

Научная статья /
Research Article

<https://elibrary.ru/CNHBOA>
УДК 82
ББК 83 + 79.3

«БОЛЬШИЕ ДАННЫЕ» ЦИФРОВОГО АРХИВА: ДИАЛОГ С РАСТРОВОЙ РУКОПИСЬЮ

© 2023 г. Л.В. Хачатурян

*Национальный исследовательский университет
«Высшая школа экономики», Москва, Россия*

Дата поступления статьи: 05 декабря 2022 г.

Дата одобрения рецензентами: 19 января 2023 г.

Дата публикации: 25 июня 2023 г.

<https://doi.org/10.22455/2500-4247-2023-8-2-334-349>

*Исследование выполнено в рамках проекта Российского научного фонда № 22-28-01569,
НИУ «Высшая школа экономики»*

Аннотация: В статье отражены современные тенденции работы с цифровым наследием русской литературы, рассмотрен процесс формирования виртуальных архивов как постепенного накопления «большой даты» научных исследований — нераспознанного информационного массива растровых документов, содержащего десятки тысяч изображений. В ходе исследования анализируется возникшая на рубеже XX–XXI вв. специфика научной работы в области эго-документального наследия (корпус дневниковых записей, рабочих тетрадей, записных книжек, переписки), принципы публикации и современные стандарты цифровизации архивного наследия. На основе изучения и практики работы трех наиболее перспективных виртуальных ресурсов по истории русской литературы середины XIX – первой половины XX вв. сформулированы конкретные задачи и методы визуализации большого корпуса растровых образов архивных документов, а также не задействованные ранее возможности автоматизации поискового аппарата. Большое внимание уделяется переходу от графических элементов растрового изображения рукописи к семантическим, позволяющим применить элементы интеллектуального анализа (data mining) для нераспознанного массива данных.

Ключевые слова: эго-документы, архивные материалы, цифровой архив, русская литература XX в., рукописное наследие, большие данные, интеллектуальный анализ данных.

Информация об авторе: Любовь Валерьевна Хачатурян — кандидат культурологии, доцент, Национальный исследовательский университет «Высшая школа экономики», ул. Мясницкая, д. 20, 101000 г. Москва, Россия.
ORCID ID: <https://orcid.org/0000-0002-2689-5186>

E-mail: rgaliz2010@yandex.ru

Для цитирования: *Хачатурян Л.В.* «Большие данные» цифрового архива: диалог с растровой рукописью // *Studia Litterarum.* 2023. Т. 8, № 2. С. 334–349.
<https://doi.org/10.22455/2500-4247-2023-8-2-334-349>



This is an open access article distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0)

Studia Litterarum,
vol. 8, no. 2, 2023

“BIG DATA” OF THE DIGITAL ARCHIVE: A DIALOGUE WITH A RASTER MANUSCRIPT

© 2023, Lyubov V. Khachaturian
*National Research University Higher School
of Economics, Moscow, Russia*
Received: December 05, 2022
Approved after reviewing: January 19, 2023
Date of publication: June 25, 2023

Acknowledgements: The research was carried out at National Research University Higher School of Economics within the framework of the project of the Russian Science Foundation no. 22-28-01569.

Abstract: The article reflects the current trends in working with the digital heritage of Russian literature, examines the process of forming virtual archives as a gradual accumulation of the “big data” of scientific research, i. e. unrecognized information array of raster documents containing tens of thousands of images. The research analyzes the specifics of scientific work in the field of ego-documentary heritage that arose at the turn of the 20th – 21st centuries (a corpus of diary entries, workbooks, notebooks, correspondence), the principles of publication and modern standards of digitization of archival heritage. The study and practicing of the three most promising virtual resources on the history of Russian literature of the mid-19th – first half of the 20th centuries allows to formulate specific tasks and methods of visualization of a large corpus of raster images of archival documents, as well as previously untapped possibilities of search engine automation. Much attention is paid to the transition from the graphical elements of the raster image of the manuscript to semantic ones, which allow the use of data mining elements for an unrecognized data array.

Keywords: ego-documentary heritage, archival materials, digital archive, Russian literature of the 20th century, handwritten heritage, big data, data mining.

Information about the author: Lyubov V. Khachaturian, PhD in Cultural Studies, Associate Professor, National Research University Higher School of Economics, Myasnitskaya 20, 101000 Moscow, Russia. ORCID ID: <https://orcid.org/0000-0002-2689-5186>

E-mail: rgaliz010@yandex.ru

For citation: Khachaturian, L.V. “‘Big Data’ of the Digital Archive: A Dialogue with a Raster Manuscript.” *Studia Litterarum*, vol. 8, no. 2, 2023, pp. 334–349. (In Russ.) <https://doi.org/10.22455/2500-4247-2023-8-2-334-349>

В своей книге «Дальнее чтение» социолог литературы Франко Моретти назвал художественное наследие «Великим Непрочтенным». «Вопрос даже не в том, *что* стоит исследовать, вопрос в том, *как*. “Я занимаюсь западно-европейской прозой”... Это не совсем так, потому что я занимаюсь лишь канонизированной ее частью, которая составляет меньше одного процента всей изданной литературы» [10, с. 79]¹. Проблема «*как* исследовать» справедлива и для литературного наследия XX в. К сожалению, архивные данные только подтверждают приведенную закономерность — более 92 % рукописных документов до настоящего времени не опубликованы и фактически не изучены; по аналогии с вышеприведенной метафорой рукописное наследие до сих пор остается «великим нераспознанным».

Традиционно в качестве завершающей стадии и основы академической публикации принимался критически установленный текст². При этом уже на рубеже XX–XXI вв. как оригинальные эго-тексты стали исследоваться и публиковаться полные корпуса записных книжек, рабочих тетрадей, дневников и переписки, ранее привлекаемые только как материал для

1 Курсив Ф. Моретти. Сама метафора восходит к работе Маргарет Коэн «Сентиментальное воспитание романа» [14]. В следующей главе («Литературная бойня») Моретти приводит более точную статистику: 0,5 % прочитанного противостоит 99,5 % «великого непрочтенного» [10, с. 106].

2 В качестве издания, подводящего итог развитию академической школы текстологии второй половины XX в., можно выделить монографию Л.А. Спиридоновой «Текстология: теория и практика» [12]. В указанной работе «критически установленный текст» понимается как «текст классического произведения, который с наибольшей полнотой отражает творческую волю автора» [12, с. 10]. При этом в качестве задачи текстологии как научной дисциплины рассматривается «подготовка к научному изданию полного собрания сочинений писателей-классиков» [12, с. 10].

критического изучения, восстановления и комментирования основного или даже «канонического» текста³. Параллельно происходил рост интереса к архивному наследию в целом. Самостоятельным и принципиально новым объектом изучения стал авантекст — от черновых автографов в блокнотах и рабочих тетрадях до маргиналий на страницах уже опубликованных книг. «Многожильный провод» литературного процесса сконцентрировал внимание на обширном корпусе материалов, относящихся к бытованию текстов в социуме, делая предметом исследования «официальные документы» (стенограммы, протоколы, анкеты и т. п.) и деловую переписку.

Дополнительным фактором, определившим новую архивную оптику, стала цифровая трансформация, произошедшая в архивах в последние десятилетия. В начале XXI в. появились два принципиально новых инструмента работы с рукописным наследием. Наравне с подлинником в исследовании и архивном деле начали использоваться электронные копии документов (собственно, с этого времени рукописное наследие и получило предикат «стать цифровым»)⁴. Работа с электронными изображениями позволила заполнить, казалось бы, невосполнимые лакуны в рукописном тексте. Графическая обработка цифровых копий предоставляет возможность проявить следы осыпавшегося карандаша или выцветших чернил и прочесть запись, ранее считавшуюся утраченной⁵. Лучший пример — оцифровка личной библиотеки писателя, правки *post publicationem*, которую автор остав-

3 См.: «Тщательно выправив этот текст с помощью сопоставления его с другими источниками, исключив правку цензора и редактора, не согласованную с писателем, текстолог получает именно тот текст, который называют *каноническим* (курсив Л.А. Спиридоновой. — Л.Х.) или окончательным» [12, с. 15].

4 Речь идет об Электронном фонде пользования (ЭФП): «Электронный фонд пользования (ЭФП) представляет собой совокупность электронных копий документов Архивного фонда РФ, записанную на цифровые носители и предназначенную для использования вместо подлинников документов, что обеспечивает сохранность документов и возможность формирования электронных ресурсов, обеспечивающих оперативность доступа к документу, в т. ч. с использованием интернет-технологий» [22, п. 2.1].

5 Бесконтактная цифровая реставрация рукописных документов — одно из наиболее перспективных направлений научно-исследовательских разработок в сфере архивоведения. В статье Л.И. Бородкина «Digital history: Применение цифровых медиа в сохранении культурного наследия?» виртуальная реконструкция и визуализация объектов культурного наследия рассматриваются как самостоятельное направление цифровой истории [3, с. 18]. Методика бесконтактной реконструкции объектов архитектурного наследия представлена в его работе «Виртуальная реконструкция исторического городского ландшафта: проблемы междисциплинарного синтеза и их решение» [2, с. 130–131].

лял на страницах своих уже опубликованных книг, по сути, создавая новую редакцию⁶. Не меньшего внимания заслуживают и маргиналии писателя в книгах, ставших материалом для его собственных произведений⁷. Вторым инструментом текстолога стал виртуальный архив, представляющий собой открытый каталог и фонд пользования рукописей, в любое время дня и ночи доступный для исследователя в Интернете. Как и реальный архив, электронное собрание позволяет увидеть источник в его бытовании — набросок рассказа вместе с письмом, в которое он был вложен; правку газетной статьи — в альбоме рецензий; первые варианты стихотворений вместе с посвященными им дневниковыми записями. Работая над созданием цифрового архива, ученый воссоздает историю текста во всей ее полноте. Он продвигается от нескольких строчек в записной книжке или подчеркнутой карандашом фразы к рабочим тетрадям, письмам, рисункам. Затем следуют черновые автографы, еще очень далекие от окончательного варианта, опять письма и дневники, и наконец, первый беловик, возможно — один из многих, а дальше, в зависимости от судьбы произведения, гранки, рукописная книга, при счастливом исходе — издание, авторские пометы на котором, вполне вероятно, заставят архивиста еще раз пройти весь этот путь.

Создание архивов растровых изображений рукописей (т. е. изображений, передаваемых исключительно графически, «нечитаемых» программным обеспечением любого класса) было вызвано активным развитием цифровых архивов и библиотек, содержащих распознанные данные, редактируемые в программной среде. В первую очередь это корпусные исследования⁸, к которым относятся Национальный корпус русского язы-

6 Показателен пример цифрового восстановления правки И.А. Бунина на страницах его Собрания сочинений. См.: «Работа с электронными изображениями позволяет не только фиксировать процесс создания текста, но и восстановить, казалось бы, полностью утраченные строки. <...> Электронные копии, созданные с разрешением от 300 до 800 TIFF, позволяют масштабировать изображение, а их последующая обработка позволяет выполнить электронное ретуширование текста (путем последовательного наложения друг на друга полностью идентичных копий) и восстановить следы осыпающегося карандаша, “проявляя” неразборчивые или стертые временем фрагменты» [8].

7 В этом отношении крайне интересен входящий в личную библиотеку М.А. Булгакова русский прозаический перевод (подстрочник) «Фауста» Гёте А. Соколовского [17]. В настоящее время ведутся переговоры с руководством РГБ о размещении полной цифровой копии этого издания на портале «Автограф. XX век».

8 Среди зарубежных корпусных исследований ведущую роль занимает Manuscripts Special Interest Group, развивающая получившее популярность на рубеже 1970–1980-х гг. се-

ка [23] и Фундаментальная электронная библиотека «Русская литература и фольклор» [31]⁹. Междисциплинарный корпус Фундаментальной электронной библиотеки (ФЭБ) можно рассматривать как связующее звено между собраниями распознанных текстов, во множестве представленными в Интернете, и цифровыми архивами последнего десятилетия — массивами растровых изображений. Тотальная фиксация источников, характерная для современных фольклорных собраний, объединяет в цифровом пространстве картографию экспедиций, фотографии, аудио- и видеофайлы¹⁰. Необходимо отметить, что постоянное обновление технических средств каталогизации и визуализации источников заложена в специфике фольклорных коллекций. Еще в 1960-х гг. В.Я. Пропп в «Методической записке по архивному хранению и систематизации фольклорных материалов» назвал архив, лишенный современного справочного аппарата, «мертвой грудой материалов»¹¹. При этом несколько глобальных изменений формата цифровых и магнитных источников произошли только за последние десять лет [13, с. 11].

мантическое кодирование текстов. Эта система стала основой широко известной программы семантической классификации текста TEI (Text Encoding Initiative, 2011) [35]. Размещенные в программе тексты проходят расшифровку, обработку и первичное кодирование, а затем вносятся в программу. Растровые рукописные массивы система TEI автоматизированно не обрабатывает.

9 Сходный принцип систематизации данных использован и в цифровых корпусах текста, создаваемых для локальных научных исследований отдельными группами ученых. В качестве одного из наиболее удачных примеров локального корпуса можно привести систему обработки рукописей «Манускрипт» (Ижевск) [20]. «Манускрипт» представляет собой собрание текстов (в данном случае — памятников древнерусской письменности) и систему электронной разметки, с помощью которой можно формировать поисковые запросы.

10 Речь идет о форматах WMA, MP3, AVI, MOV, MPEG4, miniDV, Betacam, DvCam, miniDVD. Более подробно о методике тотальной фиксации фольклорных материалов см. в публикациях о создании Экспедиционного мультимедийного программного комплекса: «Технические средства для фиксации аудио и визуальных компонентов традиционных культур используются исследователями практически с момента их появления: фотографии с середины XIX в., звукозаписи и кино — с начала XX-го. Появление цифровых аудио, видео и фото форматов коренным образом изменили процесс экспедиционной работы фольклористов и этнологов. Они позволяют фиксировать не просто “образцы” народной культуры, по необходимости экономя на магнитофонной, кино или фотоленке, как это было еще совсем недавно — буквально до 1990-х. Цифровая аппаратура дает возможность проводить что называется тотальную фиксацию — “бытовые” разговоры с исполнителями и носителями традиций, многочасовые видеозаписи обрядовых актов и т. д.» [6, с. 84–85].

11 Более точно: «При таком размахе, который приобрела в СССР собирательская работа, правильное хранение собранных материалов становится делом государственной важности. Архив без надлежащих указателей и каталогов — это мертвая гряда материалов» [21, с. 5].

Становясь частью растрового архива, источник приобретает новое качество — мобильность. Независимо от научной квалификации исследователя и состояния оригинала рукописи любой пользователь Интернета получает к нему прямой доступ: может его изучать, комментировать, цитировать в своих работах и ссылаться на сам источник, а не только на публикацию в собрании сочинений¹². Развитие технологий обусловило практически неограниченный рост цифровых архивов. С созданием дата-центров публикация тысяч и десятков тысяч цифровых изображений не представляет никаких технических затруднений¹³. Тем не менее с постоянным накоплением цифры органически связаны и слабые стороны электронной публикации. Ориентируясь на сложившиеся стандарты академического исследования¹⁴, цифровой архив просто физически не может подготовить полноценный научно-справочный аппарат для стремительно растущего объема источников. Постепенно, независимо от позиции его создателей, виртуальный архив ставит вопрос об альтернативных стандартах публикации. Авторское исследование информационного массива (вступительная статья, научная передача

12 В качестве оптимального источника возможно рассматривать предложенный Н.А. Богомоловым и В.Л. Гайдук «интегрированный комментированный текст», находящийся в непосредственной связи с опубликованной в Интернете цифровой копией рукописного автографа [1, с. 335].

13 В 2010–2013 гг. в постоянно функционировавших цифровых архивах «Объединенный электронный архив И.А. Бунина» [26], «Виртуальный архив Анны Ахматовой» [16] и «Объединенный архив Вяч. Иванова» [24] ежегодный рост информационного массива составлял около 1000 электронных изображений. В 2014 г. начал работать Цифровой архив русской литературы «Автограф. XX век» [33]. В 2014–2021 гг. его экспонентный рост составил 4 000 изображений. В 2022 г. ежегодное пополняемое количество изображений было увеличено до 5 000. Готовится к открытию портал «“Стенограмма”: Политика и литература. Цифровой архив литературных организаций 1920-х – 1930-х годов» (ИМЛИ РАН) [30], электронный массив которого составляет более 15 000 изображений.

14 Лучше всего этот тип научного комментария охарактеризовал А.В. Лавров в одной из своих недавних публикаций: «Архивную единицу можно было ввести в оборот без солидного идеологического сопровождения — просто охарактеризовав конкретную ситуацию, конкретное лицо, персонажа. Вот тогда у нас и появился новый тип комментария — развернутого, с большими дополнительными этажами материалов. Первым его освоил Гарик Суперфин еще до своей посадки: его стиль работы — к письму в четыре строки давать комментарий в четыре страницы. Образец такой работы — письма Ахматовой к Брюсову (совместно с Романом Тименчиком), где писем несколько строчек, а работа на 20 страниц» [19]. Или более академично: «В течение последнего столетия были осуществлены или продолжают осуществляться под эгидой Академии наук несколько десятков собраний сочинений русских классиков. За это время установлены и отработаны в ходе практической деятельности базовые текстологические принципы, которым призваны удовлетворять издания академического типа» [7, с. 7].

ча текста, создание реального, текстологического и контекстного комментария) функционально меняется на разработку новой модели управления информацией. Возможно, текущим решением дихотомии «буквы и цифры» может стать возвратная публикация на бумажном носителе, подразумевающая подготовку тематического издания на материалах цифрового архива¹⁵. В возникшем таким образом конгломерате традиции и новации электронный архив обеспечит необходимую полноту источниковой базы, а сборник научных материалов — академическую подготовку.

* * *

Технологически цифровое собрание представляет собой нераспознанный корпус растровых изображений, состоящий из десятков тысяч объектов — электронных форм архивных документов. Понимаемый как единый массив, виртуальный архив идеально подходит под уже ставшее классическим определение «больших данных»: структурируемые и неструктурированные данные большого объема и значительного разнообразия. В качестве определяющих характеристик для массивов big data традиционно выделяют «три V»: объем (volume), скорость приращения (velocity) и многообразие (variety). Действительно, и накопленный объем цифровых изображений, и темпы роста виртуальных архивов полностью укладываются в предложенный формат. Третий и важнейший критерий — разнообразие и унификация частного — отсылает к «Структуре научных революций» Томаса Куна: числа, собранные при отсутствии каких-либо ожидаемых закономерностей, почти наверняка останутся просто числами. В поисках закономерности, которая смогла бы превратить принципиально нерешаемые задачи исследования в «паззлы» (puzzle), доступные парадигме «нормальной науки», оправдано обращение к стандартам описания архивных документов. Разработанный в 1994 г. Основной международный

¹⁵ Подобное решение дихотомии цифрового архива было предложено в выступлениях Н.В. Корниенко, Д.С. Московской и М.Л. Спивак на Международной конференции «Литературный цифровой архив как культурная практика и социальный опыт», проходившей 20 октября 2022 г. в ИМЛИ РАН. Заслуживает внимания то, что по пути синтеза «буквы и цифры» развиваются и новейшие разработки Пушкинского Дома. В качестве примера можно привести создание в 2018 г. электронного ресурса «Объединенный цифровой архив рукописей Ф.М. Достоевского» [25], а также опубликованные работы К.А. Баршта и его научной группы [18].

стандарт архивного описания (ISAD, Оттава) [34] за два последующих десятилетия получил несколько редакций, в последнюю из которых были включены цифровые копии. International Standard Archival Description предусматривает выделение общих признаков по шести основным (страна, хранилище, шифр¹⁶, подлинность, авторство, датировка) и 26 дополнительным параметрам, варьируемым в каждой стране или даже административном округе. За исключением фондирования, включенные в международный стандарт информационные поля соответствуют архитектонике рукописного документа. В беловике и черновике, письме и дневнике мы можем выделить информационное поле даты, заголовка, подписи, рисунка и корпуса документа, то есть собственно текста. Эти поля, универсальные для любой рукописи, могут стать стабильными ориентирами, на которых базируется управление любым (и здесь уже не столь важно, распознанным или растровым) информационным массивом.

Публикация растрового источника как электронного гипертекста становится частью цифровой трансформации филологии; уже сейчас возможно говорить как минимум о нескольких виртуальных архивах, выстраивающих систему гиперссылок, предоставляющих исследователю возможность самостоятельно моделировать состав и размер информационного массива. Среди них «Объединенный цифровой архив рукописей Ф.М. Достоевского» (ИРЛИ РАН) и ««Стенограмма»: Политика и литература. Цифровой архив литературных организаций 1920-х – 1930-х годов» (ИМЛИ РАН). С некоторой оговоркой возможно в этом ряду говорить и о проекте ИРЛИ РАН «А.С. Пушкин: Электронное академическое издание. Тексты произведений» [15]; при этом необходимо отметить, что Pushkin Digital в первую очередь является уникальным цифровым научным изданием в гипертекстовом формате, и возможность использовать черновики Пушкина как растровый архив [32] — результат вторичный.

Если размер источника уже не единица хранения, а гипертекст¹⁷, меняется стратегия его исследования. Единица хранения требует пристально-

16 Именно здесь заложен основополагающий принцип фондирования: от общего (фонда) к частному (единице хранения и листу).

17 Обращения к литературе как к гипертексту И.А. Пильщиков связывает не только с работой Франко Моретти «Дальнее чтение», но и с работами русской формальной школы и близких к ней ученых (Б.В. Томашевский, Б.И. Ярхо), указанными им в качестве источника исследований Моретти. В статье «Русский квантитативный формализм 1910-х – 1930-х гг.

го текстологического изучения, «медленного чтения», академической публикации. Работа с гипертекстом преследует иные цели. Корпус структурно различных текстов предназначен для быстрого, или «дистанцированного»¹⁸ чтения, главной задачей которого становится поиск нужной информации в разнохарактерном множестве: «Когда мы просто читаем и интерпретируем тексты, то сколько текстов мы можем изучить? Один текст, десять текстов, пусть сто текстов — ну в крайнем случае тысячу, а ведь их гораздо больше. Как пишет Моретти, есть семь тысяч викторианских романов, которые никто никогда не прочтет, — но мы до сих пор читаем Теккерея или Диккенса и хотим знать, в чем их специфика. <...> Что делать? Выход — изучать эти семь тысяч романов дистантно, выявив некоторые формализуемые признаки, по которым уже сейчас компьютер может анализировать большие текстовые объемы (это и есть быстрое, “дистанцированное” чтение, *distant reading*, противопоставленное медленному, пристальному чтению — *close reading*)» [27].

Наиболее близко к управлению «большой цифрой» подходит проект «“Стенограмма”: Политика и литература». В проекте задействован нераспознанный массив официальных документов — машинописных отчетов о съездах, конференциях, заседаниях литературных институций¹⁹, протоко-

как предшественник *Digital Humanities*» Пильщиков отмечал «Удивительный факт: русские первопроходцы количественной поэтики и статистического стиховедения — Андрей Белый, Борис Исаакович Ярхо и Борис Викторович Томашевский — начали и долгое время вели свою работу независимо друг от друга» [11, с. 130].

¹⁸ Терминологическая модель, предложенная Франко Моретти. Показательно, что для характеристики работ Моретти Т.Д. Венедиктова заимствует термины из «Структуры научных революций» Томаса Куна: «Здесь перед нами — “нормальный” (в куновском смысле) литературовед-зарубежник, автор диссертации и монографии о творчестве Т.С. Элиота, — уже готовый, впрочем, к рискованной творческой эскападе» [4, с. 82]. О применении *distant reading* в библиотечном деле см. в работе В.П. Леонова: «Ф. Моретти, начав с поисков точного метода, приходит не к большим данным, с чем литературоведу и библиографу еще нужно научиться иметь дело, а к методологическому повороту в самом литературоведении и, полагаю, библиографоведении. На этом повороте мы получаем объекты и данные нового типа, надстроенные над языком пристального чтения, которые нужно не понимать, но на которые нужно смотреть. <...> Не исключаю также, что взаимодействие пристального, дальнего и масштабированного чтения в рамках цифровой и автоматизированной обработки текстов будет способствовать формированию нового исследовательского объекта — точного библиографоведения» [9, с. 65].

¹⁹ Московской ассоциации пролетарских писателей, Всероссийской ассоциации пролетарских писателей, Всесоюзного объединения ассоциаций пролетарских писателей, Российской ассоциации пролетарских писателей (1925–1932).

лов и анкет их участников, учетных листков и финансовых ведомостей — в совокупности насчитывающих десятки тысяч страниц. Ранее эти материалы не публиковались²⁰, поэтому работу с архивным корпусом проект начинает практически с чистого листа: структуры, хроники, комментария, аннотированного указателя.

Информационный массив выстроен по модели интерактивного древа — ветвящейся сети источников, которую исследователь формирует сам. От конкретной стенограммы он может перейти к персоналии и всему комплексу связанных с ней документов (а в перспективе — анкет, справок и отчетов); в равной степени исследователь может задействовать поиск по организациям и сосредоточить внимание на источниках, посвященных институциональной политике. Предусмотрена и еще одна возможность — компаративное направление, формирующие источниковый массив по ключевым словам. С помощью гибкого управления большими данными комплекс из десятков тысяч растровых страниц редуцируется до нескольких десятков отдельных страниц, непосредственно затрагивающих проблематику исследования. Остается один шаг до интеллектуального анализа данных: если унифицировать ключевые слова и сделать интерактивным поисковый аппарат, то сжатие информационного массива от десятков тысяч возможно нужных страниц до нескольких десятков совершенно необходимых можно будет выполнить автоматически. Технологии искусственного интеллекта могут быть задействованы для решения исследовательских задач и при полнотекстовом поиске. В этом случае с помощью программного модуля будет сформирован растровый паттерн искомой лексемы (графическая имитация машинописи), а затем автоматизированно сопоставлен с множеством словоформ полного машинописного корпуса стенограмм и протоколов.

«Новая оптика» цифрового архива устраняет кажущиеся противоречия между академической публикацией XX в. и нераспознанными мас-

20 Более подробно: «До сих пор эти архивные документы — стенограммы профессиональной производственной деятельности самых влиятельных для первых пореволюционных десятилетий пролетарских писательских организаций не подвергались фронтальной разработке, не вводились в научный оборот, не становились самостоятельным предметом системного изучения и научной публикации. Фонды этих **писательских союзов** хранятся в архивном собрании ИМЛИ РАН, состав их уникален и не дублируется материалами других архивохранилищ, журнальной и газетной периодикой и представляет собой бесценный источниковедческий ресурс, основополагающий для восстановления хроники литературной повседневности» [30].

сивами данных современности. Цифровая революция превратила единицу хранения в корпус текстов. Работа с «большими данными» требует современных инструментов — последовательного использования «дистанционного» и «пристального чтения». Эффективная организация работы с растровым массивом для управления дистанционным чтением становится главной задачей цифрового архива, вдумчивая научная интерпретация отобранной информации остается прерогативой пользователя.

Список литературы

Исследования

- 1 Богомолов Н.А., Гайдук В.Л. Валерий Брюсов. Дневник 1890 год (Предисловие Н.А. Богомоллова, подготовка текста и примечания В.Л. Гайдук и Н.А. Богомоллова) // *Studia Litterarum*. 2020. Т. 5, № 3. С. 328–357. DOI: <https://doi.org/10.22455/2500-4247-2020-5-3-328-357>
- 2 Бородкин Л.И. Виртуальная реконструкция исторического городского ландшафта: проблемы междисциплинарного синтеза и их решение // Историко-культурное наследие в цифровом измерении: материалы Международ. науч. конф. (Пермь, 20–22 октября 2021 г.). Пермь, 2021. 210 с. URL: <http://www.psu.ru/files/docs/science/books/sborniki/istoriko-kulturnoe-nasledie-v-cifrovom-izmerenii.pdf> (дата обращения: 30.03.2023).
- 3 Бородкин Л.И. Digital history: Применение цифровых медиа в сохранении культурного наследия? // Информационный бюллетень Методологические проблемы исторической информатики. 2012. Т. 1, № 1. С. 14–21.
- 4 Венедиктова Т.Д. Хитроумный путешественник // Новое литературное обозрение. 2018. № 2. С. 82–88.
- 5 Жиров М.С., Жирова О.Я., Кузнецова Н.С. Электронный фольклорный архив в современном социокультурном пространстве // Самарский научный вестник. 2021. Т. 10, № 1. С. 283–286.
- 6 Кляус В.Л., Воронков С.О., Семенов А.А. Проект создания «экспедиционного мультимедийного программного комплекса» // Мультимедийные и цифровые технологии в собирании, сохранении и изучении фольклора. Материалы междунар. науч. конф. (16–18 ноября 2011 г., Москва) / сост. В.Л. Кляус, Е.В. Миненок; под ред. В.М. Гацака. М.: ИМЛИ РАН, 2012. 252 с.
- 7 Лавров А.В. Тексты и комментарии. Из материалов к истории русской литературы первой трети XX века. СПб.: Пушкинский Дом, 2018. 528 с.
- 8 Ларин А.А., Хачатурян Л.В. Электронная реконструкция творческого архива: экономические и правовые аспекты. Опыт РГАЛИ и Архива русской эмиграции университета г. Лидса // Объединенный электронный архив Ивана

Бунина. Электронная текстология. URL: <http://www.bunin-rgali.ru/index.php?view=textualism&t=text1> (дата обращения: 30.03.2023).

- 9 *Леонов В.П.* Дальнее чтение как стратегия точного библиографоведения // Научные и технические библиотеки. 2019. № 10. С. 56–67.
- 10 *Моретти Ф.* Дальнее чтение / пер. с англ. А. Вдовина, О. Собчука, А. Шели; науч. ред. пер. И. Кушнарева. М.: Изд-во Ин-та Гайдара, 2016. 352 с.
- 11 *Пильщиков И.А.* Семь бесед о филологии и Digital Humanities. Интервью и дискуссии (2015–2021). М.: Изд-во Московского ун-та, 2022. 192 с.
- 12 *Спиридонова Л.А.* Текстология: теория и практика. М.: ИМЛИ РАН, 2019. 256 с.
- 13 *Черных А.В.* Фольклорный архив. Пермский край: исследовательский и издательский проект // Славянская традиционная культура и современный человек. Слово. Время. Человек / сост. В.Е. Добровольская, А.Б. Ипполитова, Т.М. Санникова, А.В. Черных. СПб.: Маматов, 2021. 416 с.
- 14 *Cohen M.* The Sentimental Education of the Novel. Princeton: Princeton University Press, 1999. 219 p.

Источники

- 15 Александр Сергеевич Пушкин. Электронное академическое издание. Тексты произведений. URL: <https://pushkin-digital.ru> (дата обращения: 30.03.2023).
- 16 Виртуальный архив Анны Ахматовой. URL: <http://www.akhmatova-rgali.ru> (дата обращения: 30.03.2023).
- 17 *Гете И.-В.* Фауст. Прозаический перевод с пометами М.А. Булгакова. СПб., 1902. 192 с. // НИОП РГБ. Ф. 562. Карт. 69. Ед. хр. II. 194 л.
- 18 Записные тетради Ф.М. Достоевского 1869–1872 гг. к роману «Бесы»: дипломатическая транскрипция. СПб.: Наука, 2021. 581 с.
- 19 *Лавров А.[В.]* «У нас все — целина: куда ни копни, все впервые» // Arzamas. 2022. 3 августа. URL: <https://arzamas.academy/mag/1108-lavrov> (дата обращения: 30.03.2023).
- 20 Манускрипт. Славянское письменное наследие. URL: <http://mns.udsu.ru> (дата обращения: 30.03.2023).
- 21 Методическая записка по архивному хранению и систематизации фольклорных материалов / отв. ред. проф. В.Я. Пропп. Вильнюс, 1964. 105 с.
- 22 Методические рекомендации по электронному копированию архивных документов и управлению полученным информационным массивом. М.: ВНИИДАД, 2012. 125 с. URL: https://archives.gov.ru/documents/rekomend_el-copy-archival-documents.shtml (дата обращения: 30.03.2023).
- 23 Национальный корпус русского языка. URL: <https://ruscorpora.ru> (дата обращения: 30.03.2023).
- 24 Объединенный архив Вяч. Иванова. URL: <http://www.ivanov-rgali.ru> (дата обращения: 30.03.2023).

- 25 Объединенный цифровой архив рукописей Ф.М. Достоевского.
URL: <https://dostoevskyarchive.pushdom.ru/about> (дата обращения: 30.03.2023).
- 26 Объединенный электронный архив Ивана Бунина. URL: <http://www.bunin-rgali.ru>
(дата обращения: 30.03.2023).
- 27 *Пильщиков И.* У цифровой филологии большое будущее // Arzamas.
URL: <https://arzamas.academy/materials/1161> (дата обращения: 30.03.2023).
- 28 Собрание П.Н. и С.П. Лукницких // РО ИРЛИ РАН. Ф. 754. Оп. 1.
- 29 Собрание П.Н. и С.П. Лукницких // Портал «Автограф. XX век».
URL: <http://gumilev.literature-archive.ru/ru/digital-archive/stihotvoreniya-i-poemy>
(дата обращения: 30.03.2023).
- 30 «Стенограмма»: Политика и литература. Цифровой архив литературных организаций 1920–1930 гг. URL: <http://stenogramma.imli.ru> (дата обращения: 30.03.2023).
- 31 Фундаментальная электронная библиотека. «Русская литература и фольклор».
URL: <http://feb-web.ru> (дата обращения: 30.03.2023).
- 32 Электронная интерактивная модель академического издания А.С. Пушкина.
URL: <https://pushkin-digital.ru/node/108> (дата обращения: 30.03.2023).
- 33 Электронный архив русской литературы «Автограф. XX век». URL: <http://literature-archive.ru> (дата обращения: 30.03.2023).
- 34 ISAD (G): Основной международный стандарт архивного описания: принят Комитетом по стандартам описания, Стокгольм, Швеция, 19–22 сентября 1999 г. / Междунар. совет архивов; гл. ред. пер.: Жабко Е.Д., 2-е изд. СПб.: Президентская библиотека им. Б.Н. Ельцина, 2011. 247 с.
- 35 Text Encoding Initiative. URL: <https://tei-c.org> (дата обращения: 30.03.2023).

References

- 1 Bogomolov, N.A., and V.L. Gaiduk. “Valerii Briusov. Dnevnik 1890 god (Predislovie N.A. Bogomolova, podgotovka teksta i primechaniia V.L. Gaiduk i N.A. Bogomolova)” [“Valery Brusov. Journal Entries of 1890 (Foreword by N.A. Bogomolov, preparation of the text and notes by V.L. Gayduk and N.A. Bogomolov)”]. *Studia Litterarum*, vol. 5, no. 3, 2020, pp. 328–357. <https://doi.org/10.22455/2500-4247-2020-5-3-328-357> (In Russ.)
- 2 Borodkin, L.I. “Virtual’naia rekonstruktsiia istoricheskogo gorodskogo landshafta: problemy mezhdistsiplinnogo sinteza i ikh reshenie” [“Virtual Reconstruction of the Historical Urban Landscape: Problems of Interdisciplinary Synthesis and Their Solution”]. *Istoriko-kul’turnoe nasledie v tsifrovom izmerenii: materialy Mezhdunarodnoi nauchnoi konferentsii (Perm’, 20–22 oktiabria 2021 g.)* [Historical and Cultural Heritage in the Digital Dimension: Materials of the International Scientific Conference (Perm, October 20–22, 2021)]. Perm, 2021. 210 p. Available at: <http://www.psu.ru/files/>

- docs/science/books/sborniki/istoriko-kulturnoe-nasledie-v-cifrovom-izmerenii.pdf (Accessed 12 January 2023). (In Russ.)
- 3 Borodkin, L.I. "Digital history: Primenenie tsifrovyykh media v sokhranении kul'turnogo naslediiia?" ["The Use of Digital Media in the Preservation of Cultural Heritage?"]. *Informatsionnyi biulleten' Metodologicheskie problemy istoricheskoi informatiki*, vol. 1, no. 1, 2012, pp. 14–21. (In Russ.)
- 4 Venediktova, T.D. "Khitroumnyi puteshestvennik" ["The Clever Traveler"]. *Novoe literaturnoe obozrenie*, no. 2, 2018, pp. 82–88. (In Russ.)
- 5 Zhirov, M.S., and O.Ia. Zhirova, and N.S. Kuznetsova. "Elektronnyi fol'klorny arkhiv v sovremennom sotsiokul'turnom prostranstve" ["Electronic Folklore Archive in the Modern Socio-cultural Space"]. *Samarskii nauchnyi vestnik*, vol. 10, no. 1, 2021, pp. 283–286. (In Russ.)
- 6 Kliaus, V.L., and S.O. Voronkov, and A.A. Semenov. "Proekt sozdaniia 'Ekspeditsionnogo mul'timediinogo programmnoogo kompleksa'." ["The Project of Creating an 'Expeditionary Multimedia Software Complex'."]. *Mul'timediinye i tsifrovye tekhnologii v sobiranii, sokhranении i izuchenii fol'klora. Materialy mezhdunarodnoi nauchnoi konferentsii (16–18 noiabria 2011 g., Moskva)* [Multimedia and Digital Technologies in Collecting, Preserving and Studying Folklore. Proceedings of the International Scientific Conference (November 16–18, 2011, Moscow)]. Moscow, IWL RAS Publ., 2012, pp. 84–94. (In Russ.)
- 7 Lavrov, A.V. *Teksty i kommentarii. Iz materialov k istorii russkoi literatury pervoi treti XX veka* [Texts and Comments. From Materials on the History of Russian Literature of the First Third of the 20th Century]. St. Petersburg, Pushkin House Publ., 2018. 528 p. (In Russ.)
- 8 Larin, A.A., and L.V. Khachaturian. "Elektronnaia rekonstruktsiia tvorcheskogo arkhiva: ekonomicheskie i pravovye aspekty. Opyt RGALI i Arkhiva russkoi emigratsii universiteta g. Lidsa" ["Electronic Reconstruction of the Creative Archive: Economic and Legal Aspects. The Experience of the RGALI and the Archive of Russian Emigration of the University of Leeds"]. *Ob"edinennyi elektronnyi arkhiv Ivana Bunina. Elektronnaia tekstologiya. Elektronnyi resurs* [Ivan Bunin's United Electronic Archive. Electronic Textology. Electronic Resource]. Available at: <http://www.bunin-rgali.ru/index.php?view=textualism&t=text1> (Accessed 12 January 2023). (In Russ.)
- 9 Leonov, V.P. "Dal'nee chtenie kak strategiiia tochnogo bibliografovedeniia" ["Distant Reading as a Strategy of Accurate Bibliography"]. *Nauchnye i tekhnicheskie biblioteki*, no. 10, 2019, pp. 56–67. (In Russ.)
- 10 Moretti, F. *Dal'nee chtenie* [Distant Reading], trans. from English by A. Vdovin, O. Sobchuk, A. Sheli, scientific ed. of trans. by I. Kushnarev. Moscow, The Gaidar Institute Publ., 2016. 352 p. (In Russ.)

- 11 Pilshchikov, I.A. *Sem' besed o filologii i Digital Humanities. Interv'iu i diskussii (2015–2021)* [Seven Conversations about Philology and Digital Humanities. Interviews and Discussions (2015–2021)]. Moscow, Moscow University Publ., 2022. 192 p. (In Russ.)
- 12 Spiridonova, L.A. *Tekstologiya: teoriia i praktika* [Textology: Theory and Practice]. Moscow, IWL RAS Publ., 2019. 256 p. (In Russ.)
- 13 Chernykh, A.V. "Fol'klorni arkhiv. Permskii krai: issledovatel'skii i izdatel'skii proekt" ["Folklore Archive. Perm Krai: Research and Publishing Project"]. *Slavianskaia traditsionnaia kul'tura i sovremennyi chelovek. Slovo. Vremia. Chelovek* [Slavic Traditional Culture and Modern Man. Word. Time. Person], comp. V.E. Dobrovol'skaia, A.B. Ippolitova, T.M. Sannikova, A.V. Chernykh. St. Petersburg, Mamatov Publ., 2021, pp. 10–21. (In Russ.)
- 14 Cohen, Margaret. *The Sentimental Education of the Novel*. Princeton, Princeton University Press, 1999. 219 p. (In English)