринимались разные попытки создать нечто подобное, ничего стоящего внимания так и не появилось, а общая картина проектов и публикаций снова оказывается неполной.

В отношении интеллектуальных прорывов на фоне других событий этого периода одно выделяется особенно. В ноябре 1987 года Нэнси Айд с помощью коллег из АСН организовала встречу по приглашениям в Колледже Вассара, Пукипси, для обсуждения возможности создания стандартной схемы кодирования для гуманитарных электронных текстов³². До сих пор предпринимались разные попытки решить проблему множества различных конфликтующих схем кодирования — ситуация, которую один из участников конференции в Вассаре назвал хаосом. Настало время действовать. Ученые уставали все больше, тратя время на реформатирование текстов под определенную программу, и испытывали все большую фрустрацию из-за недо-

32. Burnard L. Report of Workshop on Text Encoding Guidelines//Literary and Linguistic Computing. 1988. Vol. 3. № 2.

статков существующих схем. В 1986 году появилась новая схема кодирования. Стандартный общий язык разметки *SGML* (*Standard Generalized Markup Language*), выпущенный *ISO*, предложил такой механизм определения схемы разметки, который мог работать с различными типами текстов, обращаться с метаданными и представлять как сложный научный анализ, так и базовые структурные элементы документов.

Участники конференции согласовали ряд принципов («Принципы Пукипси») в качестве основы для разработки новой схемы кодирования и доверили руководство проектом Комитету управления с представителями от ассоциаций по компьютерам и гуманитарным наукам (*ACH*), по литературному и лингвистическому компьютерингу (*ALLC*) и по компьютерной лингвистике³³. Впоследствии группа собрала свыше миллиона долларов в Северной Америке и контролировала разработку «Принципов кодирования электронного текста и обмена» Инициативы по кодированию текстов (*Text Encoding Initiative, TEI*). Работа сначала велась в четырех областях, с комитетом при каждой из них. Результаты затем были объединены двумя редакторами в первый черновой вариант, распространенный в 1990 году для открытого комментирования. На следующем этапе к проекту были подключены рабочие группы, которые детально исследовали применимость этих методов к конкретным областям. Первая полная версия «Принципов» *TEI* была опубликована в мае 1994 года и распространена в печатном и электронном виде.

Размер, размах и воздействие *TEI* превзошли все ожидания участников конференции в Вассаре. Это была первая способная заинтересовать ученых систематическая попытка категоризировать и определить все свойства текстов в гуманитаристике. Всего около 400 тегов кодирования было задано в структуре, легко расширяемой для новых прикладных областей. Задавание тегов в «Принципах» проясняет некоторые вопросы, однако по ходу работы возникли более глубокие интеллектуальные проблемы. Работа в *TEI* вызвала интерес к теории разметки и к репрезентации гуманитарного знания в качестве темы как таковой. Публикация «Принципов» *TEI* совпала с появлением полнотекстовых цифровых библиотек, так что для проектов цифровых библиотек, которые до сих пор не пересекались с гуманитарным компьютерингом, оказалось естественным основывать работу на *TEI*, нежели изобретать схемы разметки с нуля.

33. *Text Encoding Initiative* (2001), URL: <http://www.tei-c.org>.

Большая часть работы *TEI* была сделана с помощью электронной почты с использованием частных и публичных групп дискуссий и через файловый сервер, куда вывешивались черновики документов. С самого начала любой, кто работал в группе *TEI*, должен был регулярно пользоваться электронной почтой, так что проект стал любопытным примером такого способа работы. Впрочем, участники вскоре поняли, что подводить итоги в электронных дискуссиях непросто, но, к счастью, имелись ресурсы и для регулярных личных встреч, с тем чтобы удостовериться, что решения действительно выполняются, а предложения по разметкам от различных рабочих групп обоснованны.

Как мы видим, помимо важных достижений в использовании персональных компьютеров, нетворкинге и *TEI* в 1970-е годы бурно развивался и гуманитарный компьютеринг, вовлекая все больше пользователей и проектов. Некоторые прикладные области постепенно отделились от обработки данных в гуманитаристике, развив собственную культуру и способы распространения. «Компьютеры и письмо» оказались темой, довольно быстро себя исчерпавшей. Компьютерная лингвистика развивалась всегда независимо от гуманитарного компьютеринга и, несмотря на усилия Дона Уокера в оргкомитете *TEI*, оставалась независимой дисциплиной. Уокер и Антонио Замполли из Института компьютерной лингвистики в Пизе много работали над объединением сообществ гуманитарного компьютеринга и компьютерной лингвистики, однако без большого успеха. Как раз когда ученые в гуманитарном компьютеринге начинали испытывать серьезную потребность в инструментах, разработанных в компьютерной лингвистике (морфологический анализ, синтаксический анализ и лексические базы данных), в компьютерной и корпусной лингвистике велась активная работа по аргументационному анализу и анализу речи. Несмотря на знаковый доклад о взаимодействии между компьютерной лингвистикой и литературным и лингвистическим компьютерингом, прочитанный Замполли и его коллегой Николеттой Кальцолари на первой объединенной конференции *ACH/ALLC* в Торонто в июне 1989 года³⁴, эти два сообщества почти не контактировали, и той пользы, какую мог бы извлечь из методик компьютерной лингвистики, гуманитарный компьютеринг не извлек.

34. Calzolari N., Zampolli A. Lexical Databases and Textual Corpora: A Trend of Convergence between Computational Linguistics and Literary and Linguistic Computing // Research in Humanities Computing / S. Hockey, N. Ide (eds). Oxford, 1991. Vol. 1.

Эра интернета: от начала 1990-х до наших дней

Одно событие по своему воздействию далеко превзошло любое другое в период 1990-х. Это был приход интернета, а особенно Всемирной паутины (*World Wide Web*). Первый графический браузер *Mosaic* появился на сцене в 1993 году. Отныне использование интернета — жизненно важная часть любой академической деятельности. Поколение студентов выросло с ним и естественно воспринимает его как главный источник любой информации.

Изначально специалистам, работающим в долгосрочных проектах гуманитарного компьютеринга, было сложно осознать возможное воздействие сети почти в тех же масштабах, что и *Microsoft*'а. Работающие с *TEI* остро почувствовали, что язык гипертекстовой разметки *HTML* (*HyperText Markup Language*) был слабой системой разметки, навсегда сохранившей все проблемы с текстовыми редакторами и с основанной на этом виде разметкой. Сеть просматривалась с любопытством, но, скорее, извне. Она была средством поиска каких-то видов информации, а не действительно серьезным инструментом гуманитарных исследований, но она стала первой возможностью для тех институций и организаций, которые первыми намерились освоить компьютеринг в гуманитарных науках. Они увидели, что сеть является и превосходным средством для распространения результатов научной работы не только среди коллег, но и среди намного более широкой аудитории. Возникла новая группа пользователей. С точки зрения производителя, преимущества интернета огромны. Формат более не ограничивается форматом печатной книги. Теоретически почти отсутствуют ограничения в объеме, и гипертекстовые ссылки предоставляют удобный способ работы с примечаниями и пр. Публикация может выстраиваться постепенно, по мере того как и когда ее части готовы к изданию. Ее можно выложить сразу, легко исправлять и обновлять.

С начала до середины 1990-х годов было заявлено множество новых проектов, некоторые из них нашли финансирование и были запущены. В частности, в области научных электронных изданий было несколько встреч и публикаций, посвященных дискуссии о том, на что может быть похожим электронное издание³⁵. Происходило это как раз в то время, когда теоретики редакторского дела работали над текстом как физическим объ-

35. *Finneran R.* The Literary Text in the Digital Age. Ann Arbor: University of Michigan Press, 1996; *Bornstein G., Tinkle T.* The Iconic Page in Ma-

ектом, который они могли бы репрезентировать в цифровых образах. За знаменательным исключением работы, проделанной Питером Робинсоном³⁶ и, возможно, еще парой других специалистов, немногие из этих публикаций увидели свет иначе, чем в виде прототипов или небольших образцов, и ко второй половине десятилетия интерес к этому несколько угас. Было предложено множество творческих идей, но, как только они оказывались на стадии, где теория должна перейти в практику, а проекты требовали усердного труда по вводу, текстовой разметке и разработке программного обеспечения, внимание переносилось на другие вещи.

Велись споры о том, как называть коллекции электронных ресурсов. Многие предпочитали термин «архив» (*archive*). В частности, так были названы «Архив Блейка» и другие проекты в Институте передовых технологий в гуманитарных науках в Университете Вирджинии. «Архив» означал коллекцию материала, где обычно пользователь должен был выбирать метод навигации. «Версия» (*edition*) подразумевает значительную долю научной прибавочной стоимости, поскольку отражает представления одного или более редактора, что может быть осуществлено благодаря предпочтению определенных методов навигации. Язык разметки *SGML*, в основном в созданных на основе *TEI* приложениях, был принят как способ снабжения добавочными блоками, на которых могли быть построены маршруты навигации. Однако оставались нерешенными серьезные задачи по проектированию и построению эффективного пользовательского интерфейса. И все же основной упор делался на навигации, а не на инструментах и методах анализа, которые в прошлом сформировали в гуманитарном компьютеринге основные сферы приложения. На заре появления сети технология распространения текстов, размеченных по стандарту *SGML*, была неуклюжей и во многих отношениях предоставляла менее удовлетворительный пользовательский интерфейс, чем тот, что может быть выработан неформатированным *HTML*. Но из-за легкости

nuscript, Print, and Digital Culture. Ann Arbor: University of Michigan Press, 1998.

36. Geoffrey Chaucer: The Wife of Bath's Prologue on CD-ROM / P. Robinson (ed.). Cambridge: Cambridge University Press, 1996; *Idem*. New Directions in Critical Editing // Electronic Text: Investigations in Method and Theory / K. Sutherland (ed.). Oxford: Clarendon Press, 1997; *Idem*. New Methods of Editing, Exploring and Reading The Canterbury Tales. 1999. URL: <http://www.cta.dmu.ac.uk/projects/ctp/desc2.html>.

просмотра этих текстов воздействие многих издательских проектов такого рода было существенным. Для большего числа людей идея технологии в гуманитарии стала привычной, хотя это в меньшей мере касалось загрузки материалов в сеть.

Хотя поначалу большинство из издательских проектов были запущены коллективами ученых, вскоре и библиотеки стали обдумывать идею выкладывания содержания своих коллекций в интернет. В США несколько институций создали коллекции электронных текстов и цифровых библиотек в качестве гуманитарных первоисточников, обычно используя поисковый механизм *OpenText SGML*³⁷. Обладая хорошими и быстрыми возможностями поиска слов (цепочек), он в то же время позволяет немногим больше, чем справочная система для поиска слов. Другие проекты использовали систему электронных книг *DynaText SGML* для распространения своих текстов. Это дало возможность более структурированного поиска, однако с не особенно удобным интерфейсом.

Совершенно новая идея электронных публикаций принадлежит проекту «Орландо», работающему с историей британской женской литературы в университетах Альберты и Гуэльфа. При существенной финансовой помощи новый материал в виде кратких биографий авторов, истории их письма и общих мировых событий был создан как совокупность документов стандарта *SGML*³⁸. Затем можно было работать над извлечением порций этих документов и преобразованием их в новый материал, например, для того, чтобы генерировать хронологии отдельных периодов или тематик. Этот проект представил собой идею совершенно новой формы академического письма, идею, принципиально отличную от всего, что было сделано до сих пор. Остается только следить за тем, действительно ли она наберет обороты.

Благодаря интернету стало реально выполнять совместные проекты ранее невозможными способами. Простая возможность участия в составлении одного и того же документа людей, находящихся в разных местах, была большим прорывом в более ранних методах работы. В проекте «Орландо» исследователи обеих институций создают дополнения к архиву документа, который раз-

37. *Price-Wilkin J.* Using the World Wide Web to Deliver Complex Electronic Documents: Implications for Libraries // *The Public-Access Computer Systems Review.* 1994. Vol. 5. № 3.

38. *Brown S. et al.* SGML and the Orlando Project: Descriptive Markup for an Electronic History of Women's Writing // *Computers and the Humanities.* 1997. Vol. 31. № 4.

работан на базе *web*-технологии, в административных целях использующей элементы разметки *SGML*. Распространялись также идеи по совместной редакции рукописных источников, при которой специалисты, находясь в разных местах, могли бы добавлять слои примечаний, например для проекта Пирс³⁹ и *Codex Leningradensis*⁴⁰. Технические стороны этой идеи достаточно прозрачны. Может быть, менее ясно управление проектом: кто контролирует или проверяет примечания и как все это можно сохранять.

Внедрение *TEI* в качестве модели в проектах цифровых библиотек затронуло некоторые любопытные вопросы относительно целой философии *TEI*, задуманной большей частью учеными, которые желали быть возможно более гибкими. Любой тег *TEI* может быть изменен, теги можно добавлять там, где это уместно. Несколько иная философия преобладает в библиотечно-информационной науке, где заданным стандартам следуют с большим тщанием, для того чтобы обеспечить читателям легкий поиск книг. К сожалению, большего вклада со стороны библиотечно-информационной науки в то время, когда создавалась *TEI*, ожидать не приходилось, но проект *TEI* был запущен задолго до того, как термин «цифровая библиотека» вошел в обиход. Среди библиотекарей многолетний опыт работы с электронными текстами был редкостью в отличие от сообщества ученых. Тем не менее *TEI* была взята в качестве модели разработчиками кодированного архивного описания *EAD* (*Encoded Archival Description*), возымевшего очень широкое воздействие в качестве стандарта поисковых вспомогательных средств в архивах и специальных коллекциях.

В начале 1990-х к электронным ресурсам гуманитарных наук добавилось дополнительное измерение — тогда появилась возможность подачи мультимедийной информации в виде изображений, аудио и видео. На заре эпохи цифровых изображений было много дискуссий о файловых форматах, цветовой разрешающей способности, других технических аспектах формирования изображения и гораздо меньше о том, что же люди могут с ними делать помимо просмотра. Безусловно, в доступе к изображениям первоисточников в сети есть множество преимуществ, однако специалисты по гуманитарному компьютерингу,

39. Neuman M., Keeler M., Kloesel C., Ransdell J., Renear A. The Pilot Project of the Electronic Peirce Consortium (abstract)//ALLC/ACH '92 Conference Abstracts and Program. Oxford, 1992.

40. Leningrad Codex Markup Project//Project «EL»: The XML Leningrad Codex. 2000. URL: <http://www.leningradensis.org/>.

воспитанные на привычке к гибкости текста с возможностью поиска, опять-таки склонны были рассматривать связанные с изображениями проекты как не совсем их дело, если только,— как в проекте «Беовульф»⁴¹,— с изображениями нельзя было работать и каким-то образом увеличивать их ценность. В настоящее время ведется интересная работа по соединению образов с текстом, вплоть до отдельных слов⁴². Когда большую часть этой работы можно будет выполнять автоматически, перед нами предстанет возможность переосмысления некоторых аспектов изучения рукописей. Потенциал других форм мультимедиа сегодня широко признан, но их применение действительно оправданно лишь при высокоскоростном доступе, а будущее состоит в их постепенном сближении с телевидением.

Распространение доступа к электронным ресурсам, порожденное сетью, вызвало интерес к другим теоретическим областям в гуманитарном компьютеринге. Электронные ресурсы стали отдельной темой исследований: их начала анализировать новая группа ученых, среди которых и люди без большого опыта работы с техническими аспектами ресурсов. В частности, внушительное количество теоретиков привлекало гипертекст. Это помогло расширить спектр интереса дискуссий о компьютеринге в гуманитарных науках, но в то же время, возможно, повлияло и на формирование превратных представлений о том, что происходит на самом деле при создании и использовании такого ресурса. Снова возникли проблемы с двумя культурами — той, которая ресурс создает, и другой, предпочитающей об этом говорить.

Введение академических программ служит обычно еще одним подтверждением принятия какой-либо дисциплины более широкой академической средой. В случае гуманитарного компьютеринга это начало происходить к концу 1990-х, хотя, наверное, любопытно будет отметить, что очень немногие из программ содержали слова «гуманитарный компьютеринг» в самом названии. Королевский колледж в Лондоне предлагает студентам-бакалаврам гуманитарных специальностей дополнительную программу по прикладному компьютерингу с набором гуманитарных дисциплин, а его новая магистерская программа при Центре гуманитарного компьютеринга также называется «Прикладной

41. *Kiernan K.* Digital Image Processing and the Beowulf Manuscript // *Literary and Linguistic Computing*. 1991. Vol. 6. № 1.

42. *Zweig R.* Lessons from the Palestine Post Project // *Literary and Linguistic Computing*. 1998. Vol. 13. № 2.

компьютинг». МакМастерский университет в Канаде предлагает бакалавриат по специальности «мультимедиа». Магистерская программа, которую вскоре откроет Университет Вирджинии, называется «Цифровые гуманитарные науки» и находится под покровительством программы «Исследования медиа». Университет Альберты, насколько мне известно, первый открывает программу с «гуманитарным компьютерингом» в названии, хотя в Университете Глазго на магистратуре философского факультета специальность «история и компьютеринг» существует уже много лет. Поскольку интернет способствовал все более широкому применению компьютеров в гуманитарных науках, к процессу начали подключаться другие организации, что привело в дальнейшем к некоторым попыткам определить это дисциплинарное поле или по крайней мере его текущие исследовательские задачи. В 1996 году отдел информатики в истории искусств музея Гетти выпустил, на мой взгляд, очень интересную программу исследований сетевого культурного наследия⁴³. В ней восемь статей, которые охватывают такие области, как цифровые библиотеки, гуманитарные исследования и преподавание. Каждая из областей могла бы сформировать отдельную исследовательскую программу, однако инициатива не была развита. Тем временем *ALLC* и *ACH* продолжали устраивать ежегодные конференции, где преобладали доклады о разметке и других технических вопросах. Попытка выработать план действий и новые направления в гуманитарном компьютеринге на конференции 2002 года в Германии дала в результате полезный обзор⁴⁴, однако едва ли новый: к работе, возможно, стоило бы привлечь более широкий круг специалистов. Однако задействовать другие научные круги становилось все проблематичнее в эпоху, когда все больше электронных ресурсов гуманитарных наук создавались за пределами сообщества гуманитарного компьютеринга.

Заключение

Если нужно выделить самое важное направление гуманитарного компьютеринга, то, на мой взгляд, это должна быть *TEI*. Она представляет собой самый существенный интеллектуальный

43. Research Agenda for Networked Cultural Heritage/D. Bearman (ed.). Santa Monica, CA: Getty Art History Information Program, 1996.

44. Robey D. Introduction: new directions in humanities computing//Literary and Linguistic Computing. 2003. Vol. 18. № 1.

прорыв среди всего, что было сделано в нашей области. *TEI* оказала воздействие на все сообщество, занимающееся разметкой. Она привлекла внимание лидирующих специалистов *SGML* в то время, когда разрабатывался расширяемый язык разметки *XML*, и Майкл Сперберг МакКуин, один из редакторов *TEI*, был приглашен в качестве соредактора нового стандарта *XML*. Сделанная в *TEI* работа по связыванию через гиперссылки послужила основой связывающих механизмов в пределах *XML*. Во многих отношениях *TEI* была впереди своего времени, поскольку лишь с быстрым внедрением *XML* на протяжении последних двух-трех лет потребность в описательной разметке была признана более широкими кругами. Между тем выделившаяся из *TEI* группа теоретиков разметки продолжает ставить актуальные вопросы по репрезентации знания.

До сих пор остаются области, пока еще тщательно не исследованные. Гуманитарный компьютеринг может оказать существенное влияние на растущий интерес к выкладыванию культурного наследия в интернете — не только для пользователей академических кругов, но и для занимающихся самообразованием, для широкой аудитории в целом. Инструменты и методы, разработанные в гуманитарном компьютеринге, облегчат изучение этого наследия, а внедрение методов компьютерной лингвистики, как показывает проект «Персей»⁴⁵, может добавить и новые области работы. Наши инструменты и методы также могут помочь исследованиям, облегчая процессы оцифровки и кодирования там, где нам нужно находить пути снижения затрат на создание данных без потери научного значения или функциональности. Благодаря интернету гуманитарный компьютеринг привлекает гораздо более широкую аудиторию, и студенты, которые оканчивают новые предложенные программы, смогут заниматься не только наукой, но и электронной издательской деятельностью, образовательными технологиями и разработкой мультимедиа. На протяжении своей истории гуманитарный компьютеринг продемонстрировал здоровый аппетит к воображению и инновации, продолжая в то же время придерживаться высоких научных стандартов. Сейчас, когда интернет является важным элементом повседневности, для гуманитарного компьютеринга существует возможность достичь гораздо большего, чем было возможно до сих пор.

45. *Rydberg-Cox J. Co-occurrence Patterns and Lexical Acquisition in Ancient Greek Texts // Literary and Linguistic Computing. 2000. Vol. 15. № 2.*

The History of Humanities Computing

Susan Hockey. Emeritus Professor of Library and Information Studies (SLAIS) at the Department of Information Studies of the University College London (UCL). Address: Gower Street, London WC1E 6BT, United Kingdom. E-mail: s.hockey@ucl.ac.uk.

Keywords: computing; methodology; digitization; encoding; data.

The author traces the history of the Digital Humanities (humanities computing) from 1950 to 1990. She starts from Roberto Busa's pioneering endeavour to render the works of Thomas Aquinas automatically treatable, and goes on to look at the Text Encoding Initiative (TEI). The article covers a variety of methods and a large range of topics, with a clear interdisciplinary inclination. By its very nature, humanities computing has had to embrace "the two cultures," to bring in the rigour and the unambiguous, systematic, procedural methodologies characteristic of the sciences to address

problems within the humanities that had hitherto been most often treated in an intuitive and serendipitous fashion. The pinnacle of achievement for this period has been the TEI, which has influenced the professional community as a whole. Humanities computing can contribute substantially to the growing interest in putting humanity's cultural heritage on the internet, not only for academic users, but also for lifelong learners and the general public. Our tools and techniques can also assist researchers in facilitating the digitization and encoding processes, where ways to reduce the costs of data creation, without loss of scholarly value or functionality, have to be found. Throughout its history, humanities computing has shown a healthy appetite for imagination and innovation, while continuing to maintain high scholarly standards. Now that the internet is such a dominant feature of everyday life, the opportunity exists for humanities computing to reach out much further than has hitherto been possible.

References

- Berry-Rogghe G., Crawford T. Developing a Machine-independent Concordance Program for a Variety of Languages. *Computer and Literary Studies* (eds A. J. Aitken, R. W. Bailey, N. Hamilton-Smith), Edinburgh, Edinburgh University Press, 1973, pp. 309–316.
- Bessinger J., Parrish S. *Literary Data Processing Conference Proceedings*, White Plains, New York, IBM, 1965.
- Bornstein G., Tinkle T. *The Iconic Page in Manuscript, Print, and Digital Culture*, Ann Arbor, University of Michigan Press, 1998.
- Brown S. et. al. SGML and the Orlando Project: Descriptive Markup for an Electronic History of Women's Writing. *Computers and the Humanities*, 1997, vol. 31, no. 4, pp. 272–284.
- Brunner T. Classics and the Computer: The History of a Relationship. *Accessing Antiquity: The Computerization of Classical Studies* (ed. J. Solomon), Tucson, London, University of Arizona Press, 1993, pp. 10–33.
- Burnard L. CAFS: A New Solution to an Old Problem. *Literary and Linguistic Computing*, 1987, vol. 2, no. 2, pp. 7–12.
- Burnard L. Principles of Database Design. *Information Technology in the Humanities* (ed. S. Rahtz), Chichester, Ellis Horwood, 1987, pp. 54–68.
- Burnard L. Report of Workshop on Text Encoding Guidelines. *Literary and Linguistic Computing*, vol. 3, no. 2, pp. 131–133.
- Burton D. Automated Concordances and Word Indexes: Machine Decisions and Editorial Revisions. *Computers and the Humanities*, 1982, vol. 16, no. 4, pp. 195–208.

- Burton D. Automated Concordances and Word Indexes: The Early Sixties and the Early Centers. *Computers and the Humanities*, 1981, vol. 15, no. 2, pp. 83–100.
- Burton D. Automated Concordances and Word Indexes: The Fifties. *Computers and the Humanities*, 1981, vol. 15, no. 1, pp. 1–14.
- Burton D. Automated Concordances and Word Indexes: The Process, the Programs, and the Products. *Computers and the Humanities*, 1981, vol. 15, no. 3, pp. 139–154.
- Busa R. *Index Thomisticus: Sancti Thomæ Aquinatis operum indices et concordantiæ: In 49 vols*, Stuttgart, Frommann-Holzboog, 1974–.
- Busa R. Picture a Man... *Literary and Linguistic Computing*, 1999, vol. 14, no. 1, pp. 5–9.
- Busa R. The Annals of Humanities Computing: The Index Thomisticus. *Computers and the Humanities*, October 1980, vol. 14, no. 2, pp. 83–90.
- Calzolari N., Zampolli A. Lexical Databases and Textual Corpora: A Trend of Convergence between Computational Linguistics and Literary and Linguistic Computing. *Research in Humanities Computing* (eds S. Hockey, N. Ide), Oxford, 1991, vol. 1, pp. 273–307.
- Conner P. Networking in the Humanities: Lessons from Ansaxnet. *Computers and the Humanities*, 1992, vol. 26, no. 3, pp. 195–204.
- Conner P. The Beowulf Workstation: One Model of Computer-assisted Literary Pedagogy. *Literary and Linguistic Computing*, 1991, vol. 6, no. 1, pp. 50–58.
- De Tollenaere F. The Problem of the Context in Computer-aided Lexicography. *The Computer and Literary Studies* (eds A. J. Aitken, R. W. Bailey, N. Hamilton-Smith), Edinburgh, Edinburgh University Press, 1973, pp. 25–35.
- Ellegård A. *A Statistical Method for Determining Authorship: The Junius Letters 1769–1772*, Gothenburg, Elanders Boktryckeri Aktiebolag, 1962.
- Finneran R. *The Literary Text in the Digital Age*, Ann Arbor, University of Michigan Press, 1996.
- Geoffrey Chaucer: *The Wife of Bath's Prologue on CD-ROM* (ed. P. Robinson), Cambridge, MA, Cambridge University Press, 1996.
- Gorcy G. L'informatique et la mise en œuvre du trésor de la langue française (TLF), dictionnaire de la langue du 19e et du 20e siècle (1789–1960). *The Possibilities and Limits of the Computer in Producing and Publishing Dictionaries* (eds A. Cappelli, A. Zampolli), Pisa, Giardini, 1984, p. 119–144.
- Hamilton-Smith N. A Versatile Concordance Program for a Textual Archive. *The Computer in Literary and Linguistic Research* (ed. R. Wisbey), Cambridge, MA, Cambridge University Press, 1971, pp. 235–244.
- Healey A. The Corpus of the Dictionary of Old English: Its Delimitation, Compilation and Application. *Fifth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary* (September 18–19, 1989, St. Catherine's College, Oxford), Waterloo, ON, UW Centre for the New OED, 1989, pp. 113–124.
- Hockey S. The History of Humanities Computing. *A Companion to Digital Humanities* (eds S. Schreibman, R. Siemens, J. Unsworth), Oxford, Wiley-Blackwell, 2004, pp. 3–19.
- Hockey S. Workshop on Teaching Computers and the Humanities Courses. *Literary and Linguistic Computing*, 1986, vol. 1, no. 4, pp. 228–229.
- Hockey S., Marriott I. The Oxford Concordance Project (OCP). Part 1–3. *ALLC Bulletin* 7, 1979, pp. 35–43, 155–164, 268–275.
- Hockey S., Marriott I. The Oxford Concordance Project (OCP). Part 4. *ALLC Bulletin* 8, 1980, pp. 28–35.
- Holmes D., Forsyth R. The Federalist Revisited: New Directions in Authorship Attribution. *Literary and Linguistic Computing*, 1995, vol. 10, no. 2, pp. 111–127.

- Kiernan K. Digital Image Processing and the Beowulf Manuscript. *Literary and Linguistic Computing*, 1991, vol. 6, no. 1, pp. 20–27.
- Leningrad Codex Markup Project. *Project “EL”: The XML Leningrad Codex*, 2000. Available at: <http://leningradensis.org/>.
- Lord R. Studies in the History of Probability and Statistics: viii. de Morgan and the Statistical Study of Literary Style. *Biometrika*, 1958, vol. 45, no. 1–2, p. 282.
- McCarty W. Humanist: Lessons from a Global Electronic Seminar. *Computers and the Humanities*, 1992, vol. 26, no. 3, pp. 205–222.
- Mendenhall T. C. A Mechanical Solution of a Literary Problem. *The Popular Science Monthly*, 1901, vol. 60, no. 7, pp. 97–105.
- Morton A. *The Authorship of the Pauline Epistles: A Scientific Solution*, Saskatoon, 1965.
- Morton A., Winspear A. *It's Greek to the Computer*, Montreal, Harvest House, 1971.
- Mosteller F., Wallace D. *Inference and Disputed Authorship: The Federalist*, Reading, MA, Addison-Wesley, 1964.
- Neuman M., Keeler M., Kloesel C., Ransdell J., Renear A. The Pilot Project of the Electronic Peirce Consortium (abstract). *ALLC/ACH '92 Conference Abstracts and Program*, Oxford, 1992, pp. 25–27.
- Parrish S. Problems in the Making of Computer Concordances. *Studies in Bibliography*, 1962, vol. 15, no. 1, pp. 1–14.
- Price-Wilkin J. Using the World Wide Web to Deliver Complex Electronic Documents: Implications for Libraries. *The Public-Access Computer Systems Review*, 1994, vol. 5, no. 3, pp. 5–21.
- Proud J. *The Oxford Text Archive*, London, British Library R&D Report, 1989.
- Research Agenda for Networked Cultural Heritage* (ed. D. Bearman), Santa Monica, CA, Getty Art History Information Program, 1996.
- Robey D. Introduction: New Directions in Humanities Computing. *Literary and Linguistic Computing*, 2003, vol. 18, no. 1, pp. 3–9.
- Robinson P. M. W. New Directions in Critical Editing. *Electronic Text: Investigations in Method and Theory* (ed. K. Sutherland), Oxford, Clarendon Press, 1997, pp. 145–171.
- Robinson P. M. W. *New Methods of Editing, Exploring and Reading The Canterbury Tales*, 1999. Available at: <http://ctadmu.ac.uk/projects/ctp/desc2.html>.
- Russell D. *COCOA—A Word Count and Concordance Generator for Atlas*, Chilton, UK, Atlas Computer Laboratory, 1967.
- Rydberg-Cox J. Co-occurrence Patterns and Lexical Acquisition in Ancient Greek Texts. *Literary and Linguistic Computing*, 2000, vol. 15, no. 2, pp. 121–129.
- The Computer in Literary and Linguistic Research* (ed. R. Wisbey), Cambridge, MA, Cambridge University Press, 1971.
- The Humanities Computing Yearbook 1989–1990: A Comprehensive Guide to Software and Other Resources* (ed. I. Lancashire), Oxford, Clarendon Press, 1991.
- Thomae Aquinatis Opera Omnia Cum Hypertextibus in CD-ROM* (ed. R. Busa), Milano, 1992.
- Tweedie F. J., Singh S., Holmes D. I. Neural Network Applications in Stylogometry: The Federalist Papers. *Computers and the Humanities*, 1996, vol. 30, no. 1, pp. 1–10.
- Wisbey R. The Analysis of Middle High German Texts by Computer: Some Lexicographical Aspects. *Transactions of the Philological Society*, 1963, vol. 62, no. 1, pp. 28–48.
- Zweig R. Lessons from the Palestine Post Project. *Literary and Linguistic Computing*, 1998, vol. 13, no. 2, pp. 89–97.