

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное  
образовательное учреждение высшего образования  
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

Н. Г. Поврозник

# ВЕБ-ИСТОРИЯ СОЦИУМА И ЕГО ИНСТИТУТОВ

*Допущено методическим советом  
Пермского государственного национального  
исследовательского университета в качестве  
учебного пособия для студентов, обучающихся  
по направлению подготовки магистров  
«История»*



Пермь 2022

УДК 930.25: 004.7(075.8)

ББК 63.2л63+32.81

П427

**Поврозник Н. Г.**

П427 Веб-история социума и его институтов : учебное пособие / Н. Г. Поврозник ; Пермский государственный национальный исследовательский университет. – Пермь, 2022. – 156 с.

ISBN 978-5-7944-3815-4

Учебное пособие является первым в России, призванным представить основы работы с веб-архивами при проведении исторических исследований. Предназначено для углубленного изучения веб-архивов как исторических источников и возможностей их использования в исследованиях. В главах книги показана специфика веб-истории как междисциплинарного исследовательского поля, описан процесс веб-архивирования, продемонстрировано влияние веб-архивов на складывание исторических источников нового типа, представлен краткий обзор исторических исследований, проведенных на основе использования ресурсов веб-архивов, рассмотрен инструментарий и методы проведения исследования в области веб-истории.

Учебное пособие предназначено студентам исторических специальностей, исследователям, изучающим социальную, культурную, экономическую и политическую историю современности, а также историю информационных технологий, сети Интернет и Всемирной паутины. Также пособие будет полезно студентам и специалистам в области социальных и гуманитарных наук, использующим ресурсы Интернета и веб-архивов в профессиональной деятельности.

УДК 930.25: 004.7(075.8)

ББК 63.2л63+32.81

*Издается по решению ученого совета историко-политологического факультета  
Пермского государственного национального исследовательского университета*

*Рецензенты:* кафедра истории России и зарубежных стран Челябинского государственного университета (зав. кафедрой – д-р ист. наук, доцент **С. А. Баканов**); доцент кафедры исторической информатики МГУ им. М. В. Ломоносова, канд. ист. наук, доцент **А. Ю. Володин**; доцент кафедры информационных технологий в гуманитарных исследованиях Института международных отношений и мировой истории Национального исследовательского Нижегородского государственного университета им. Н. И. Лобачевского, канд. ист. наук, доцент **А. А. Акашева**

ISBN 978-5-7944-3815-4

© ПГНИУ, 2022

© Поврозник Н. Г., 2022

## ОГЛАВЛЕНИЕ

<b>Введение .....</b>	<b>6</b>
<b>Глава 1. Веб-история как дисциплина .....</b>	<b>10</b>
1.1. Информационное общество и его история.....	10
1.2. Междисциплинарный характер веб-истории .....	12
1.3. Зачем сохранять веб?.....	14
1.4. Всемирная паутина как исторический источник .....	17
Рекомендуемая литература.....	21
Источники.....	22
Вопросы для самопроверки .....	23
<b>Глава 2. Веб-архивирование как процесс:</b>	
<b>Проекты, уровни, стратегии .....</b>	<b>24</b>
2.1. Архивирование веб-ресурсов .....	24
2.2. Веб-архив как информационный ресурс.....	25
2.3. Обзор веб-архивных инициатив.....	27
2.4. Уровни архивирования веб-ресурсов .....	29
2.5. Стратегии отбора ресурсов для сохранения.....	31
2.5.1. Этика и правовые основания сохранения веба.....	33
2.5.2. Ценностный подход к сохранению веб-ресурсов.....	35
2.5.3. Доменное сканирование.....	40
2.5.4. Временные рамки как характеристика отбора ресурсов.....	42
Рекомендуемая литература.....	43
Источники.....	44
Вопросы для самопроверки:.....	46
<b>Глава 3. Сбор, сохранение, документирование</b>	
<b>и обеспечение доступа к данным в жизненном цикле веб-архива ...</b>	<b>47</b>
3.1. Сбор данных.....	47
3.2. Сохранение.....	49
3.3. Документирование.....	50

3.4. Обеспечение доступа.....	51
3.5. Сложности веб-архивирования и их негативное влияние на выборку .....	53
3.6. Социальные медиа и веб-архивирование: Значение сохранения, подходы и ограничения.....	59
Рекомендуемая литература.....	66
Источники.....	67
Вопросы для самопроверки.....	69
<b>Глава 4. Веб-история социума и его институтов:</b>	
<b>Обзор направлений исследований .....</b>	<b>71</b>
4.1. Социальная веб-история: Общество и общественные объединения.....	73
4.2. История государственных и общественных институтов .....	77
4.3. История культуры и эволюция способов ее репрезентации в сети ..	80
4.4. Веб-история как история развития медиaprостранства.....	86
Рекомендуемая литература.....	90
Источники.....	94
Вопросы для самопроверки .....	95
<b>Глава 5. Предмет и источниковая база исследования .....</b>	<b>97</b>
5.1. Выбор и формулировка исследовательского вопроса.....	97
5.2. Формирование источниковой базы исследования.....	100
5.3. Сервисы для формирования источниковой базы.....	104
5.3.1. Wayback Machine.....	104
5.3.2. Archive-It .....	107
5.3.3. Memento .....	109
5.3.4. Webrecorder .....	110
5.3.5. GLAM Workbench.....	110
5.3.6. Другие сервисы.....	111
5.4. Внешняя и внутренняя критика источников .....	112
Рекомендуемая литература.....	119

Источники.....	120
Вопросы для самопроверки.....	122
<b>Глава 6. Методы и инструментарий исследования.....</b>	<b>124</b>
6.1. Моделирование данных.....	126
6.2. Методы количественного анализа.....	127
6.3. Сетевой анализ данных.....	131
6.4. Методы качественного анализа.....	134
6.5. Визуализация данных.....	137
Рекомендуемая литература.....	138
Источники.....	141
Вопросы для самопроверки.....	141
<b>Заключение. Рекомендации преподавателю .....</b>	<b>144</b>
<b>Дополнительная литература и источники.....</b>	<b>152</b>
<b>Приложения .....</b>	<b>153</b>
Приложение 1. Глоссарий.....	153
Приложение 2. Веб-архивы, репозитории и сервисы с открытым доступом к веб-архивным материалам .....	154
Приложение 3. Специализированные сервисы и приложения.....	155

## ВВЕДЕНИЕ

Учебное пособие «Веб-история социума и его институтов» является первым в России, призванным представить основы работы с веб-архивами при проведении исторических исследований. Веб-история как дисциплина только начинает формироваться, определяются исследовательское поле, методология, создается источниковая база, методика использования материалов веб-архивов и их источниковедческой критики.

Одним из показателей незавершенности процессов складывания дисциплины является недостаточность применения ресурсов веб-архивов в гуманитарных исследованиях, на что обращают внимание специалисты<sup>1</sup>. Существует целый комплекс причин, препятствующих широкому использованию ресурсов веб-архивов и развитию исследований в области веб-истории. Сложность развития веб-истории как дисциплины заключается в высокой степени междисциплинарности и трансдисциплинарности. Междисциплинарное поле, связанное с веб-историей, представляет собой пересечение с изучением медиа (media studies), комплексом дисциплин о коммуникации и способах передачи информации, информационными технологиями, особенно в отношении процессов поиска информации и организации информационных систем, а также более широкими гуманитарными и социальными науками, включая историю, социологию, психологию и другие дисциплины. Н. Брюггер и Р. Шредер<sup>2</sup> указывают на проблемы с методикой использования ресурсов веб-архивов, обращая внимание на сложность критики цифровых исторических веб-ресурсов. С. Беловари отмечает в качестве одной из причин отсутствие сформированной образовательной экосистемы для веб-историков<sup>3</sup>, специализированных курсов и тренингов, которые позволили бы развить необходимые навыки работы с веб-архивами. Кроме того, использование веб-архивов в работе историка является довольно сложным процессом, даже своего рода вызовом исследователям ввиду специфики критики цифровых источников, подбора инструментария и методов работы, а также трудностями, связанными с формулировкой проблематики исследования. Исследовательский

---

<sup>1</sup> Winters J. (2018) Negotiating the archives of UK web space. In: The Historical Web and Digital Humanities: the Case of National Web Domains, eds. Niels Brügger and Ditte Laursen. Routledge.

<sup>2</sup> Brügger N., Schroeder R. (2017) The Web as History. Using Web Archives to Understand the Past and the Present. London: UCL Press.

<sup>3</sup> Belovari S. (2017) Historians and Web Archives. In: Archivaria 83. P. 61. URL: <https://archivaria.ca/index.php/archivaria/article/view/13600/14985> (дата обращения: 01.05.2022).

вопрос зачастую довольно затруднительно сформулировать изначально, а затем реализовать его ввиду особенностей сбора, организации и анализа источниковой базы.

Стимулировать исследовательский интерес и расширить применение ресурсов веб-архивов в гуманитарных исследованиях способны разработка и внедрение в академические программы для историков специализированных дисциплин, направленных на использование в профессиональных целях источников этого типа. Развитие образовательных курсов может существенно продвинуть формирование веб-истории как дисциплины.

Данный учебный курс «Веб-история социума и его институтов» разработан для магистерской программы по направлению «История» и предназначен для углубленного изучения веб-архивов как исторических источников и возможностей их использования в исследованиях.

Цель учебного пособия – показать возможности работы с веб-архивами при проведении исторических исследований, предложить краткий обзор исторических исследований, проведенных на основе использования ресурсов веб-архивов, а также рассмотреть инструментарий и методы проведения исследования в области веб-истории. В учебном пособии главы, посвященные разным теоретическим аспектам, содержат также и практическое наполнение – представлены обзоры инструментов и сервисов работы с веб-архивами.

Курс предназначен для развития необходимых компетенций специалистов-историков, стимулирования их профессионального интереса к современной истории, истории веб-пространства, веб-архивным материалам. Задачами курса также являются: разработка методики анализа информационных ресурсов веб-архивов для ее использования в образовательных программах, адаптация существующих методов, этапов и технологий проведения исторических исследований на основе изучения материалов веб-архивов. Наличие понятного инструментария, разбор кейсов исследований в рамках данной дисциплины для магистратуры призваны стимулировать исследовательский интерес студентов к работе с веб-архивами и будут способствовать формированию специалистов в области веб-истории.

Логика построения данного образовательного курса основана на необходимом теоретическом и практическом фундаменте, рассмотрении конкретных кейсов, используемой методологии и методики анализа веб-архивов как исторических источников.

В качестве важных итоговых компетенций, которыми должны овладеть обучающиеся данного курса, выделены следующие:

- понимание устройства веб-архивов, знание основ современной цифровой инфраструктуры информационных ресурсов разного типа;
- развитое критическое мышление, умение находить информацию в веб-архивах и проверять ее, критически анализировать источники, средства и способы сбора и организации информации;
- владение на высоком уровне методами проведения исторического исследования, умение самостоятельно формулировать исследовательские цели и задачи, подбирать адекватные методы для их достижения и использовать соответствующие инструменты;
- владение расширенным диапазоном цифровых инструментов для сбора, организации и анализа данных;
- умение интерпретировать результаты, извлекать новое знание из полученной информации.

Структура данного учебного пособия выстроена во взаимосвязи этапов освоения принципов работы с веб-архивами от знакомства до профессионального использования в деятельности историков. Для проведения исследований в области веб-истории необходимо рассмотреть особенности веб-истории как междисциплинарного исследовательского поля (см. Гл. 1), изучить основы веб-архивирования (см. Гл. 2) и влияния веб-архивов на складывание исторических источников нового типа (см. Гл. 3), проанализировать уже реализованные исследовательские проекты (см. Гл. 4), освоить специфику постановки исследовательского вопроса в области веб-истории и организации источниковой базы исследования (см. Гл. 5) и применить соответствующие методы анализа веб-архивного материала (см. Гл. 6).

Разработанный курс позволяет органично дополнить широкий круг компетенций современных специалистов в области истории, создает благоприятную основу для углубления специализации магистерских программ в направлении веб-истории, создания слоя профессионалов-историков с междисциплинарной подготовкой.

Формирование веб-истории как дисциплины требует понимания и формулирования специфики исследовательского поля, особенностей методов работы с источниками, проработки терминологии и т.д. При разработке курса и его внедрении в образовательный процесс, так или иначе,



приходится сталкиваться с этими проблемами. Постепенное решение данных вопросов, поиск ответов на них позволят сформировать специалистов, которые будут продолжать работать в направлении веб-истории, проводя исследования, ставя новые задачи, разрабатывая проблемное поле, что благоприятно скажется на формировании веб-истории как дисциплины, и в этом смысле создание курса «Веб-история социума и его институтов» может рассматриваться как вклад в основу этого процесса.

Учебное пособие предназначено студентам исторических специальностей, историкам, изучающим социальную, культурную, экономическую и политическую историю современности, а также историю информационных технологий, в том числе, Интернета и Всемирной паутины. Также пособие будет полезно студентам и специалистам в области социальных и гуманитарных наук, использующим ресурсы Интернета и веб-архивов в профессиональной деятельности.

Разработка курса «Веб-история общества и его институтов» для магистерской программы по направлению «История» поддержана грантом благотворительного фонда Владимира Потанина. Автор выражает благодарность Фонду и его команде за всестороннюю поддержку в реализации проекта.

# ГЛАВА 1. ВЕБ-ИСТОРИЯ КАК ДИСЦИПЛИНА

## 1.1. Информационное общество и его история

Современное общество как информационное общество и его историю невозможно рассматривать в отрыве от контекста цифровой революции и развития информационного пространства Интернета как среды для коммуникации, хранения и передачи информации. Ю. Дэниел ввел термин «цифровая современность»<sup>4</sup>, который позволяет связать развитие общества с технологическими инновациями, относящимися, в первую очередь, к компьютеризации, информатизации, интернетизации.

Важную роль в процессе современной коммуникации играет поиск, извлечение, обработка и передача информации, а также ее производство. До появления Всемирной паутины (World Wide Web или WWW) источники информации, которыми пользовалось общество, были традиционными и привычными для того времени. Люди использовали печатные энциклопедии вместо Википедии, тома справочников вместо баз данных и электронных сервисов в сети, кулинарные книги вместо блогов от шеф-поваров и т.д. С проникновением Интернета в повседневную жизнь человека и развитием веб-пространства принципиально меняется подход к созданию, поиску и получению информации.

Виртуальная среда стала важным фактором нашей повседневности, в разных масштабах стало менять жизнь человека, очень значительно, если не сказать принципиально. Воздействие на социум открытой, доступной и разнообразной информации, способы ее производства еще предстоит изучить, однако мы видим, что происходящие изменения кардинальным образом влияют на общество, трансформируя его. Можно привести массу примеров, когда ресурсы Всемирной паутины оказывали решающее воздействие на развитие событий в офлайне, становились важным инструментом в борьбе разных взглядов, позиций, установок, в информационных войнах, направляли и даже изменяли ход истории.

Характерной чертой современности стало не только усложнение процессов коммуникации, но и появление дополнительных информационных слоев, прежде являвшихся не столь распространенными, очевидными

---

<sup>4</sup> Danyel J. (2012) Zeitgeschichte der Informationsgesellschaft, in: Zeithistorische Forschungen/Studies in Contemporary History, Online-Ausgabe, 9 (2012), H. 2, Druckausgabe, pp. 186–211. URL: <https://zeithistorische-forschungen.de/2-2012/4441>, DOI: <https://doi.org/10.14765/zsf.dok-1598> (дата обращения: 01.05.2022).

и легко воспроизводимыми. Если сравнивать с широким распространением газет во второй половине XIX в. – позицию читателя вряд ли можно назвать активной. Несмотря на то, что в редакции поступали письма от читателей, это скорее было исключением, хотя прогрессивное значение газет для того периода не будем отрицать и сравнивать с другими факторами, как распространение грамотности. Альтернативные периодической печати способы передачи информации, как, например, радио или телевидение, также были ориентированы скорее на пассивное потребление. В информационном обществе эта созерцательная роль получателя информации трансформируется радикально. Читатель и зритель становится не просто активным, он начинает производить дополнительную информацию сам. Так электронные ресурсы начинают аккумулировать новые, ценные, уникальные сведения в виде обратной связи от пользователей, их реакций, комментариев, данных о репостингах и других действиях.

Особое значение в процессах коммуникации приобретают социальные сети (см. Гл. 3). В современной истории производство контента нарастает ускоренными темпами, и в его структуре все более заметную часть составляет информация, созданная не только авторизованными учреждениями, различными организациями и сообществами, но и индивидами. Именно реакции индивидов непосредственно влияют на развитие информационных ресурсов, определяют их положение в рейтингах, поисковой выдаче и в определенном смысле способствуют формированию информационного поля, в котором мы живем.

Развитие и распространение веб-технологий изменяют привычные паттерны поведения человека, влияют на мыслительные процессы. Все большее значение приобретают информационные и информационно-аналитические технологии, используемые для принятия решений, причем не только долговременных стратегических, но и сиюминутных. П. Р. Сمارт эту особенность современного общества назвал «разумом, расширенным посредством веба» (“the Web-Extended Mind”<sup>5</sup>), подчеркивая значение веб-технологий и ресурсов Всемирной паутины для развития когнитивных способностей человека и их использования в повседневной жизни. В этой связи многочисленные веб-ресурсы становятся ценным источником для изучения истории информационного общества.

---

<sup>5</sup> Smart P. R. (2012) the Web-Extended Mind. In: *Metaphilosophy*, Vol. 43, No. 4, Special Issue: PHILOWEB: Toward a Philosophy of the Web (July 2012). Pp. 446–463.

Веб-архивы и сохраненные в них материалы или «порции веба», в той или иной мере предназначены отразить историю информационного общества. Соответственно, от качества сохранения информации в веб-архивах, ее полноты, репрезентативности, обеспечения ее доступности и целого ряда других факторов зависит то, что исследователь будущего будет знать и понимать об обществе, в котором мы живем, о его развитии, какие представления о прошлом будут у будущих поколений.

## **1.2. Междисциплинарный характер веб-истории**

Веб-архивы хранят уникальные источники, которые позволяют реконструировать историю самого веба как процесс развития информационных ресурсов, а также служить основой для изучения истории – политической, социальной, экономической, культурной и др. – за последние четверть века, предлагая для анализа огромное количество свидетельств событий своего времени. Веб-архив позволяет раскрыть и другие важные стороны существования человека, в том числе трансформацию способов коммуникации, изменение речи и т.д. Таким образом, веб может выступать одновременно объектом исследований и важным источником по истории современности, а веб-архив является ресурсом для получения сведений по широкому полю социальных и гуманитарных дисциплин.

Специфика веб-истории как только еще формирующейся дисциплины заключается в пересечении ее исследовательского поля с другими дисциплинами социально-гуманитарного цикла, в том числе с политологией, социологией, психологией. Междисциплинарность может проявляться на основе использования одних ресурсов, общих методов, схожих исследовательских вопросов, требующих для историка знаний и компетенций из других областей для интерпретации анализируемых данных.

Веб-архивы хранят цифровые источники, работа с которыми подразумевает необходимость применения соответствующих подходов, методов и инструментария к их обработке при проведении исследований в области веб-истории. Указанное обстоятельство крепко связывает веб-историю с цифровыми гуманитарными науками и цифровой историей в частности. Источниковедческая критика как непрменный исследовательский компонент требует от историка знания основ веб-архивирования.

Веб-история развивается на пересечении дисциплин, связанных с исследованиями Интернета<sup>6</sup> (Internet Studies), что подразумевает изучение развития не только самой сети, но и социальных аспектов, к ней относящихся. В плане использования ресурсов Всемирной паутины можно выделить отдельные исследовательские направления, посвященные анализу определенных типов таких ресурсов. В первую очередь, это медийные источники. В этом отношении веб-история развивается во взаимосвязи с дисциплинами в области медиа-археологии и изучения средств массовой информации и коммуникации, направленных на поиск сведений, анализ каналов и процессов коммуникации, способов передачи информации. В этом плане важен и социальный аспект: необходимо понимать, кто производит контент, уметь выявлять их цели и задачи, а также аудитории, для которых создаваемые информационные ресурсы предназначаются. Широкое распространение и доступность веб-пространства реализовали принципиально новые возможности для разных социальных категорий граждан и сообществ создавать информацию и распространять ее. Вопросами изучения социальных процессов, связанных с производством и потреблением веб-ресурсов, занимаются социологи. Специалисты из других областей анализируют другие прикладные аспекты, например культурологического характера.

Значимым компонентом в этом меж- и трансдисциплинарном союзе являются информационные технологии, рассмотрение которых особенно важно в отношении процессов поиска и обработки информации, устройства и функционирования информационных систем. Критически важно знать не только то, как устроен поиск в информационных системах того или иного типа и назначения, но и понимать человеческий фактор: под влиянием каких обстоятельств пользователи ищут и находят информацию, какие средства являются наиболее подходящими, удобными, понятными и эффективными для ее получения, на какие аудитории рассчитаны системы, какие у них цели и потребности, из чего складывается социальная значимость производимого веб-контента и т.д. Этот человеческий фактор напрямую связан с эволюцией взаимодействия человека и компьютера (computer-human interaction) и теми возможностями, которые дает прогресс в области информационных технологий и их влияния на развитие веб-ресурсов.

---

<sup>6</sup> См. Dutton W. H. (2013) Internet Studies: The Foundations of a Transformative Field. The Oxford Handbook of Internet Studies. DOI: 10.1093/oxfordhb/9780199589074.013.0001. URL: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199589074.001.0001/oxfordhb-9780199589074-e-1> (дата обращения: 01.05.2022); Колозариди П. В. Интернет-исследования: Исторический обзор и производство знания // Социология власти. 2018. № 3. С. 69–92.

### 1.3. Зачем сохранять веб?

Веб-архивирование – это форма преднамеренного и целенаправленного сохранения веб-материалов, что подразумевает понимание значимости сохраняемых ресурсов, поскольку веб является историческим источником комплексного характера и представляет собой цифровое культурное наследие. В этом отношении веб-архивирование выступает как единственный способ сохранить цифровую часть культуры современности, хотя бы в какой-то мере, для будущих поколений.

Н. О. Финнеман предлагает подразделять цели долговременного архивирования на: связанные с необходимостью сохранить веб-ресурсы как цифровое наследие для будущих поколений, потенциальные коммерческие и исследовательские перспективы<sup>7</sup>. В нижеследующем обзоре обратим внимание и на другие причины.

Веб-архивирование решает задачи сохранения веб-ресурсов для их последующего использования в научном, образовательном, культурном, социальном, политическом и других направлениях. Представление о веб-ресурсах как о цифровом культурном наследии было официально обозначено в хартии ЮНЕСКО 2003 г. Статья 1 «Сфера охвата» указанного документа гласит:

*«Цифровое наследие состоит из уникальных ресурсов человеческих знаний и форм выражения. Оно охватывает ресурсы, относящиеся к области культуры, образования, науки и управления, а также информацию технического, правового, медицинского и иного характера, которые создаются в цифровой форме либо переводятся в цифровой формат путем преобразования существующих ресурсов на аналоговых носителях. В случае «цифрового происхождения» ресурсы существуют лишь в виде цифрового оригинала.*

*Цифровые материалы включают в себя текстовые документы, базы данных, неподвижные и движущиеся изображения, звуковые и графические материалы, программное обеспечение и веб-страницы, представленные в значительном и непрерывно увеличивающемся количестве форматов. Зачастую эти материалы фиксируются на короткий срок и тре-*

---

<sup>7</sup> Finnemann N. O. (2018) Web archives and Knowledge organization. In: Birger Hjørland (Ed.) ISKO Encyclopedia of Knowledge Organization & Knowledge Organization. URL: [https://www.isko.org/cyclo/web\\_archive](https://www.isko.org/cyclo/web_archive) (дата обращения: 01.05.2022).

*буют принятия целеустремленных мер, направленных на их создание, сохранение и управление ими.*

*Многие из этих ресурсов имеют непреходящую ценность и значимость и, таким образом, представляют собой наследие, которое необходимо сберечь и сохранить для нынешнего и будущих поколений. Такое непрерывно увеличивающееся наследие может существовать на любом языке, в любой части мира и относиться к любой сфере человеческих знаний и форм выражения»<sup>8</sup>.*

Создаваемые в архивном формате коллекции веб-контента предназначены для научно-исследовательских и образовательных целей, могут использоваться в общественно-политических проектах и решении многих других задач.

Веб-страницы и веб-сайты в их многообразии представляют собой уникальные материалы, произведенные изначально в цифровом формате. По образному сравнению И. Миллигана, если бы не было веб-архивов, мы бы так и изучали прошлое по печатным газетам, полностью упустив как исчезнувшие интернет-ресурсы, так и способы взаимодействия и коммуникации людей в сети<sup>9</sup>.

Происхождение веб-ресурсов имеет изначально цифровую природу и связано с электронной средой. Эти материалы создаются в цифровом измерении, и зачастую транзит в аналоговый формат представления невозможен без потери целостности информации и ее качества. Таким образом, сохранение в неких традиционных аналоговых форматах не представляется даже в малой степени эффективным, что создает угрозу исчезновения цифровых ресурсов при отсутствии должного разработанного подхода. Если сравнивать с традиционными материальными источниками (например, книгами на бумажных носителях), то угроза утраты цифровых ресурсов, не имеющих аналогового представления, весьма существенна. Так веб-ресурсы уже по своему формату и происхождению представляют собой уникальные источники, которые требуют разработанного инструментария для их долговременного сохранения, и эта функция выполняется, в первую очередь, веб-архивами.

---

<sup>8</sup> Хартия ЮНЕСКО о сохранении цифрового наследия. Принята 15 окт. 2003 г. URL: [https://www.un.org/ru/documents/decl\\_conv/conventions/digital\\_heritage\\_charter.shtml](https://www.un.org/ru/documents/decl_conv/conventions/digital_heritage_charter.shtml) (дата обращения: 01.05.2022).

<sup>9</sup> Milligan I. (2016). Lost in the infinite archive: The promise and pitfalls of web archives. *International Journal of Humanities and Arts Computing*, 10 (1). Pp. 78–94.

Веб-архивирование имеет важный утилитарный смысл и для библиографического использования: в веб-архивах генерируются постоянные ссылки на ресурсы, которые работают независимо от активности и статуса сайта в «живом» вебе. Даже если страница будет перемещена на другой адрес или сайт совсем прекратит свое существование, ссылки и заархивированный материал останутся доступными для пользователей. В научно-исследовательской работе такие постоянно действующие ссылки важны для цитирования и формирования актуальной библиографии. В этом отношении веб-архивы становятся базами знаний, которым можно доверять – факты «закрепляются» на хранении в веб-архивах, что подразумевает их содержание в неизменном виде, что впрочем не отменяет необходимости источниковедческой критики сохраненного контента.

Веб-ресурсы, представленные в «живом» интернет-пространстве, характеризуются уязвимостью к внешним воздействиям и цензуре, которая исторически влияла на контент медиа. В «живом» вебе ресурсы могут быть изменены в любой момент времени, отредактированы, перемещены или удалены. Размещение материалов в веб-архивах снижает возможность внешнего вмешательства в заархивированный контент, способствует сохранению важных данных, обеспечению общественного доступа к ним для последующего просмотра, изучения и анализа. По сравнению с «живым» вебom, контент современных веб-архивов менее поддается редактированию, поскольку одной из их основных функций является сохранение веба в неизменном виде. Однако в отдельных случаях информация может быть удалена (например, персональные данные или сведения в рамках «права на забывание»), но для совершения действий по редактированию веб-архивных ресурсов требуется провести определенные процедуры – апелляции, проверки, согласования, иногда даже судебные тяжбы. Если в отношении физических лиц, не имеющих отношения к государственной службе, эти процессы довольно просты, то корректирование информации в веб-архивах об официальных государственных и общественных структурах – дело намного более сложное, требующее судебных решений. Поэтому материалы в веб-архивах менее подвержены корректировкам в плане содержания и испытывают меньшее воздействие со стороны внешних сил и цензуры, в отличие от «живого веба».

Сведения о редактировании контента, произведенном после сохранения в веб-архиве, чаще всего фиксируются в метаданных, что позволяет



видеть привнесенные сведения и осуществлять аналитику этих изменений. Логирирование интервенций (т.е. создание логов, записей о корректировках в специальном журнале событий) имеет особый смысл, поскольку отследить эти изменения «вручную» является довольно трудоемкой или часто даже невыполнимой задачей для человека. В «живом» вебе изменения, обновления и дополнения информации на страницах сайта также могут быть зафиксированы во время архивирования и представлены, в частности, в статистике по сайту, как это реализовано на сервисе Wayback Machine. Кроме того, эта информация является ценной еще и по той причине, что позволяет оценить регулярность обновлений на ресурсе.

Сам факт сбора и сохранения информации для постоянного представления в веб-архивах способствует формированию ответственного отношения представителей бизнеса, корпораций, общественных объединений, политических деятелей, публичных персон к своим публикациям и высказываниям. Таким образом, веб-архивирование может, с одной стороны, фиксировать опубликованный в сети контент, а с другой – служить поводом для самоцензуры.

Обозначенные аспекты, связанные с сохранением веб-ресурсов, также непосредственно влияют и на возможности использования этих материалов в научных исследованиях и должны приниматься во внимание. Исследовательский потенциал веб-архивов существенный, поскольку они содержат в себе важную информацию о современности, представляя уникальный материал для использования в гуманитарных и социальных науках (подробнее о значении веб-архива как исторического источника см. в Гл. 2 данного пособия).

Материалы веб-архивов могут иметь существенное утилитарное назначение, являться весомыми доказательствами в судах, использоваться для восстановления контента при нарушении структуры и потере контента сайтов под действием различных факторов.

#### **1.4. Всемирная паутина как исторический источник**

Веб-ресурсы представляют собой особый тип исторического источника – это материал изначально «рожденный» цифровым (born-digital), то есть созданный изначально в электронной среде и таким образом не имеющий аналогового формата. Эта особенность цифровых источников (в противовес

оцифрованным<sup>10</sup>) делает их особенно уязвимыми, подверженными исчезновению без возможности восстановления. Сохранить веб для будущего призваны веб-архивы. Архивированные веб-ресурсы (материалы веб-архивов) относятся к перерожденным цифровым ресурсам (*reborn digital*<sup>11</sup>), которые образуются в условиях, когда оцифрованный или цифровой материал собран, сохранен и представлен заново, и могут существенно отличаться от оригинала. При такой репрезентации контента несколько изменяется его изначальное содержание, и ресурсы отражают лишь частично ту реальность, в которой появились и функционировали. Одной из таких особенностей является то, что архивированные веб-ресурсы приобретают новые черты, теряя динамичность веб-контента и приобретая статичность. Термин «*reborn digital*», по мнению Н. Брюггера, лучше отражает определенную точку своего развития. И эту особенность необходимо учитывать при проведении источниковедческой критики (см. Гл. 5).

Веб-архив может выступать как исследовательский объект или как источник знаний о другом исследовательском объекте<sup>12</sup>. Изучение веба как исторического источника формирует источниковедение веб-истории. Всемирная паутина является важным ресурсом для изучения происходящих в реальном пространстве политических, экономических событий и явлений культурной жизни. В свою очередь, веб-архив выступает средством документирования различных событий, жизни общества и его индивидов.

Источниковедение веб-ресурсов и критика цифровых источников являются неотъемлемыми элементами исследования в области современной истории, позволяют учесть внешние и содержательные особенности источника при проведении научных изысканий, а также адекватно оценить информационный потенциал источника для изучения тех или иных тем в социально-гуманитарных областях знания. Такие исследования могут основываться на тематических коллекциях, образованных на основе имеющихся материалов веб-архивов ответственными за сохранение веба

---

<sup>10</sup> См. Salmi H. *What is Digital History?* Medford: Polity Press, 2020. 130 p.; Поврозник Н. Г. Цифровая история, цифровые исторические источники и критика: Рецензируя книгу Ханну Салми “What is Digital History?” // Историческая информатика. 2021. № 1. С. 182–188. DOI: 10.7256/2585-7797.2021.1.35282. URL: [https://nbpublish.com/library\\_read\\_article.php?id=35282](https://nbpublish.com/library_read_article.php?id=35282) (дата обращения: 01.05.2022).

<sup>11</sup> Brügger N. (2012). When the present web is later the past: Web historiography, digital history and Internet studies // *Historical Social Research*, 37(4). Pp. 102–117.

<sup>12</sup> Brügger N., Schroeder R. (Eds.): *The Web as History. Using Web Archives to Understand the Past and the Present*. London: UCL Press 2017.

организациями (часто эти функции выполняют крупные национальные библиотеки). Другим вариантом создания тематических коллекций является целенаправленный сбор ресурсов при архивировании согласно стратегии сохранения, значимости и масштабов фиксируемых событий или руководствуясь другими аргументами (в т.ч. ценности ресурсов, см. Гл. 2).

Веб фиксирует события реальной, офлайновой жизни, которые могут становиться центральными темами подборок архивированного веба. Тематические коллекции фокусируются на разных вопросах политического, экономического, культурного или иного характера. При формировании коллекций ключевыми характеристиками могут являться не только событие, но и персона, сообщество, организация, тема (сформулированная широко или более узко), жанр, формат контента или тип медиаконтента.

Представление тематических коллекций реализовано на разных платформах архивного веба. Крупнейшей является платформа Archive-It<sup>13</sup>, на которой размещено более 8 тыс. подборок заархивированных веб-ресурсов, распределенных по темам, датам и организациям, которые осуществляли сохранение. Указанная платформа расширяет возможности пользователей, заинтересованных в изучении тематических подборок, и предлагает поиск по ключевым словам или фразам (возможен поиск по совпадению терминов в названии или аннотации к ресурсу), по общей тематике коллекций (например, выделены такие тематические блоки, как «Общество и Культура», «Политика и выборы», «Искусство и Гуманитарные науки») и другим характеристикам.

Тематические коллекции являются важным способом представления контента и на других веб-архивных ресурсах, как, например, глобальном Интернет-архиве<sup>14</sup>, австралийском TROVE<sup>15</sup> или британском UKWA<sup>16</sup>.

Вехи истории и знаковые события могут лежать в основе тематических коллекций веб-архивов или даже быть в центре целых программ и проектов по сохранению конкретных сегментов веба. Событийность может быть зафиксирована в ресурсах разного типа – новостных сайтах, блогерских постах и платформах, специализированных ресурсах. Осознание значимости происходящего подталкивает создавать такие коллекции од-

---

<sup>13</sup> Archive-It. URL: <https://archive-it.org> (дата обращения: 01.05.2022).

<sup>14</sup> Internet Archive. URL: <https://archive.org> (дата обращения: 01.05.2022).

<sup>15</sup> TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au> (дата обращения: 01.05.2022).

<sup>16</sup> The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk/> (дата обращения: 01.05.2022).

новременно с разворачивающимися событиями. Именно так были сформированы политически ориентированные коллекции, посвященные президентским выборам в России<sup>17</sup>, Украине<sup>18</sup> или других странах. Часть из заархивированных ресурсов приобрели особое значение, поскольку подверглись манипуляциям тех или иных политических сил и были удалены из «живого» веба. Кроме указанных событий, в фокусе внимания при формировании тематических коллекций были и другие критические события, связанные с проявлениями терроризма (например, коллекция ресурсов о событиях 9 сентября 2001 г. в США<sup>19</sup>), с катастрофами (коллекция веб-ресурсов, посвященная ядерному взрыву на реакторе в Фукусиме, ликвидации последствий аварии и социальным аспектам происшествия<sup>20</sup>), с протестными движениями (коллекция ресурсов движения против расизма<sup>21</sup>) и т.д. Политическими событиями тематические коллекции не ограничиваются, и в веб-архивах представлены коллекции, связанные с обществом, гендерной повесткой, культурой, здравоохранением, наукой и технологиями и другими областями современной жизни.

Создание тематических коллекций веб-архивов все менее характеризуется как случайностное, многие коллекции создаются с прицелом на фиксацию событий с максимальной подробностью деталей. Такие тематические коллекции способны не только передать общий смысл происходящих событий, но и настроить максимально приближенную к реальности оптику для их последующего изучения.

Тематические коллекции предназначены для документирования и сохранения подборок в области определенной темы или проблемы и могут служить в перспективе источником информации для решения определенных исследовательских целей и задач, поэтому необходимо критически оценивать и анализировать полноту выборки, определять предвзятости при ее формировании, в том числе, связанные с выражением разных точек зрения на происходящие события, и другие факторы.

---

<sup>17</sup> Presidential Elections – Russia – 2008. Archive-It. URL: <https://archive-it.org/collections/901> (дата обращения: 01.05.2022).

<sup>18</sup> Центральный державный электронный архив Украины. Internet Archive. URL: [https://web.archive.org/web/\\*/https://tsdea.archives.gov.ua](https://web.archive.org/web/*/https://tsdea.archives.gov.ua) (дата обращения: 01.05.2022).

<sup>19</sup> National Disaster and Digital Preservation: 9/11 and the Internet. Archive-It. URL: <https://archive-it.org/collections/4486> (дата обращения: 01.05.2022).

<sup>20</sup> Japan, Social movements after the Fukushima Nuclear Power Plants crisis, March 11, 2011. Archive-It. URL: <https://archive-it.org/collections/2954> (дата обращения: 01.05.2022).

<sup>21</sup> #blacklivesmatter Web Archive. Archive-It. URL: <https://archive-it.org/collections/4783> (дата обращения: 01.05.2022).

## Рекомендуемая литература

1. Гаков В. Рыцарь Интернета // Системный администратор. 2014, 9 (142). С. 80–83. URL: [https://www.elibrary.ru/download/elibrary\\_21919595\\_45798623.pdf](https://www.elibrary.ru/download/elibrary_21919595_45798623.pdf) (дата обращения: 01.05.2022).
2. Колозарики П. В. Интернет-исследования: Исторический обзор и производство знания // Социология власти. 2018, № 3. С. 69–92.
3. Миронова Г. В., Миронов А. Л. Динамика терминологии в области информационных сетей: “net” vs “web” // Филологические науки. Вопросы теории и практики (входит в перечень ВАК). Тамбов: Грамота, 2020. № 2. С. 174–178. URL: <https://www.gramota.net/materials/2/2020/2/34.html> (дата обращения: 01.05.2022).
4. Поврозник Н. Г. Цифровая история, цифровые исторические источники и критика: Рецензируя книгу Ханну Салми “What is Digital History?” // Историческая информатика. 2021. № 1. С. 182–188. DOI: 10.7256/2585-7797.2021.1.35282. URL: [https://nbpublish.com/library\\_read\\_article.php?id=35282](https://nbpublish.com/library_read_article.php?id=35282) (дата обращения: 01.05.2022).
5. Юлдашев Л. Как изучать историю интернета? Методы, подходы, идеи. М., 2020.
6. Belovari S. (2017) Historians and Web Archives. In: Archivaria 83. P. 61. URL: <https://archivaria.ca/index.php/archivaria/article/view/13600/14985> (дата обращения: 01.05.2022).
7. Brügger N., Schroeder R. (2017) The Web as History. Using Web Archives to Understand the Past and the Present. London: UCL Press.
8. Brügger N. (2012). When the present web is later the past: Web historiography, digital history and Internet studies // Historical Social Research, 37(4). Pp. 102–117.
9. Danyel J. (2012) Zeitgeschichte der Informationsgesellschaft, in: Zeithistorische Forschungen/Studies in Contemporary History, Online-Ausgabe, 9 (2012), H. 2, Druckausgabe. Pp. 186–211. URL: <https://zeithistorische-forschungen.de/2-2012/4441>, DOI: <https://doi.org/10.14765/zzf.dok-1598> (дата обращения: 01.05.2022).
10. Dutton W.H. (2013) Internet Studies: The Foundations of a Transformative Field. The Oxford Handbook of Internet Studies. DOI: 10.1093/oxfordhb/9780199589074.013.0001. URL: <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199589074.001.0001/oxfordhb-9780199589074-e-1> (дата обращения: 01.05.2022).
11. Finnemann N. O. (2018) Web archives and Knowledge organization. In: Birger Hjørland (Ed.) ISKO Encyclopedia of Knowledge Organization & Knowledge

Organization. URL: [https://www.isko.org/cyclo/web\\_archive](https://www.isko.org/cyclo/web_archive) (дата обращения: 01.05.2022).

12. Milligan I. (2016). Lost in the infinite archive: The promise and pitfalls of web archives. *International Journal of Humanities and Arts Computing*, 10 (1). Pp. 78–94.

13. Salmi H. *What is Digital History?* Medford: Polity Press, 2020. 130 p.

14. Smart P.R. (2012) the Web-Extended Mind. In: *Metaphilosophy*, Vol. 43, No. 4, Special Issue: PHILOWEB: Toward a Philosophy of the Web (July 2012). Pp. 446–463.

15. Winters J. (2018) Negotiating the archives of UK web space. In: *The Historical Web and Digital Humanities: the Case of National Web Domains*, eds. Niels Brügger and Ditte Laursen. Routledge.

### **Источники**

1. Хартия ЮНЕСКО о сохранении цифрового наследия. Принята 15 октября 2003 года. URL: [https://www.un.org/ru/documents/decl\\_conv/conventions/digital\\_heritage\\_charter.shtml](https://www.un.org/ru/documents/decl_conv/conventions/digital_heritage_charter.shtml) (дата обращения: 01.05.2022).

2. Центральный державный электронный архив Украины. Internet Archive. URL: [https://web.archive.org/web/\\*/https://tsdea.archives.gov.ua](https://web.archive.org/web/*/https://tsdea.archives.gov.ua) (дата обращения: 01.05.2022).

3. #blacklivesmatter Web Archive. Archive-It. URL: <https://archive-it.org/collections/4783>

4. Archive-It. URL: <https://archive-it.org> (дата обращения: 01.05.2022).

5. Internet Archive. URL: <https://archive.org> (дата обращения: 01.05.2022).

6. Japan, Social movements after the Fukushima Nuclear Power Plants crisis, March 11, 2011. Archive-It. URL: <https://archive-it.org/collections/2954> (дата обращения: 01.05.2022).

7. National Disaster and Digital Preservation: 9/11 and the Internet. Archive-It. URL: <https://archive-it.org/collections/4486> (дата обращения: 01.05.2022).

8. Presidential Elections – Russia – 2008. Archive-It. URL: <https://archive-it.org/collections/901> (дата обращения: 01.05.2022).

9. The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk/> (дата обращения: 01.05.2022).

10. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au> (дата обращения: 01.05.2022).

## **Вопросы для самопроверки**

1. Назовите особенности современной истории для ее изучения.
2. Объясните роль и значение информации для современного человека.
3. В чем заключается междисциплинарный характер веб-истории?
4. Назовите причины сохранения веба.
5. В чем заключается значение веба как цифрового культурного наследия?
6. Сравните «живой» веб и веб-архивные ресурсы. В чем заключается их сходство и различие?
7. Сформулируйте определение краулера. Какую функцию краулер выполняет?
8. Какие организационные особенности веб-архивов позволяют выявить информационный потенциал веба как исторического источника?
9. Назовите особенности цифровых ресурсов типа born-digital (рожденный цифровым) и reborn-digital (перерожденный в цифровом формате).

## ГЛАВА 2. ВЕБ-АРХИВИРОВАНИЕ КАК ПРОЦЕСС: ПРОЕКТЫ, УРОВНИ, СТРАТЕГИИ

### 2.1. Архивирование веб-ресурсов

В 1989 г. англичанин сэр Тимоти Джон Бернерс-Ли и бельгиец Роберт Кайо, работая в CERN (Европейская организация по ядерным исследованиям), изобрели для локального использования систему ссылок-команд, позднее ставшую известной как Всемирная паутина, или World Wide Web. В 1991 г. был создан первый сайт<sup>22</sup>, посвященный Всемирной Паутине, ее устройству и создателям, а двумя годами позже CERN разместила веб-технологию в открытом доступе и, тем самым, дал согласие на ее всеобщее использование. В этот период и начинается расширение «вселенной» веб-ресурсов. Согласно статистике Internet Live Stats, с 1991 г. и появления первого веб-сайта количество сайтов постоянно растет. В 1996 г. было зарегистрировано уже 257 601 сайтов, в 2000 г. их больше 17 млн, а в 2016 г. число веб-сайтов превысило 1 млрд<sup>23</sup>.

Первые программы по веб-архивированию появились лишь в 1996 г., а сайты, созданные до этого времени и прекратившие свое существование, могли сохраниться лишь на локальных носителях или были утрачены.

*Веб-архивирование – это процесс сбора фрагментов Всемирной паутины (WWW), сохранения коллекций в архивном формате, а также обеспечения доступа к ним для их использования.*

Интернет-архив (Internet Archive)<sup>24</sup> стал первым и наиболее масштабным проектом по архивированию веба, одними из основателей которого стали Брюстер Кейл и Брюс Джиллиат. Архивирование веба началось с коллекции веб-сайтов, посвященных кандидатам на пост Президента США в предвыборной гонке 1996 г. Вскоре стало очевидно, что сохранение только определенного списка сайтов сильно ограничит возможности последующего изучения процессов ввиду связей скачиваемых сайтов с внешними ресурсами, на которые содержалось большое количество ссылок. Так родилась идея архивировать также взаимосвязанные на основе гиперссылок сайты, что переросло в идею глобального интернет-архива.

---

<sup>22</sup> World Wide Web. URL: <http://info.cern.ch/hypertext/WWW/TheProject.html> (дата обращения: 01.05.2022).

<sup>23</sup> Internet Live Stats. URL: <https://www.internetlivestats.com/total-number-of-websites/#trend> (дата обращения: 01.05.2022).

<sup>24</sup> Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).



Интернет-архив – это некоммерческая организация, целью которой стало сохранение порций веба и обеспечение к ним доступа, и сегодня он является крупнейшим архивом цифрового контента в мире.

## **2.2. Веб-архив как информационный ресурс**

Веб-архив – это особое хранилище архивированного веба, в котором реализованы возможности доступа к сохраненным материалам.

Веб-архив можно сравнить с другими типами электронных ресурсов – в первую очередь, с электронной библиотекой и электронным архивом как хранилищами информации. Обратим внимание на особенности веб-архива, указанные Н. Брюггером при определении сущности веб-архива как информационного ресурса<sup>25</sup>. По мнению исследователя, различия, определяющие специфику информационных ресурсов, следует искать в типах публикаций, хранимых и предоставляемых обществу. Так, архив относится к учреждению, специализирующемуся на сохранении и предоставлении доступа к личным документам (среди которых дневники, фотографии, письма) или материалам учреждений и организаций (в том числе, внутренним документам – личным делам сотрудников, стенограммам заседаний и др., предназначенным для сохранения согласно законодательству). Естественно, документы, хранимые в архиве, существенно более разнообразны по своему характеру, чем описано выше, однако Н. Брюггер подчеркивает, что архив – это хранилище материалов, изначально не созданных для публичной демонстрации, и их трансфер в широкое общественное пространство происходит с существенной задержкой по времени от момента их создания ввиду понимания их ценности и определения исторической значимости. Библиотеки созданы как публичные пространства, и основной целью их функционирования является доступ к текстовому, а также аудиовизуальному и иному наследию, предназначенному для общественного использования.

Целью функционирования веб-архивов является сохранение и обеспечение доступа к ресурсам, находящимся в публичном домене, то есть открытыми для свободного просмотра пользователями. Это обстоятельство делает веб-архив более похожим на библиотеку, чем на архив. Поэтому Н. Брюггер предлагает вместо веб-архива использовать термин

---

<sup>25</sup> Brügger N. (2016) Webraries and Web archives: the Web between public and private. In W. Evans and D. Baker (Eds), *The End of Wisdom? The Future of Libraries in a Digital Age*. Oxford: Chandos. pp. 185–190.

“webrary” («веб-библиотека» может служить русскоязычным аналогом предложенного термина).

Н. Брюггер указывает на размывание границ информационных ресурсов библиотек и архивов, которое стало типичным процессом для современности. Веб-архивы также начинают сохранять информацию, не относящуюся изначально к публичному контуру информации (частной жизни граждан или внутренним вопросам организаций, например, датский веб-архив сохраняет внутреннюю информацию из локальных сетей бизнес-учреждений). Указанное обстоятельство может служить основой для разграничения веб-архивов по типу представленной информации на веб-библиотеки (“webrary”), предоставляющие доступ к вебу, сохраненному из публичного контура, и на веб-архивы, материалы которых собраны из закрытых, внутренних сетей или ограниченных по доступу. Поэтому многие проекты по веб-архивированию являются гибридными, если принимать во внимание указанное разграничение. Чаще всего к таким проектам относятся национальные веб-архивы, сохраняющие как публичный сектор веба, так и ограниченный по доступу.

Вместе с тем не всегда представляется возможным четко определить предназначение того или иного материала, представленного во Всемирной паутине, классифицировать точно – является ли он публичным или частным. Так, документы могут быть обнародованы в открытом вебе по ошибке или в нарушение авторских прав. Поэтому сложно дать определение веб-архиву как информационному ресурсу, специализирующемуся на сохранении данных открытого контура, ввиду особенностей их наполнения. Кроме того, из-за отсутствия четкой типологии веб-ресурсов и наличия широкого разнообразия сохраняемых материалов бывает сложно определить специфику того или иного веб-архива, классифицировать. Такие классификации попросту отсутствуют. Таким образом, порог избирательности у веб-архивов может быть низким для создания благоприятных условий для сохранения разного рода материалов, что впоследствии создает трудности типизации веб-архива как информационного ресурса. В итоге, это размывание границ может стать поводом для переосмысления специфики ставших традиционными ресурсов архивов и библиотек, многие функции и характеристики которых стали схожими в настоящее время.

### 2.3. Обзор веб-архивных инициатив

Проект Интернет-архив (Internet Archive) является одной из самых ранних и самой крупной инициативой по архивированию веба в глобальном масштабе. На сайте Интернет-архива сформулирована миссия существования организации, которая заключается в том, чтобы «обеспечить универсальный доступ ко всему знанию»<sup>26</sup>. Веб – не единственный тип контента, сохраняемого Интернет-архивом. В нем также аккумулируются и другие виды цифрового контента – оцифрованные и электронные книги, рукописи и другие текстовые источники, аудио- и видеозаписи, изображения, игры, наборы данных и программное обеспечение различного назначения. По сути, Интернет-архив сегодня представляет собой электронный архив веб-сайтов и других ресурсов в цифровом формате. Доступ к веб-материалам Интернет-архива осуществляется с помощью платформы Wayback Machine<sup>27</sup>, которая до сих пор остается самой известной и наиболее используемой платформой для работы с архивным вебом по всему миру.

В рамках рассматриваемого проекта был разработан и первый краулер (сканер) Heritrix<sup>28</sup>, представляющий собой программное обеспечение с открытым программным кодом, который можно использовать для качественного архивирования веба любого масштаба.

Достоинством Интернет-архива является то, что весь сохраняемый контент находится в открытом доступе в совокупности с метаданными, статистической и иной информацией. Иные веб-архивы часто имеют ограничения на предоставление онлайн-доступа к ресурсам, что может быть связано с политикой организации, производящей веб-архивирование, с ограничениями по авторскому праву или другими причинами.

На национальном уровне реализуются проекты разного масштаба, но в первую очередь, следует обратить внимание на проекты по сохранению национального веба. Одними из наиболее ранних стали проекты PANDORA Национальной библиотеки Австралии (сейчас веб-архив интегрирован в систему TROVE<sup>29</sup> и является частью национальной программы по сохранению цифрового культурного наследия, осуществляемой Национальной

---

<sup>26</sup> About the Internet Archive. URL: <https://archive.org/about/> (дата обращения: 01.05.2022).

<sup>27</sup> Wayback Machine. URL: <https://web.archive.org/> (дата обращения: 01.05.2022).

<sup>28</sup> Heritrix. Internet Archive. URL: <http://crawler.archive.org/index.html> (дата обращения: 01.05.2022).

<sup>29</sup> Trove. National Library of Australia. URL: <https://trove.nla.gov.au/> (дата обращения: 01.05.2022).

библиотекой Австралии в сотрудничестве с другими учреждениями хранения) и Kulturarw3 Национальной библиотеки Швеции<sup>30</sup>. Оба проекта были запущены еще в 1996 г. и, если их сравнивать, то австралийский национальный веб-архив выглядит более наполненным контентом и удобным для использования. В то время как шведский краулер в отдельные годы не запускался для сканирования веб-контента, на данный момент не имеет онлайн-доступа к ресурсам и открыт для пользователей только в локальной сети библиотеки. Австралийский TROVE имеет поисковую систему, позволяющую делать запросы по ключевым словам, дате и веб-адресу, а также предоставляет пользователю возможность выбирать нужные ресурсы по тематическим коллекциям.

Основными организациями, занимающимися веб-архивированием на национальном уровне и ответственными за этот процесс, являются библиотеки и университеты. Национальные программы веб-архивирования были запущены несколько позже первых веб-архивных инициатив середины 1990-х. В начале 2000-х один за другим появляются национальные веб-архивы Норвегии (2001), Франции и Японии (2002), Хорватии и Исландии (2004), Великобритании, Дании, Латвии и Южной Кореи (2005). Расширение сотрудничества в области веб-архивирования, в первую очередь, отвечало необходимости разработки и использования оптимальных способов архивирования веба в крупном масштабе, и привело к установлению Международного Консорциума по сохранению Интернета (the International Internet Preservation Consortium (ИПС) в 2003 г.<sup>31</sup>

Если обратиться к карте распространения веб-архивных инициатив, представленной в интернет-энциклопедии Wikipedia и созданной на основании данных списка проектов в этой сфере<sup>32</sup>, то все еще видны белые пятна и существуют целые регионы, в которых нет программ по архивированию веба. В Латинской Америке, Африке, Ближнем и Среднем Востоке и в других частях земного шара все еще не функционируют такие программы, а значит, безвозвратно исчезает ценная информация. Частично эта проблема решается за счет деятельности Интернет-архива и часть контента все же сохраняется. Кроме того, ряд веб-архивных проектов,

---

<sup>30</sup> Kulturarw3. The National Library of Sweden. URL: <http://www.kb.se/om/projekt/Svenska-webbsidor---Kulturarw3> (дата обращения: 01.05.2022).

<sup>31</sup> The International Internet Preservation Consortium (ИПС). URL: <http://netpreserve.org> (дата обращения: 01.05.2022).

<sup>32</sup> List of Web archiving initiatives. Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives) (дата обращения: 01.05.2022).

инициированные организациями из разных частей мира, способствуют созданию тематических коллекций веба тех стран, где отсутствуют национальные программы по веб-архивированию. Например, Университет Техаса (США) инициировал программу по архивированию веба Латинской Америки LAWAP (Latin American Web Archiving Project)<sup>33</sup>, которая функционирует с 2005 г., и на данный момент создано 4 тематических коллекции. Естественным образом, при использовании таких веб-архивных материалов необходимо обращать внимание на специфику деятельности этих организаций, внимательно анализировать подходы к сохранению информации, уметь видеть возможные предвзятости в формировании подборок и проводить качественную источниковедческую критику.

## 2.4. Уровни архивирования веб-ресурсов

Существует три основных уровня архивирования веб-материалов (по Н. Брюггеру<sup>34</sup>):

- 1) веб-элемент – отдельный элемент веб-страницы, представленный в виде текста, изображений, аудио-, видео- и других форматах;
- 2) веб-страница – это отдельный документ во Всемирной паутине или часть веб-сайта, имеющая уникальный адрес, и которую пользователь может прокручивать для просмотра и взаимодействия с помощью веб-браузера;
- 3) веб-сайт – организованная совокупность HTML-страниц, основанная на семантике, структурных и формальных взаимосвязях.

Можно еще выделить промежуточный уровень между веб-страницей и веб-сайтом – это раздел сайта, представляющий собой набор структурно, тематически, хронологически или любым другим образом связанных страниц.

Веб-архивы можно классифицировать по уровню охвата сохраняемой информации: на глобальные (или транснациональные), национальные, региональные и локальные. Н. Брюггер<sup>35</sup> предлагает также выделить отдельную группу веб-архивов, ориентированных на конкретные исследо-

---

<sup>33</sup> The Latin American Web Archiving Project (LAWAP). URL: <http://lanic.utexas.edu/project/archives> (дата обращения: 01.05.2022).

<sup>34</sup> Brügger N. (2011). Web archiving–Between past, present and future. In: M. Consalvo & C. Ess (Eds.) The handbook of internet studies. Chichester: Wiley-Blackwell. P. 25.

<sup>35</sup> Brügger N. (2018) The Archived Web: Doing History in the Digital Age. Cambridge, MA: The MIT Press. 200 p.

вания, которые создаются в отдельных библиотеках, университетах и архивах.

К глобальным веб-архивным инициативам относится, прежде всего, Интернет-архив<sup>36</sup>, крупнейший некоммерческий веб-архивный проект. Примерами национальных инициатив являются проекты по сохранению национального сегмента веба – это веб-архивы Австрии<sup>37</sup>, Австралии<sup>38</sup>, Великобритании<sup>39</sup>, Финляндии<sup>40</sup>, Новой Зеландии<sup>41</sup> и других стран. К региональным веб-архивам относятся проекты университетов, например, которые не имеют намерения создать настолько масштабные проекты, как национальные веб-архивы, и сфокусированные на решении определенных исследовательских и образовательных задач. Таковыми, например, являются веб-архивные проекты университетской библиотеки Братиславы (Словакия)<sup>42</sup>, библиотек Стенфордского университета (США)<sup>43</sup> и университета Йорка (Канада)<sup>44</sup>.

Поддерживает инициативы по веб-архивированию Международный консорциум по сохранению Интернета (International Internet Preservation Consortium, ИПС), главная цель которого – «обеспечить сохранение богатого интернет-контента со всего мира таким образом, чтобы его можно было архивировать, гарантировать защиту и получать к нему доступ с течением времени»<sup>45</sup>. Консорциум способствует выработке единых подходов по веб-архивированию, преодолению проблем, связанных со сбором, хранением и обеспечением доступа к данным.

---

<sup>36</sup> Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).

<sup>37</sup> Webarchiv Österreich. URL: <https://webarchiv.onb.ac.at/> (дата обращения: 01.05.2022).

<sup>38</sup> Trove. National Library of Australia. URL: <https://trove.nla.gov.au/> (дата обращения: 01.05.2022).

<sup>39</sup> The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk/> (дата обращения: 01.05.2022).

<sup>40</sup> Finnish Web Archive. URL: <https://verkkoarkisto.kansalliskirjasto.fi> (дата обращения: 01.05.2022).

<sup>41</sup> New Zealand Web Archive. URL: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive> (дата обращения: 01.05.2022).

<sup>42</sup> Digital Resources. University Library in Bratislava. URL: <https://www.webdepozit.sk> (дата обращения: 01.05.2022).

<sup>43</sup> Web-archiving. Stanford University Libraries. URL: <https://library.stanford.edu/projects/web-archiving> (дата обращения: 01.05.2022).

<sup>44</sup> Web archives. York University Digital Library. URL: <https://digital.library.yorku.ca> (дата обращения: 01.05.2022).

<sup>45</sup> International Internet Preservation Consortium, ИПС. URL: <http://netpreserve.org/about-us/> (дата обращения: 01.05.2022).

Понимание процессов, лежащих в основе сохранения веба, позволяет корректнее работать с цифровыми источниками, видеть информационный исследовательский потенциал в заархивированном материале и качественнее проводить источниковедческую критику.

Веб-архивирование представляет собой сложный процесс, который состоит из следующих стадий:

- 1) выбор ресурсов для сохранения;
- 2) сбор данных;
- 3) сохранение;
- 4) документирование;
- 5) обеспечение доступа.

Рассмотрим указанные стадии более подробно.

## **2.5. Стратегии отбора ресурсов для сохранения**

Существует два основных уровня веб-архивирования – это микро- и макроархивирование, различающиеся между собой масштабом предпринятого проекта.

Проекты, реализуемые на макроуровне, являются более масштабными и чаще всего преследуют цель сохранения национального веба или цифрового наследия в самом широком смысле. Такие проекты предпринимаются и проводятся крупными академическими организациями и национальными центрами знаний, как библиотеки, например. Макроуровневые проекты довольно сложные по структуре и контенту, который может варьироваться по тематике, форматам и разнообразию материалов, и могут обеспечивать сохранение источников для их использования, в частности, служить основой для исследований в области истории и других гуманитарных и социальных дисциплин.

Микроуровень подразумевает реализацию проектов по веб-архивированию отдельными исследовательскими группами, которые имеют более частные задачи и цели для сбора веб-ресурсов и их последующего изучения.

Определение критериев для отбора материалов с целью их долгосрочного сохранения является принципиальным этапом любого проекта по веб-архивированию. Идеальной картиной веб-архивирования мог бы послужить подход со сплошным сохранением всех созданных ресурсов, чтобы уберечь их от исчезновения и обеспечить максимальную полноту и

разнообразие источников. Однако такой подход вряд ли можно считать состоятельным. С одной стороны, сохранить весь веб является недостижимой задачей, так как к настоящему моменту накоплен огромный объем данных<sup>46</sup>, который продолжает постоянно увеличиваться, и его невозможно даже оценить с высокой степенью точности, а соответственно, и определить хотя бы приблизительно технические характеристики ресурсов, требуемых для сохранения и обслуживания, также не представляется реальным. С другой стороны, динамизм появления и развития веб-ресурсов настолько интенсивный, что процесс сохранения даже в самой оптимистичной перспективе становится бесконечным.

Вопрос о том, какие ресурсы должны сохраняться в первую очередь, относится к первостепенной важности, поскольку определяет рамки и характеристики сохраняемого веба для будущих поколений, и это содержимое, несомненно, должно иметь ценность. Таким образом, на этапе отбора ресурсов для веб-архивирования необходимо руководствоваться выработанной стратегией, которая может быть обусловлена политикой в области культурного наследия (и цифрового культурного наследия в т.ч.) и подходами к развитию цифровых инфраструктур на государственном уровне. Так, разрабатываемые национальные цифровые инфраструктуры часто содержат веб-архив в качестве одного из своих компонентов. Такие инфраструктуры обеспечивают удобство использования сохраняемого цифрового (born-digital) и оцифрованного (digitized) контента, а также обеспечивают комфортность доступа к этим материалам разным аудиториям пользователей.

Основные стратегии по отбору ресурсов для веб-архивирования могут быть основаны на:

- 1) определении качественных характеристик ресурсов, их значимости, авторитетности, информационного потенциала и прогнозирования того, насколько ресурс станет востребован в будущем;
- 2) сохранении ресурсов определенного доменного пространства;
- 3) определении временных рамок.

---

<sup>46</sup> См. Internet Live Stats. URL: <https://www.internetlivestats.com/total-number-of-websites/#trend> (дата обращения: 01.05.2022).



### 2.5.1. Этика и правовые основания сохранения веба

Веб-архивирование может быть встроено в систему государственной политики сохранения национальных публикаций, в том числе, выпущенных в электронном формате. Речь идет об обязательном экземпляре документов (т.н. *legal deposit*)<sup>47</sup>. Перечни и характеристики таких экземпляров определяются национальным законодательством, обозначаются в стратегиях цифровизации и могут содержать разнообразный материал, создаваемый в электронном формате, как, например, книги или веб-сайты. В Российской Федерации действует ФЗ «Об обязательном экземпляре документов» (от 29 дек. 1994 г. № 77-ФЗ), который «определяет политику государства в области формирования обязательного экземпляра документов как ресурсной базы комплектования полного национального библиотечно-информационного фонда документов Российской Федерации и развития системы государственной библиографии»<sup>48</sup>. Частью такого депозита могут быть и веб-сайты. Так, например, Национальное управление архивов и документации США (NARA) регулярно запрашивает у федеральных департаментов снапшоты сайтов для обеспечения их долговременного хранения. NARA выпустило меморандум, в котором говорится о том, что следует обязательно сохранять снапшоты сайтов с материалами Президента и Конгресса США<sup>49</sup>, другие агентства могут предоставлять эту информацию по отдельному запросу или совсем не предоставлять ее. Программа реализуется с 2004 г., и коллекции регулярно пополняются. Обеспечен свободный доступ к веб-архиву, и просмотр заархивированного контента возможен на платформе WebHarvest<sup>50</sup>. Во многих других случаях сохранение веба строится на добровольных началах.

Следует обратить внимание и на правовую основу веб-архивирования. Круг ресурсов может определяться авторизованными для сохранения веба организациями, а ответ на вопрос о том, необходимо ли получать согласие на сканирование веба у владельцев контента (частных лиц и

---

<sup>47</sup> Legal Deposit. International Internet Preservation Consortium. URL: <https://netpreserve.org/web-archiving/legal-deposit> (дата обращения: 01.05.2022).

<sup>48</sup> Федеральный закон от 29 дек. 1994 г. № 77-ФЗ «Об обязательном экземпляре документов» (с изменениями на 8 июня 2020 г.).

<sup>49</sup> Web Harvest Background Information. National Archives and Records Administration. URL: <https://www.archives.gov/records-mgmt/memos/nwm13-2008-brief.html> (дата обращения: 01.05.2022).

<sup>50</sup> Congressional & Federal Government Web Harvests. URL: <https://www.webharvest.gov/> (дата обращения: 01.05.2022).

организаций), зависит от масштаба веб-архива, его назначения и действующего национального законодательства. Так, Национальной библиотеке Новой Зеландии не требуется дополнительного согласия собственников веб-сайтов, поскольку библиотека обладает законным правом на сохранение веб-контента, находящегося в открытом доступе. Аналогичным образом особого разрешения не требуется ответственным за национальные программы веб-архивирования организациям в США и Великобритании. Если же веб-архивы являются частными инициативами, то наиболее вероятно, что разрешение на сканирование веб-контента может потребоваться. Интернет-архив не просит разрешения на сканирование и сохранение контента, однако владелец веб-сайта может отправить запрос на исключение своего сайта из информационной системы архива<sup>51</sup>, которое будет рассмотрено в определенном порядке для удаления сведений.

В отношении материалов социальных сетей, которые создаются частными лицами, законодательство значительно сложнее и строже. При сборе данных с личных страниц в социальных сетях необходимо внимательно ознакомиться с национальным законодательством о персональных данных. Такая необходимость может быть связана с научно-исследовательскими задачами. В этом случае важным требованием является очистка сохраняемых данных, т.е. удаление персональной информации, а также отказ от их опубликования в чистом виде. В отдельных случаях, распространенной практикой является шифрование данных персонального характера. Кроме того, использование должно быть мотивировано исследовательской целью.

Другим поводом для удаления сохраненной в веб-архивах информации о персоне может стать «право на забывание» (“Right to be forgotten”)<sup>52</sup>, которое связано с возможностью частного лица сделать защищенными от доступа данные о себе или удалить их полностью из социальных сетей, веб-ресурсов, в том числе и из веб-архивов. В этом случае человек может реализовать свое право на управление информацией, и необходимо уважать такое решение, несмотря на то, что оно делает невозможным использование этих данных, в том числе, в научных целях, сужая источниковую

---

<sup>51</sup> Using The Wayback Machine. URL: <https://help.archive.org/hc/en-us/articles/360004651732-Using-The-Wayback-Machine> (дата обращения: 01.05.2022).

<sup>52</sup> См. Dulong de Rosnay M., Guadamuz A. (2017) Memory Hole or Right to Delist? Implications of the Right to be Forgotten for Web Archiving // RESET, Vol. 6. Online since 30 October 2016. DOI: <https://doi.org/10.4000/reset.807>, URL: <http://journals.openedition.org/reset/807> (дата обращения: 01.05.2022).

базу и ограничивая возможную тематику гуманитарных и социальных исследований. В целом, вопросы о балансе частного и общественного, личного и научного и других перспектив являются предметом постоянных дискуссий<sup>53</sup>.

Таким образом, как при реализации программ веб-архивирования, так и в процессе создания отдельных репозиторий и даже наборов данных необходимо обращать внимание на работу с персональными данными и на авторские права держателей, субъектов прав и третьих лиц.

### **2.5.2. Ценностный подход к сохранению веб-ресурсов**

Одним из важнейших критериев отбора веб-ресурсов для сохранения является ценность сохраняемых ресурсов, которая определяется их исторической, культурной, общественной, политической, научной, образовательной значимостью.

Представление о том, что в будущем может быть востребовано пользователями, в том числе и исследователями, влияет на выбор ресурсов для веб-архивирования. Историк Т. Хичкок ёмко назвал современные веб-архивы «объектом изучения воображаемого ученого»<sup>54</sup>. Эксперты из проекта Kulturarw3<sup>55</sup> по архивированию цифрового наследия из Швеции обратили внимание на тот факт, что никто не знает достоверно, какая информация окажется важной и востребованной будущими поколениями, и архивирование может основываться лишь на предположениях о будущих запросах и исследовательских целях<sup>56</sup>. В этом отношении важно предположить потенциальный интерес со стороны разных общественных групп в ресурсах, понять будущие нужды, попытаться их определить заранее. Поэтому эта стратегия веб-архивирования основана на прогнозировании бу-

---

<sup>53</sup> См. например, Engle E. Legal Issues in Web Archiving. The blog post. In: The Signal. May 30, 2012. URL: <https://blogs.loc.gov/thesignal/2012/05/legal-issues-in-web-archiving> (дата обращения: 01.05.2022); Law Enforcement Requests. The blog post. Internet Archive. URL: <https://help.archive.org/hc/en-us/articles/360004716011-Law-Enforcement-Requests> (дата обращения: 01.05.2022).

<sup>54</sup> Hitchcock T. (2015) The UK Web Archive, Born-Digital Sources, and Rethinking the Future of Research. In: Web Archives for Historians (blog), 5 June 2015. URL: <https://webarchivehistorians.org/2015/06> (дата обращения: 01.05.2022).

<sup>55</sup> Kulturarw3. The National Library of Sweden. URL: <http://www.kb.se/om/projekt/Svenska-webbsidor---Kulturarw3> (дата обращения: 01.05.2022).

<sup>56</sup> Arvidson A., Persson K., Mannerheim J. (2000) The Kulturarw3 Project: The Royal Swedish Web Archiw3e: An example of 'complete' collection of Web pages. In: 66th International Federation of Library Associations and Institutions Council and General Conference Jerusalem, Israel, 13-18 August, 2000. URL: <https://archive.ifla.org/IV/ifla66/papers/154-157e.htm> (дата обращения: 01.05.2022).

дущих запросов и на понимании долговременных ценностей, определяющих значимость контента, в котором эти ценности отражены.

В отчете особой группы по изучению устойчивости процессов сохранения и доступности цифрового контента (Blue Ribbon Task Force 2010) указано, что в веб-архивах «интерес будущих пользователей слабо представлен в отобранных материалах для сохранения», и учреждения, обеспечивающие доступ к этим ресурсам (в первую очередь, архивы и библиотеки), должны быть крайне внимательны, определяя круг архивируемых ресурсов для потенциальных пользователей<sup>57</sup>. Сложность с отбором сегментов для сохранения заключается еще и в том, что зачастую трудно или даже невозможно предсказать дальнейшее развитие веба. Проблематично определить ресурсы, которые будут развиваться, и те, что исчезнут как невостребованные.

Очевидно, что представления о ценности ресурсов для сохранения варьируются у разных заинтересованных сторон, свой интерес есть и у историков, и у других гуманитариев. Поэтому очевидно, что для максимальной эффективности программы по сохранению веба должны основываться на учете мнений разных категорий пользователей и на сотрудничестве профессионалов из разных отраслей – библиотекарей, архивистов, математиков, информатиков и др. М. Киршенбаум обратил внимание на то, что важное место в этом процессе занимают специалисты из области цифровых гуманитарных наук. Они способствуют развитию информационного пространства и созданию инфраструктуры для гуманитарных исследований. Эти специалисты вносят свой вклад в продвижение материалов веб-архивов для использования в разных гуманитарных и смежных областях, апробируя среду веб-архивов и созданные инструменты поиска и анализа, а также создают условия для трансформации этой среды в сторону большего удобства для решения исследовательских задач<sup>58</sup>.

---

<sup>57</sup> Blue Ribbon Task Force, The. 2010. Sustainable Economics for a Digital Planet: Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access Ensuring Long-Term Access to Digital Information. February 2010. Funded by U.S. National Science Foundation, The Andrew W. Mellon Foundation, the U.S. Library of Congress, the U.K. Joint Information Systems Committee, the Electronic Records Archives Program of the National Archives and Records Administration, and the Council on Library and Information Resources. 2010, p. 2. URL: [https://web.archive.org/web/20200208074650/http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](https://web.archive.org/web/20200208074650/http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf) (дата обращения: 01.05.2022).

<sup>58</sup> Kirschenbaum, M. (2013). The .txtual condition: digital humanities, born-digital archives, and the future literary // Digital Humanities Quarterly, Vol. 7(1), URL: <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html> (дата обращения: 01.05.2022).

**Сохранение данных во всей полноте**, естественным образом, представляется как идеальный подход. Однако уже упоминалось ранее, что он не может быть реализован на пространстве всей планеты. С другой стороны, стратегии сохранения национального веба могут развиваться в этом направлении. Так, в Великобритании на основании закона об обязательном экземпляре<sup>59</sup> сохраняются сайты из британских доменных зон (к ним относятся, такие домены как *.uk*, *.london*, *.scot* и другие). Кроме того, под программу сохранения попадают также сайты, хранящиеся на серверах, расположенных на территории Великобритании, сайты, содержащие в контактных данных британский адрес, или по содержанию которых понятно, что контент был произведен в этой стране. Подход Великобритании к сохранению веба заключается в том, чтобы сканировать и сохранять в веб-архиве (the UK Web Archive, UKWA<sup>60</sup>) материалы как минимум один раз в год. Преимущество такого подхода очевидно – это сохранение материалов в их полноте. Однако заархивированный массив едва ли поддается разделению какие-то определенные группы – тематические, по типу ресурса или организации, что создает сложности в навигации по сохраненным данным и проблематичность его последующего использования.

Несколько другим подходом к отбору ресурсов является представление о необходимости сохранения ресурсов в их **разнообразии** на основе отдельных избранных порций национального веба. В рамках такого подхода выявляются ресурсы отдельных тем и типов. Особым вариантом является стремление зафиксировать не только положительные и выдающиеся примеры, но и обозначить другие характерные стороны современной жизни, в том числе, характеризующиеся со знаком «минус», как, например, криминал, порнография, мошеннические ресурсы и т.д. Так, например, французский национальный проект по веб-архивированию, который реализуется Национальной библиотекой Франции (Bibliothèque nationale de France, BnF)<sup>61</sup>, направлен на сохранение такого разнообразия в рамках национальных французских доменов, таких как *.fr* и *.re*, а также сайтов в домене *.com*, зарегистрированных на территории Франции. Закон Фран-

---

<sup>59</sup> Joint Collecting Framework for UK Legal Deposit, 2015-2020 (LDLC 15-15 (IG 335)). URL: <https://www.bl.uk/britishlibrary/~media/bl/global/legal%20deposit/joint-collecting-framework-for-uk-legal-deposit.pdf> (дата обращения: 01.05.2022).

<sup>60</sup> The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk> (дата обращения: 01.05.2022).

<sup>61</sup> Bibliothèque nationale de France. URL: <https://www.bnf.fr/fr> (дата обращения: 01.05.2022).

ции о наследии (Code du Patrimoine)<sup>62</sup> предполагает создание и наполнение особого национального депозита, который включает помимо документов, имеющих аналоговый формат и оцифрованных для сохранения, компонент французского веба, предназначенный к сбору и архивированию на регулярной основе. Причем Франция не предпринимает сплошного сохранения веба, а выбирает отдельные «образцы» из разнообразных по типу, тематике и т.д. областей французского веб-пространства.

**Репрезентативность** представленного контента основывается не только на оценке значимости содержания ресурса, но и на определении других характеристик отбираемых примеров. Такой подход направлен на сохранение важных примеров из большой выборки однотипного веб-контента, что позволяет уменьшить объемы архивируемого материала. Выбор примеров для национального веб-архива Франции основан на признании необходимости сохранения тех ресурсов, которые «отражают французское общество и культуру в их разнообразии вне зависимости от научной значимости или популярности публикаций»<sup>63</sup>. С этой точки зрения, должны сохраняться как примеры лучшего и выдающегося контента (академического, культурного и другого), так и примеры не самых лучших практик (навязчивой рекламы, запрещенных товаров и услуг и т.д.), чтобы сохранить разнообразие, а не уменьшить или исказить его.

Однако возникает вопрос о том, как представить это разнообразие в отношении веб-ресурсов, их типов, содержания, тематики и проблематики, затрагиваемой на сайтах или в отдельных их разделах. Н. О. Финнеман в качестве ключевых критериев разнообразия предлагает следующий набор характеристик: авторство, культурные и социальные практики, коммуникативные жанры, визуальные и звуковые свойства, поисковые возможности, интерфейсы и веб-дизайн, ссылки и сетевые отношения, темы и проблематика, чувствительность ресурсов ко времени, а также координация синхронных и диахронных перспектив<sup>64</sup>. Кроме того, разнообразие ресурсов должно достигаться и в направлениях использования этих

---

<sup>62</sup> Code du Patrimoine. URL: [https://www.legifrance.gouv.fr/codes/texte\\_lc/LEGITEXT000006074236](https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006074236) (дата обращения: 01.05.2022).

<sup>63</sup> Lasfargues, F., Oury, C., & Wendland, B. (2008). Legal deposit of the French Web: Harvesting strategies for a national domain. Proceedings from IAWAW '08: 8th International Workshop for Web Archiving, Denmark. URL: <https://hal-bnf.archives-ouvertes.fr/hal-01098538/document> (дата обращения: 01.05.2022).

<sup>64</sup> Finnemann N.O. (2018) Web archives and Knowledge organization. In: Birger Hjørland (Ed.) ISKO Encyclopedia of Knowledge Organization & Knowledge Organization. URL: [https://www.isko.org/cyclo/web\\_archive](https://www.isko.org/cyclo/web_archive) (дата обращения: 01.05.2022).

ресурсов – от исследовательских и образовательных задач, связанных с анализом цифрового культурного наследия и исторических источников, до возможных коммерческих целей, от документирования развития гражданского общества до представления индивидуальных интересов и отдельных персональных убеждений. Разнообразие также может измеряться по разным шкалам – в микро и макроизмерениях, событиях и тематике, значимых локально или глобально.

**Качество и авторитетность** ресурса являются одними из оснований для сохранения в веб-архиве, поскольку позволяют спрогнозировать востребованность этого ресурса в будущем. **Качество** подразумевает наличие определенных свойств у ресурса, которые характеризуют его содержательно и функционально и обеспечивают возможность удовлетворять интересы пользователей в наиболее полной мере по сравнению с другими аналогичными ресурсами. **Авторитетность** подчеркивает значимость ресурса, содержащего надежную информацию, заслуживающую доверия, и производимого понятным, авторизованным издателем, организацией или индивидом. К таким ресурсам часто относятся сайты с официальной информацией – например, сайты и порталы органов власти и местного самоуправления, политических партий, личные сайты депутатов, представителей академических или образовательных кругов и т.д. Действительно, правительственные сайты часто становятся в центр внимания при создании тематических коллекций, поскольку отражают проводимую государством политику в разных направлениях и содержат официальную версию происходящих событий. Например, единственная программа по архивированию веба в России связана с сохранением именно этого типа ресурсов – сайтов федеральных и региональных органов власти, а также аккаунты в социальных сетях некоторых первых лиц государства<sup>65</sup>.

Форматы цифрового контента, сохраняемого в качестве отдельных элементов веб-страниц, могут быть различными – текстовый, изобразительный, аудио, видео, мультимедиа и некоторые другие. Интернет-архив, в частности, сканирует контент с платформы YouTube и формирует коллекцию видео<sup>66</sup>, доступную для просмотра независимо от наличия или отсутствия этих записей на известном видеохостинге. Также в Интернет-

---

<sup>65</sup> Консервированное государство. Национальный цифровой архив России. URL: <https://ruarxive.org/projects/pegov/> (дата обращения: 01.05.2022).

<sup>66</sup> Youtube Videos. Internet Archive. URL: <https://archive.org/details/youtubecrawl> (дата обращения: 01.05.2022).

архиве представлены разные типы оцифрованного и цифрового контента, в том числе, книги, журналы, газеты и другие периодические издания, фильмы, аудио, датасеты и программное обеспечение<sup>67</sup>.

Часто подходы к отбору ресурсов для сохранения могут быть гибридными и сочетать, например, тему и определенный формат контента. Так, например, в коллекции Indigenous Australians веб-архива TROVE представлены три субподборки, в которых выделены видеофайлы, новости и «Национальные извинения»<sup>68</sup>.

### 2.5.3. Доменное сканирование

Доменное сканирование используется преимущественно для архивирования национального веба. Такой способ сохранения основан на сплошном сканировании сайтов определенного домена верхнего уровня, выделенного для конкретной страны. Если краулер встречается со ссылкой, ведущей на другой домен, программа не продолжает работу в этом направлении. Так, например, Национальная библиотека Франции сохраняет сайты из доменов *.fr* и *.re*, являющимися французскими национальными доменами.

Доменное сканирование связано с сохранением внушительных объемов веб-ресурсов, что требует особых подходов для организации, хранения и обеспечения эффективного доступа к ним. Для удобной навигации по заархивированным ресурсам могут создаваться тематические коллекции, отбор в которые основан на содержательных особенностях материалов. Крупномасштабные проекты, какими являются национальные веб-архивы, и особенно основанные на доменном сканировании, вовлекают автоматизированную генерацию метаданных, в том числе, отражающих содержание ресурсов. Такой подход представляется более эффективным по сравнению с «ручным» режимом отбора ресурсов для их определения в ту или иную тематическую коллекцию, хотя национальные программы веб-архивирования и не освобождены от «ручного» курирования данных специалистом.

---

<sup>67</sup> Internet Archive Blogs. Internet Archive. URL: <http://blog.archive.org/category/video-archive> (дата обращения: 01.05.2022).

<sup>68</sup> Archived Webpage Collections. Indigenous Australians. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au/collection/15012> (дата обращения: 01.05.2022).



Доменное сканирование имеет ряд преимуществ по сравнению с другими стратегиями сохранения веба:

1. Доменное сканирование как стратегия сохранения веба снимает проблему выбора ресурсов для архивирования. Определение исключительно ценных с какой-либо точки зрения ресурсов является критичной задачей при создании тематических коллекций на основе выборочного сохранения. Поэтому подход, основанный на сканировании доменов, создает существенно более широкие возможности для последующего использования заархивированных материалов, в т.ч. формулирования более разнообразных и масштабных исследовательских задач.

2. В результате использования такого подхода формируется широкая база заархивированных веб-сайтов, которые в более полной мере по сравнению с выборочным сканированием отражают разные стороны современной жизни – культурной, политической, экономической и других сфер. Кроме того, доменное сканирование позволяет сделать тематические коллекции более полными по сравнению с теми, которые формируются куратором вручную.

3. Доменное сканирование априори снимает проблему определения ценности отдельных ресурсов. Поскольку исследовательские и поисковые задачи будущего невозможно спрогнозировать с высокой степенью точности, многие заархивированные ресурсы могут приобрести ценность намного позже их попадания в веб-архив.

4. Высокая степень автоматизации процессов путем назначения домена для сканирования и решение этим проблемы выборки.

5. Архивирование веб-сайтов с сохранением их взаимосвязей в пределах домена (гиперссылочной массы и веб-контента) помимо последующего воспроизведения отсканированных веб-ресурсов и просмотра информации на них позволяет осуществлять аналитику связей между сайтами, в т.ч. изучать архитектуру веб-ресурсов, их взаимосвязи с разными ресурсами внутри домена и с внешними доменами, выявляя эти сетевые структуры.

6. Возможность применения разнообразных методов исследования на большом массиве данных, в первую очередь, количественных.

7. Доменное сканирование чаще всего встроено в национальные программы сохранения веба и используется в качестве долговременной стратегии, что позволяет выявлять «срезы» и анализировать эволюцию домена и отдельных его ресурсов.

#### **2.5.4. Временные рамки как характеристика отбора ресурсов**

Определение временных рамок для сканирования ресурсов конкретной тематики или типа является несколько иной стратегией веб-архивирования. Временные рамки для сохранения могут быть установлены на определенный период – день, неделю, месяц, квартал и т.д., если известны сроки, в пределах которых может быть получена определенная информация. Такие временные границы могут быть определены предвыборной гонкой, например, фестивалем или каким-то мероприятием.

Чувствительность к временной шкале может быть значимой для изучения динамичных ресурсов, например, социальных медиа или форумов. Коллекции, полученные в рамках этого подхода, специфичны тем, что полученные в результате сохранения материалы отражают динамику ресурса за определенный период времени с визуализацией этих изменений – привнесений, удалений, исправлений и корректировки ссылок и взаимосвязей в архитектуре сайтов и внешних связей с другими ресурсами. Такие веб-архивные проекты сохраняют набор дополнительных метаданных, касающихся обновляемости ресурсов, внесения изменений в веб-сайты и их элементы (в т.ч. удаление и добавление веб-страниц, редактирование текстов на страницах, изображений, добавление или удаление аудио- и видео-контента). Этот подход представляется интересным и имеет большой потенциал при изучении, например, политически ангажированных ресурсов.

\*\*\*

Большинство крупных веб-архивных проектов чаще всего используют комбинацию перечисленных стратегий. Пример Интернет-архива является одним из наиболее интересных, поскольку использует подходы по сканированию национального американского веба и его отдельных фрагментов, а также выборочно сканирует веб-ресурсы других стран. После терактов 11 сентября 2001 г. была создана первая тематическая коллекция, основанная на этих трагических событиях, и стратегия выбора ресурсов по критерию связи с тем или иным событием стала важной частью стратегии веб-архивирования данного проекта. На данный момент на сайте проекта можно обнаружить целую совокупность коллекций, посвященных тем или иным событиям.

## Рекомендуемая литература

1. Arvidson A., Persson K., Mannerheim J. (2000) The Kulturarw3 Project: The Royal Swedish Web Archiw3e: An example of 'complete' collection of Web pages. In: 66th International Federation of Library Associations and Institutions Council and General Conference Jerusalem, Israel, 13–18 August, 2000. URL: <https://archive.ifla.org/IV/ifla66/papers/154-157e.htm> (дата обращения: 01.05.2022).
2. Brügger N. (2011). Web archiving – Between past, present and future. In: M. Consalvo & C. Ess (Eds.) The handbook of internet studies. Chichester: Wiley-Blackwell. p. 25.
3. Brügger N. (2016) Webraries and Web archives: the Web between public and private. In W. Evans and D. Baker (Eds), The End of Wisdom? The Future of Libraries in a Digital Age. Oxford: Chandos. Pp. 185–190.
4. Brügger N. (2018) The Archived Web: DOIing History in the Digital Age. Cambridge, MA: The MIT Press. 200 p.
5. Dulong de Rosnay M., Guadamuz A. (2017) Memory Hole or Right to Delist? Implications of the Right to be Forgotten for Web Archiving // RESET, Vol. 6. Online since 30 October 2016. DOI: <https://doi.org/10.4000/reset.807>, URL: <http://journals.openedition.org/reset/807> (дата обращения: 01.05.2022).
6. Engle E. Legal Issues in Web Archiving. The blog post. In: The Signal. May 30, 2012. URL: <https://blogs.loc.gov/thesignal/2012/05/legal-issues-in-web-archiving> (дата обращения: 01.05.2022)
7. Finnemann N. O. (2018) Web archives and Knowledge organization. In: Birger Hjørland (Ed.) ISKO Encyclopedia of Knowledge Organization & Knowledge Organization. URL: [https://www.isko.org/cyclo/web\\_archive](https://www.isko.org/cyclo/web_archive) (дата обращения: 01.05.2022).
8. Hitchcock T. (2015) The UK Web Archive, Born-Digital Sources, and Re-thinking the Future of Research. In: Web Archives for Historians (blog), 5 June 2015. URL: <https://webarchivehistorians.org/2015/06> (дата обращения: 01.05.2022).
9. Kirschenbaum M. (2013). The .txtual condition: digital humanities, born-digital archives, and the future literary // Digital Humanities Quarterly, Vol. 7(1), URL: <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html> (дата обращения: 01.05.2022).
10. Lasfargues F., Oury, C., & Wendland, B. (2008). Legal deposit of the French Web: Harvesting strategies for a national domain. Proceedings from IAWW '08: 8th International Workshop for Web Archiving, Denmark. URL: <https://hal-bnf.archives-ouvertes.fr/hal-01098538/document> (дата обращения: 01.05.2022).

## Источники

1. Об обязательном экземпляре документов: Фед. закон от 29 дек. 1994 г. № 77-ФЗ (с изменениями на 8 июня 2020 г.).
2. Консервированное государство. Национальный цифровой архив России. URL: <https://ruarxive.org/projects/pregov/> (дата обращения: 01.05.2022).
3. About the Internet Archive. URL: <https://archive.org/about/> (дата обращения: 01.05.2022).
4. Archived Webpage Collections. Indigenous Australians. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au/collection/15012> (дата обращения: 01.05.2022).
5. Bibliothèque nationale de France. URL: <https://www.bnf.fr/fr> (дата обращения: 01.05.2022).
6. Blue Ribbon Task Force, The. 2010. Sustainable Economics for a Digital Planet: Final Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access Ensuring Long-Term Access to Digital Information. February 2010. Funded by U.S. National Science Foundation, The Andrew W. Mellon Foundation, the U.S. Library of Congress, the U.K. Joint Information Systems Committee, the Electronic Records Archives Program of the National Archives and Records Administration, and the Council on Library and Information Resources. 2010, p. 2. URL: [https://web.archive.org/web/20200208074650/http://brtf.sdsc.edu/biblio/BRTF\\_Final\\_Report.pdf](https://web.archive.org/web/20200208074650/http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf) (дата обращения: 01.05.2022).
7. Code du Patrimoine. URL: [https://www.legifrance.gouv.fr/codes/texte\\_lc/LEGITEXT000006074236](https://www.legifrance.gouv.fr/codes/texte_lc/LEGITEXT000006074236) (дата обращения: 01.05.2022).
8. Congressional & Federal Government Web Harvests. URL: <https://www.webharvest.gov/> (дата обращения: 01.05.2022).
9. Digital Resources. University Library in Bratislava. URL: <https://www.webdepozit.sk> (дата обращения: 01.05.2022).
10. Finnish Web Archive. URL: <https://verkkoarkisto.kansalliskirjasto.fi> (дата обращения: 01.05.2022).
11. Helitrix. Internet Archive. URL: <http://crawler.archive.org/index.html> (дата обращения: 01.05.2022).
12. International Internet Preservation Consortium, ИПС. URL: <http://netpreserve.org/about-us/> (дата обращения: 01.05.2022).
13. Internet Archive Blogs. Internet Archive. URL: <http://blog.archive.org/category/video-archive> (дата обращения: 01.05.2022).
14. Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).
15. Internet Live Stats. URL: <https://www.internetlivestats.com/total-number-of-websites/#trend> (дата обращения: 01.05.2022).
16. Joint Collecting Framework for UK Legal Deposit, 2015-2020 (LDLC 15-15 (IG 335)). URL: <https://www.bl.uk/britishlibrary/~media/bl/global/legal%20>

deposit/joint-collecting-framework-for-uk-legal-deposit.pdf (дата обращения: 01.05.2022).

17. Kulturarw3. The National Library of Sweden. URL: <http://www.kb.se/om/projekt/Svenska-webbsidor---Kulturarw3> (дата обращения: 01.05.2022).

18. Law Enforcement Requests. The blog post. Internet Archive. URL: <https://help.archive.org/hc/en-us/articles/360004716011-Law-Enforcement-Requests> (дата обращения: 01.05.2022).

19. Legal Deposit. International Internet Preservation Consortium. URL: <https://netpreserve.org/web-archiving/legal-deposit> (дата обращения: 01.05.2022).

20. List of Web archiving initiatives. Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives) (дата обращения: 01.05.2022).

21. New Zealand Web Archive. URL: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive> (дата обращения: 01.05.2022).

22. The International Internet Preservation Consortium (IIPC). URL: <http://netpreserve.org> (дата обращения: 01.05.2022).

23. The Latin American Web Archiving Project (LAWAP). URL: <http://lanic.utexas.edu/project/archives> (дата обращения: 01.05.2022).

24. The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk> (дата обращения: 01.05.2022).

25. Trove. National Library of Australia. URL: <https://trove.nla.gov.au/> (дата обращения: 01.05.2022).

26. Using The Wayback Machine. URL: <https://help.archive.org/hc/en-us/articles/360004651732-Using-The-Wayback-Machine> (дата обращения: 01.05.2022).

27. Wayback Machine. URL: <https://web.archive.org/> (дата обращения: 01.05.2022).

28. Web archives. York University Digital Library. URL: <https://digital.library.yorku.ca> (дата обращения: 01.05.2022).

29. Web Harvest Background Information. National Archives and Records Administration. URL: <https://www.archives.gov/records-mgmt/memos/nwm13-2008-brief.html> (дата обращения: 01.05.2022).

30. Webarchiv Österreich. URL: <https://webarchiv.onb.ac.at/> (дата обращения: 01.05.2022).

31. Web-archiving. Stanford University Libraries. URL: <https://library.stanford.edu/projects/web-archiving> (дата обращения: 01.05.2022).

32. World Wide Web. URL: <http://info.cern.ch/hypertext/WWW/TheProject.html> (дата обращения: 01.05.2022).

33. Youtube Videos. Internet Archive. URL: <https://archive.org/details/youtubecrawl> (дата обращения: 01.05.2022).

## Вопросы для самопроверки

1. Назовите изобретателей Всемирной паутины.
2. Кто является основателями Интернет-архива?
3. Что такое веб-архивирование?
4. Какие веб-архивные инициативы вам известны? Назовите их особенности.
5. Какие существуют уровни веб-архивирования? Раскройте специфику этих уровней.
6. Назовите основные типы веб-архивов по уровню охвата сохраняемой информации и дайте характеристику каждого типа.
7. Перечислите стадии веб-архивирования.
8. Назовите особенности микро- и макроуровней веб-архивных проектов.
9. Какие существуют основные стратегии по отбору ресурсов для веб-архивирования?
10. Перечислите основные критерии отбора ресурсов для их сохранения.
11. Как влияет закон об обязательном экземпляре на возможности сохранения веба?
12. Приведите примеры разных национальных практик сохранения веба на основе законов об обязательном экземпляре.
13. Какие особенности и ограничения существуют для веб-архивирования на персональном уровне?
14. В чем заключается ценностный подход к сохранению веба?
15. Приведите примеры веб-архивных проектов с разными подходами к пониманию ценности сохраняемых ресурсов.
16. В чем заключаются общие и особенные черты в стратегиях архивирования национального веба в Великобритании и Франции?
17. Какие существуют особенности сохранения веб-ресурсов по критерию репрезентативности?
18. Раскройте особенности веб-архивирования на основе качества и авторитетности ресурсов.
19. Перечислите форматы контента, который может быть сохранен в веб-архиве?
20. Раскройте специфику доменного сканирования.
21. Какие существуют преимущества у доменного сканирования?
22. Как влияют временные рамки на сохранение ресурсов, и какую функцию выполняет установление временных пределов для сканирования?

## **ГЛАВА 3. СБОР, СОХРАНЕНИЕ, ДОКУМЕНТИРОВАНИЕ И ОБЕСПЕЧЕНИЕ ДОСТУПА К ДАННЫМ В ЖИЗНЕННОМ ЦИКЛЕ ВЕБ-АРХИВА**

### **3.1. Сбор данных**

Сбор данных осуществляется с помощью особых программ, краулеров, сканирующих сервера и сохраняющих коды сайтов, тексты, изображения и другое их содержимое. Краулер следует по гиперссылкам, переходя с одной страницы на другую, сохраняя взаимосвязи веб-страниц, чтобы в дальнейшем можно было воспроизвести структуру и наполнение сайта в том виде, в котором он был отсканирован в определенный момент времени. Одновременно со сканированием веб-сайта краулер создает метаданные о процессе сканирования и сохраняет их. Не весь процесс сканирования может быть осуществлен в автоматическом режиме, бывают случаи, когда краулер ввиду программных ограничений не может загрузить определенный контент, в том числе, потоковое видео прямых эфиров и включений и некоторые типы файлов, тогда требуется сохранение этого контента вручную при наличии технической возможности для этого.

С методологической точки зрения, существует несколько основных подходов к сканированию – это сохранение отдельных элементов на страницах, самих страниц и их совокупностей, а также доменное сканирование. Выбор способа сохранения зависит от значимости сканируемого контента и масштаба проекта.

В веб-архиве могут быть сохранены отдельные элементы веб-страниц, например, таковыми могут быть аудио-, видео-, мультимедийный контент, записи социальных сетей и т.д. Постраничное сканирование используется, например, на платформе Wayback Machine, когда сканер может не сканировать сайт на всю глубину его иерархии, а захватывать только верхние уровни или даже отдельные страницы. Амбициозная цель проекта – отсканировать весь веб, но на практике это неосуществимо на данный момент. Сканер Интернет-архива сканирует и сохраняет отдельные снапшоты с некоторой периодичностью. Время, когда были сделаны снапшоты конкретной страницы, представлены на динамической инфографике, формируемой для каждого сохраняемого сайта (рис. 1).

Очередность и приоритеты по сохранению отдельных страниц и сайтов могут зависеть и от пользователей. Так, при выборе ресурсов для веб-архивирования учитываются голоса пользователей, которые предлагают

сохранять те или иные веб-сайты. Такую опцию предлагает, в том числе, и Internet Archive, которая размещается под кнопкой “Save Page Now” на платформе веб-архива.

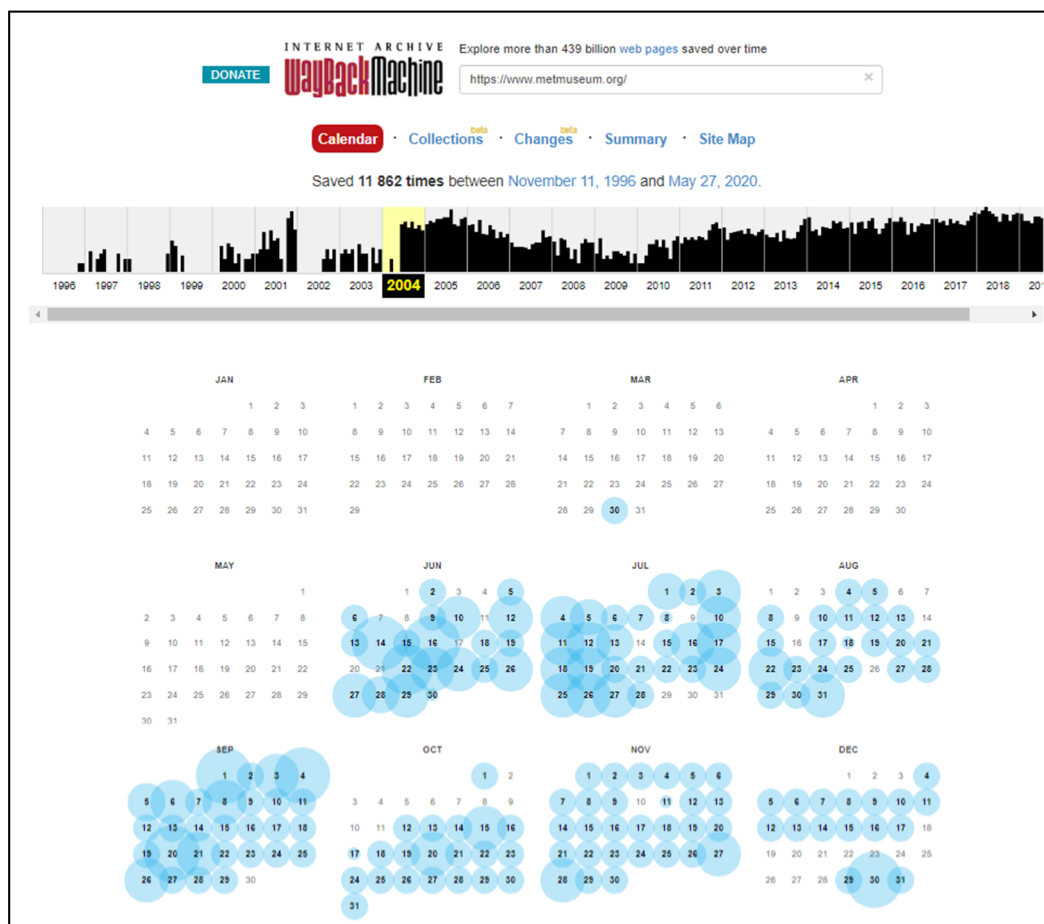


Рис. 1. Календарь сканирования страницы <https://www.metmuseum.org/> на платформе Wayback Machine

Формальные критерии отбора для сканирования могут быть относительно легко автоматизированы, например, возможно переложить на язык программного кода формат видеофайлов или блог как тип контента. Однако определение тематики коллекции требует больших усилий и экспертного участия человека в этом процессе. Создание коллекций требует формирования перечня релевантных выбранной теме веб-сайтов и страниц, согласно которому будет осуществляться сканирование. Поэтому в ручном режиме довольно сложно составить исчерпывающий список ресурсов, и краулер не охватит сайты, появившиеся после запуска программного кода и не вошедшие в список сканирования.

Перспективным в плане улучшения качества веб-архивирования является применение технологий искусственного интеллекта и, в первую



очередь, машинного обучения<sup>69</sup>, что позволит автоматизировать процесс отбора материалов для последующего сканирования и формирования веб-архивных коллекций.

### 3.2. Сохранение

Веб-архив сохраняет снимки веб-страниц, которые не следует путать со скриншотами ввиду существенной разницы между этими понятиями. Скриншот – это изображение экрана (видимого на мониторе фрагмента веб-страницы), тогда как снимок представляет собой копию веб-страницы, с которой можно осуществлять действия – переходить по ссылкам, копировать текст и изображения. Таким образом, преимущество снимка перед скриншотом очевидно – снимок сохраняет динамичность контента с «рабочими» ссылками, копируемым текстом и другим содержимым в противовес статичности формата изображения скриншота.

Важной задачей веб-архивирования является сохранение отсканированных сайтов в их изначальном виде, обеспечение условий, когда изменения, модификации и другие манипуляции для изменения структуры и содержания заархивированного контента становятся невозможными. То есть задача веб-архивов – обеспечивать целостность и аутентичность сохраняемого веб-контента.

Ресурс в веб-архиве сохраняется в комплексе с метаданными, такими как адрес в сети Интернет (URL), дата и время сканирования, количество отсканированных страниц и размер контента (в байтах, мегабайтах, гигабайтах и т.д.). Важно сохранение и архитектуры сайта, в том числе иерархической взаимосвязи компонентов, которая определяется гиперссылками и их размещением по дереву сайта. Начальной единицей сохранения веб-сайта является веб-страница. Ее положение в структуре веб-сайта определяется ссылками – *внутренними*, ведущими на другие страницы, а также *внешними*, которые ведут на эту страницу с других страниц. Совокупность страниц в их взаимосвязи образует структуру сканируемого контента. При повторном сканировании, краулер может распознать изменение структуры веб-сайта и отразить это изменение в сохраненных данных.

---

<sup>69</sup> Ayala B. R., Caragea C. (2015) Towards building a collection of web archiving research articles. Proceedings of the American Society for Information Science and Technology. URL: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101150> (дата обращения: 01.05.2022).

### 3.3. Документирование

Традиционный подход к документированию архивных данных не может в полной мере применяться в веб-архивировании, поскольку не предназначен для сохранения информации о гипертекстовых связях и динамичном контенте. Поэтому чаще всего для этих целей используется особая система иерархичного описания. Online Computer Library Center (OCLC)<sup>70</sup>, научно-исследовательская организация, целью которой является продвижение инноваций и лучших практик в библиотеках, разработала систему рекомендаций по метаописанию для материалов веб-архивов<sup>71</sup>. При разработке этих рекомендаций, перед OCLC стояла задача разработать систему, нейтральную по отношению к существующим стандартам описания (в т.ч. MARC и Dublin Core), но в то же время, чтобы эту систему могли применять организации, уже использующие эти стандарты. Метаописание необходимо для документирования ресурсов и обеспечения поисковых возможностей. При этом высокая степень детализации метаданных и точности описания являются основой для получения более релевантных результатов при осуществлении запросов пользователями при поиске контента.

В большинстве случаев генерирование метаданных происходит в автоматическом режиме во время сканирования и на основании данных сохраняемого контента. Метаданные содержат информацию о размере отсканированных материалов, дате и времени, адресах, типах контента (текст, изображение и т.д.). Часть метаданных экстрактируется краулером из тэгов, опубликованных на страницах (при их наличии). Дополнительно могут извлекаться ключевые слова на основе сохраненных текстов, что в дальнейшем может применяться для классификации ресурсов, определении принадлежности к тематическим коллекциям и оптимизации поисковых возможностей.

Некоторые веб-архивы, как например, ресурс, разрабатываемый Тайваньским Национальным Университетом, реализует широкий диапазон метаданных, позволяющий документировать и вторичную информацию – комментарии пользователей, оценки, рейтинги и т.д. Такой подход

---

<sup>70</sup> Online Computer Library Center (OCLC). URL: <https://www.oclc.org/research/about.html> (дата обращения: 01.05.2022).

<sup>71</sup> Descriptive Metadata for Web Archiving. Online Computer Library Center. URL: <https://www.oclc.org/research/publications/2018/oclcresearch-descriptive-metadata/recommendations.html> (дата обращения: 01.05.2022).

дает возможность использовать эти сведения для более детального анализа заархивированных веб-ресурсов<sup>72</sup>.

Интернет-архив оперирует довольно большим набором метаданных в системе веб-архива. Схема метаданных доступна на платформе проекта<sup>73</sup> и описывает следующие характеристики: идентификатор, тип медиа, название, коллекция, описание, загрузчик, тип объекта, дата создания, автор, язык, временной диапазон, дата публикации, дата сканирования, размер, спонсор, сканер, сканирующая организация, источник, доступ, распознавание, издатель и другие. В интернет-архиве также предусмотрены наборы тегов для курирования коллекций, описания файлов различного формата (например, видео- и аудиофайлы), документов разного типа (в том числе книг), сведений об авторских правах.

### 3.4. Обеспечение доступа

Обеспечение доступа к материалам веб-архивов предполагает создание условий, в которых возможно использование данных ресурсов пользователями, их поиск и просмотр. Доступ к сохраненным ресурсам может быть глобальным (реализующимся посредством открытых сетей) и ограниченным (локальным – требующим авторизации, либо определенной аффилиации).

Открытый доступ к материалам национальных веб-архивов обеспечивается в Австралии<sup>74</sup>, Новой Зеландии<sup>75</sup>, Хорватии<sup>76</sup> и некоторых других странах. Распространенной практикой является предоставление доступа к материалам исключительно через локальные сети организаций, обеспечивающих (контролирующих) веб-архивирование, как например, в Венгрии и Швеции.

Ограничение доступа чаще всего связано с законодательством об авторском праве, которое должно учитываться при веб-архивировании. По-

---

<sup>72</sup> Chen K., Chen Y., Ting P. (2008) Developing National Taiwan University Web Archiving System. Proceedings of International Web Archiving Workshop. URL: <http://ntur.lib.ntu.edu.tw/retrieve/168049/> (дата обращения: 01.05.2022).

<sup>73</sup> Internet Archive Metadata. Internet Archive. URL: <https://archive.org/services/docs/api/metadata-schema/index.html> (дата обращения: 01.05.2022).

<sup>74</sup> TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au> (дата обращения: 01.05.2022).

<sup>75</sup> New Zealand Web Archive. URL: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive> (дата обращения: 01.05.2022).

<sup>76</sup> Hrvatski Arhiv Web. URL: <https://haw.nsk.hr> (дата обращения: 01.05.2022).

этому на практике, чтобы избежать нарушений, доступ может быть лимитирован по отношению ко всему контенту, либо могут быть сделаны исключения для материалов, свободных от авторского права или расположенных в публичном домене как часть всеобщего достояния. Так, в Великобритании часть национального заархивированного веба, свободная от авторских прав на цифровой контент или являющаяся частью публичного пространства, размещена онлайн с возможностью просмотра данных материалов через сайт веб-архива<sup>77</sup>. Другая часть, более объемная, может быть изучена только через локальную сеть авторизованных для этих целей учреждений (преимущественно это библиотеки Соединенного королевства). Примечательна в этом плане и политика Сингапура в отношении архивного веба – открытый доступ к контенту может быть предоставлен, в тех случаях, когда владельцы веб-сайтов дали на это согласие<sup>78</sup>.

Доступ к веб-архивным материалам, открытым для использования онлайн, осуществляется с помощью простого браузера, а поисковые системы предлагают две основные опции – это поиск по адресам в сети Интернет (URL) и названиям веб-ресурсов. Многие веб-архивы устроены таким образом, что пользователь должен знать URL сайта (или веб-страницы), который требуется найти в веб-архиве. Если же пользователю неизвестен адрес искомой страницы или в его задачу не входит искать информацию на каком-то определенном сайте, либо необходимо осуществлять поиск более сложного характера, то могут возникнуть трудности при работе с веб-архивом, что существенным образом лимитирует доступность материалов.

Одним из ограничений большинства веб-архивов является зависимость от URL и отсутствие контекстного поиска по сохраненным веб-сайтам. В ходе изучения поисковых алгоритмов веб-архивов Анат Бен-Давид и Хьюго Хурдеман показали необходимость создания сервиса с возможностями контекстного поиска<sup>79</sup>. В ходе реализации проекта WebArt (2012–2016) был создан сервис WebART: Web Archive Retrieval Tools<sup>80</sup>,

---

<sup>77</sup> The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk/> (дата обращения: 01.05.2022).

<sup>78</sup> Web Archive Singapore. URL: <https://eresources.nlb.gov.sg/webarchives/landing-page> (дата обращения: 01.05.2022).

<sup>79</sup> Ben-David A., Huurdeman H. (2014) Web Archive Search as Research: Methodological and Theoretical Implications // *Alexandria*, Vol. 25(1). Pp. 93–111, DOI: 10.7227/ALX.0022

<sup>80</sup> WebArt: Web Archive Retrieval Tools. URL: <http://www.webarchiving.nl> (дата обращения: 01.05.2022).

открывающий широкие поисковые возможности для работы с веб-архивом Нидерландов. WebArt по сути является tool-driven исследованием, которое позволило понять больше о содержании нидерландского веб-архива с помощью созданного сервиса. Однако доступ к разработанному инструментарию ограничен, что делает невозможным его использование другими исследователями.

Поскольку основные поисковые решения не позволяют удовлетворять запросы большинства пользователей в навигации по сохраненному контенту, веб-архивные проекты стали включать тематические коллекции. Таковым, например, является платформа Archive-It<sup>81</sup> (см. Гл. 5).

### **3.5. Сложности веб-архивирования и их негативное влияние на выборку**

Какая бы стратегия ни находилась в основе процессов веб-архивного проекта – всегда будут те или иные ограничивающие факторы, которые могут быть связаны с подходами к сканированию, формированию коллекций, техническими или иными сложностями. И эти факторы необходимо уметь определять для проведения качественной источниковедческой критики. Перечислим далее эти ограничители:

Ранее уже обращалось внимание на то, что в веб-архивах всегда представлено ограниченное количество заархивированных материалов. При изучении веб-пространства и его динамики в исторической ретроспективе чрезвычайно важно обладать данными об истоках тех или иных процессов, явлений, проследивать эволюцию с максимально возможной степенью детализации как в пространстве и содержании, так и во времени.

Одна из ключевых проблем, связанных с архивным вебом и его использованием, заключается в том, что многие веб-материалы исчезают, так и не оказавшись заархивированными. На данный момент существует разнообразие программ по веб-архивированию, но все же прилагаемых усилий недостаточно. Веб-архивы до сих пор характеризуются высокой степенью централизации, а сохранение данных о 1990-х гг. истории веба сконцентрировано в основном в хранилищах двух веб-архивов – Интернет-архива и ресурса TROVE Австралийской национальной библиотеки. И если первый ресурс представляет собой глобальный веб-архив, о чем говорилось ранее, то второй сфокусирован на национальном сегменте ав-

---

<sup>81</sup> Archive-It. URL: <https://archive-it.org/> (дата обращения: 01.05.2022).

стралийского веб-пространства и его истории. Было проведено исследование на предмет проверки предвзятости данных на примере Интернет-архива<sup>82</sup>. В этом исследовании сравнивалась репрезентация разных типов ресурсов из разных сегментов национальных доменов. Результаты показали, что в Интернет-архиве несбалансированность представленных данных очень существенная. Например, было сохранено 92% коммерческих сайтов США и только 58% – Китая (из выборки).

Другой пример исследования полноты представленной в веб-архиве выборки связан с анализом URL-адресов страниц, сохраненных в избранном на одном из сервисов. На основе использования инструментов Memento Api удалось показать диапазон сохраненных страниц, который составил от 35 до 90%<sup>83</sup>. Причем было выявлено существенное влияние посещаемости страниц на вероятность архивирования. Чем менее популярен контент, тем меньше вероятность, что он будет сохранен в веб-архиве. Гипотеза о влиянии посещаемости также подтвердилась на примере архивирования сайта TripAdvisor – Интернет-архив сохранил только 24% от всего объема сайта на определенный момент времени. Кроме того, была обнаружена предвзятость в сторону сохранения тех веб-страниц на ресурсе, которые представляют наиболее известные локации.

Можно ли веб сохранить полностью? Ответ на этот вопрос очевиден. Веб огромен, чтобы его можно было автоматически измерить даже для того, чтобы подсчитать необходимые ресурсы для полного сохранения. И даже если будет создана программа, способная измерить всю Всемирную паутину, данные измерений все равно будут неполными ввиду постоянно увеличивающегося количества веб-ресурсов и разрастания информации на них. Поэтому веб-архивы характеризуются неполнотой, какая-то часть веба всегда отсутствует. Поэтому этот фактор необходимо принимать во внимание, оценивая выборку при проведении исследования.

Иным фактором, препятствующим сохранению веба в его аутентичном виде, является **динамичность ресурсов**. Она проявляется в быстрой трансформации сведений, редактировании и дополнении данных, когда одни сведения сменяются другими, удаляются и постоянно модифицируются. Высокой степенью динамизма отличаются, например, новостные

---

<sup>82</sup> Hale S. A., Blank G., Alexander V. D. (2017) Live versus archive: Comparing a web archive to a population of web pages. In: Niels Brügger, Ralph Schroeder (Eds.) The Web as History. Using Web Archives to Understand the Past and the Present. Pp. 45-61.

<sup>83</sup> Там же.

проекты, блоги, форумы, чаты и комнаты для дискуссий и другие ресурсы. Поэтому при использовании традиционных подходов к сохранению веба существенная часть материалов между очередными сканированиями просто исчезает. Динамичность контента также свойственна социальным медиа, поскольку пользователи легко могут создавать, редактировать, обновлять, удалять посты, изображения, видео и иной контент и производить многие другие действия.

На данный момент краулеры не в состоянии в краткие сроки создать качественную архивную копию крупного веб-сайта, который меняется очень активно, то есть имеет динамичный характер. Пока программа сканирует сайт, страницы, отсканированные в самом начале, могут уже измениться в живом вебе. Так, в процессе веб-архивирования динамичность как ключевое свойство Всемирной паутины подменяется статичностью снапшота, что в корне меняет саму суть контента, а не только рамки существования. В итоге получается архивная копия сайта, который в таком виде никогда не существовал. Можно утверждать, что веб-архивы производят «химер», и это обстоятельство необходимо иметь в виду, проводя исследования в области веб-истории.

Одной из первых попыток сохранения динамичного веба стал проект датского веб-архива Netarchive<sup>84</sup>, реализуемый на основе закона об обязательном экземпляре (Danish Legal Deposit Law). С 2005 г. этот проект реализуется Королевской библиотекой Дании и сочетает несколько подходов к веб-архивированию, в том числе, тематическое архивирование, основанное на необходимости сохранить сведения об определенных событиях, и специальное. Последняя из перечисленных стратегий ориентирована на сохранение тех ресурсов, изменения которых трудно отследить – например, если ресурс содержит видеоряды и потоковое видео, которое трудно зафиксировать с технологической точки зрения. Кроме того, к этой стратегии относится отбор тех ресурсов, которые должны быть вскоре закрыты (например, если проект заканчивается), но содержат важный материал, а также ресурсы, рекомендованные пользователями к сохранению. Следует подчеркнуть, что в проекте Netarchive принята стратегия ежедневного сканирования целого перечня сайтов для их архивирования, что позволяет смягчить проблему динамичности веба. В частности, в этот список входят новостные порталы, сайты политических партий, организаций и ассоциа-

---

<sup>84</sup> Danish Web Archive. The Netarchive. URL: <http://netarkivet.dk> (дата обращения: 01.05.2022).

ций, министерств и общественных советов, а также отдельные персональные аккаунты в социальных сетях<sup>85</sup>.

**Закрытость** и лимитированность доступа к ресурсам также является проблемой для веб-архивирования. В «живом» веб-пространстве находится ограниченное количество ресурсов, доступ к которым полностью открыт, многие другие ресурсы требуют авторизации, регистрации или подписки для использования (часто платной). Перечисленные ограничения также препятствуют реализации программ веб-архивирования в более полном виде. Разработки последнего времени позволяют отчасти преодолеть это ограничение. Так, сервис Conifer<sup>86</sup> (ранее известный как Webrecorder.io) позволяет сохранять веб-страницы и веб-сайты и их содержимое после авторизации на сохраняемом ресурсе, а затем распространять заархивированный материал без раскрытия учетных данных. Существенным ограничением является то, что этот ресурс предназначен скорее для индивидуального использования и создания частных репозиториев, чем для веб-архивирования в его широком смысле.

Существенная сложность связана и с **метаданными**, их перечнем, разнообразием, широтой, что непосредственно влияет на последующие возможности использования сохраненных ресурсов и навигации по ним. Часто в автоматическом режиме удастся сформировать очень ограниченный список метаданных, связанных в основном с физическими характеристиками отсканированного материала (как, например, количество URL-ссылок или размер собранных текстов и изображений). Поэтому требуется дополнительная работа с метаданными после сканирования, их курирование и существенная работа по их детализации в «ручном» режиме. Исследователи проекта ARCOMEM разработали инструментарий, связанный с обработкой семантики сканируемой информации и извлечения дополнительной информации из заголовков и содержания сохраняемых веб-сайтов<sup>87</sup>. Такой подход позволяет обогатить метаданные сведениями о

---

<sup>85</sup> The Netarchive. URL: <http://netarkivet.dk/om-netarkivet> (дата обращения: 01.05.2022).

<sup>86</sup> Conifer. Collect and revisit web pages. URL: <https://conifer.rhizome.org> (дата обращения: 01.05.2022).

<sup>87</sup> Thomas R., Demidova E., Dietze S., Peters W., Papailiou N., Doka K., Stavrakas Ya., Plachouras V., Senellart P., Carpentier F., Mantrach A., Cautis B., Siehndel P., Spiliotopoulos D. (2014) The ARCOMEM Architecture for Social- and Semantic-Driven Web Archiving // Future Internet. Vol. 6. Pp. 688–716. URL: <http://www.mdpi.com/1999-5903/6/4/688/htm> (дата обращения: 01.05.2022).



контенте веб-страниц, их тематике, событиях и вопросах, которые там поднимаются.

**Разнообразие видов веб-ресурсов** и множество платформ с отличающимся назначением и функциями также осложняют процесс их сохранения. Виртуальное пространство и его социально-культурные сегменты очень гибкие, а многие веяния в искусстве и культуре только начинают свое становление. К таким ресурсам, например, относятся сайты, связанные с цифровым искусством, созданием и публикацией электронных объектов. Для сохранения этих платформ достаточно сложно сформулировать требования для сканирования, выделить ключевые метаданные ввиду вариативности ресурсов, форматов, жанров и их структур.

Веб-ресурсы могут обладать **гетерогенным характером** и многократно дублировать уже существующую информацию. Это происходит из-за того, что не существует единого или сколько-то ограниченного количества «точек» загрузки контента. Так пользователи могут размещать одинаковые ресурсы на различных сайтах и платформах, контент может копироваться, переписываться и многократно дублироваться на разных ресурсах.

Другой сложностью при веб-архивировании и последующем использовании веб-ресурсов является проблема **авторизованности** организаций и лиц, создающих тот или иной контент. Понимание происхождения сайта, круга инициаторов, организаторов и разработчиков проекта является ключевым аспектом, связанным с доверием сайту и его контенту. В то же время, такие сведения не всегда содержатся на самих ресурсах, и не всегда ясно, кто является владельцем контента.

Ограничения **технического характера**, связанные со сканированием и сохранением ресурсов, распространены очень широко. Часто еще в процессе работы краулера собираются **«сломанные» ссылки** или ведущие на страницу с ошибкой. Неработающие ссылки сохраняются в таких случаях, когда гиперссылка на ресурсе направляла на внешние сайты, которые не были указаны в списке для сканирования, и поэтому они не были захвачены краулером. Другой распространенной причиной является **технический сбой при сканировании**. Проблемы технического характера негативно влияют на полноту собранной информации и возможности последующего воспроизведения сайтов.

Вариантом технической погрешности является ситуация, когда оператором назначена недостаточная **глубина сканирования сайта**. Сканирование на определенную глубину по иерархии сайта было широко распространенной практикой особенно на ранних стадиях развития веб-архивирования. Кроме того, зачастую краулеры не имеют задачи захватить весь контент страниц, а только данные определенного формата (например, тексты и изображения). Материалы остальных форматов не сканируются, что становится причиной появления «белых пятен» отсутствующей информации. Например, Интернет-архив сохранил снимок сайта Государственного музея изобразительных искусств им. А. С. Пушкина за 7 декабря 2000 г.<sup>88</sup>, где мы можем увидеть приглашение посетить виртуальный тур по выставке. Однако ознакомиться с самим туром невозможно, поскольку контент не был сохранен краулером. Так несохраненная информация впоследствии становится утраченной.

Другой существенной проблемой технического характера является сложность воспроизведения заархивированного контента. Для этих целей требуется **программное обеспечение**, подходящее для оперирования теми или иными файлами согласно их назначению. Многие форматы на данный момент уже устарели и зачастую уже не поддерживаются современными компьютерами. В результате получается, что контент сохранен, но это не означает, что он становится доступным. Сегодня эта проблема может быть частично решена на основе внедрения в структуру платформ веб-архивов различных эмуляторов (в т.ч. программного обеспечения, операционных систем, информационных сред, и т.д.). Так, инфраструктура Интернет-архива позволяет воспроизвести целый ряд устаревших программ и приложений. У пользователей появляется шанс ознакомиться с контентом, произведенным в 1980-1990-е гг. для распространения на разных носителях, таких как гибкие диски, гиперкарты и т.д. Например, эмулятор среды Macintosh Интернет-архива позволяет открыть содержимое гиперкарты, выпущенной в 1991 г. для воспроизведения в среде Macintosh, которое посвящено истории кельтов (см. рис. 2). При этом пользователь может ознакомиться с контентом полностью, в том числе прочитать тексты или проиграть аудиофайлы.

---

<sup>88</sup> Панорамы экспозиций. Снимок сайта Государственного музея изобразительных искусств им. А. С. Пушкина. Интернет-архив. URL: <http://web.archive.org/web/20001118051000/http://www.museum.ru/gmii/defrus.htm> (дата обращения: 01.05.2022).



Рис. 2. Эмуляция гиперкарты “Celtic History Museum”  
в среде Интернет-архива<sup>89</sup>

### 3.6. Социальные медиа и веб-архивирование: Значение сохранения, подходы и ограничения

Распространение социальных медиа произвело в определенном смысле революцию в коммуникации между людьми. В первую очередь это проявилось в том, что возникла возможность создавать и публиковать контент всем пользователям или лицам, прошедшим регистрацию и авторизацию на той или иной платформе, предназначенной для общения. С появлением социальных медиа возможность производить, публиковать и распространять контент стала настолько доступной, что главными требованиями для этих целей стали лишь доступ в сеть Интернет и наличие минимального уровня цифровой грамотности.

Социальные сети и платформы для неформальной коммуникации стали появляться еще с середины 1990-х гг. Одним из первых таких ресурсов стал американский Classmates.com, объединивший однокурсников и выпускников американских колледжей, который функционирует до сих

<sup>89</sup> Celtic History Museum. 11 November 1991. Internet Archive. URL: [https://archive.org/details/hypercard\\_celtic-history-museum](https://archive.org/details/hypercard_celtic-history-museum) (дата обращения: 01.05.2022).

пор. Затем появились и другие платформы – SixDegrees.com<sup>90</sup> (одна из первых глобальных социальных сетей), AsianAvenue.com<sup>91</sup> (предназначена для американцев китайского происхождения), многие из которых прекратили свое существование в настоящее время. На начало 2000-х гг. пришелся бум социальных сетей. Появляются такие сети, как MySpace и LinkedIn (2003 г.), Facebook (Meta)<sup>92</sup> (2004 г.), Youtube (2005 г.), ВКонтакте и Twitter (2006 г.), Instagram (2010 г.).

Аудитория социальных сетей по оценкам платформы Statista на конец 2020 г. составила 3.6 млрд чел. и по прогнозам превысит 4 млрд пользователей в 2023 г.<sup>93</sup> Не вызывает сомнения, что социальные сети являются значимой частью жизни современного общества, и без анализа их содержания реконструировать в будущем картину современной нам действительности вряд ли будет возможно во всей полноте.

Социальные медиа являются важной составляющей современной повседневности, представляют собой площадки для общения, места обмена информацией, формирования и потребления новостной повестки. Социальные медиа имеют существенное значение в структуре товаро-денежных отношений как маркетплейсы, фиксируют социально-культурные процессы и выполняют другие функции. Анализ социальных сетей в исторической ретроспективе позволяет больше понять формирующуюся общественную структуру современного типа и ее развитие, особенности бизнес-процессов, взаимодействия людей и сообществ и т.д.

Эволюция социальных сетей, расширение их разнообразия способствовали появлению множества разных по формату и типу цифровых «посланий» – текстовых сообщений разной длины, изображений, аудиомате-

---

<sup>90</sup> SixDegrees.com. Снапшоты ресурса доступны в Интернет-архиве. URL: <https://web.archive.org/web/19971210214122/http://www.sixdegrees.com/> (дата обращения: 01.05.2022).

<sup>91</sup> AsianAvenue.com. Снапшоты ресурса доступны в Интернет-архиве. URL: [https://web.archive.org/web/\\*/asianavenue.com](https://web.archive.org/web/*/asianavenue.com) (дата обращения: 01.05.2022).

<sup>92</sup> 28 марта 2022 г. Тверской районный суд г. Москвы вынес решение «запретить деятельность американской транснациональной холдинговой компании Meta Platforms Inc. по реализации продуктов-социальных сетей Facebook и Instagram на территории Российской Федерации по основаниям осуществления экстремистской деятельности». См. УИД 77RS0027-02-2022-005636-57, Дело № 02-2473/2022. Решение Именем Российской Федерации от 21 марта 2022 г. Тверской районный суд. Суды общей юрисдикции г. Москва. URL: <https://mosgorsud.ru/rs/tverskoj/cases/docs/content/b23582e0-ae54-11ec-9dc4-77eec1d7bd14> (дата обращения: 09.04.2022).

<sup>93</sup> Number of social network users worldwide from 2017 to 2025(in billions). Statista. URL: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users> (дата обращения: 01.05.2022).

риала, видеороликов, анимированного контента, которые дополнились реакциями и комментариями других пользователей и подписчиков, ссылками на сторонние материалы, хештегами и т.д. Все эти элементы социальных сетей так или иначе отражают современность и становятся важными свидетельствами истории.

Необходимо обратить внимание и на то, что социальные медиа оказывают существенное влияние на содержание веб-сайтов, могут его существенно редуцировать. Многие организации начинают создавать особый контент для социальных сетей, который не предназначен для публикации на сайтах. Следовательно, если не сохранена информация из социальных сетей – она будет с большой долей вероятности утрачена в дальнейшем, и невозможно будет реконструировать слои информации, связанные с этой организацией во всей полноте.

Социальные медиа как потенциальный исторический источник имеют свои особенности. Они более уязвимы к внешним воздействиям, и возможности сохранения очень ограничены. Эти ресурсы редко находятся в поле внимания глобальных неспециализированных программ по сохранению веба ввиду целого ряда причин, в числе которых жесткие требования законодательства о персональных данных, технологические лимиты для загрузки информации (закрытый характер API) и другие факторы. Поэтому много ценных сведений исчезает вместе с закрывающимися страницами и удаленными аккаунтами из социальных сетей.

К сожалению, сохранение материалов социальных сетей для будущего использования – процесс более сложный, чем сохранение обычного веба. Ранее указывалось на сложности, связанные с архивированием веб-ресурсов (см. Гл. 2). Проблемы авторизации на ресурсе, вопросы этики, персональных данных, их обработки и хранения актуальны для сохранения социальных сетей. Кроме того, проекты по их архивированию осложнены высокой степенью динамичности ресурсов, постоянным нарастанием информационных слоев, расширением контента, который непрерывно изменяется, удаляется, дополняется и корректируется.

Динамичность социальных сетей – это существенный вызов для веб-архивирования, поскольку веб-архив, сохраняя ресурс, делает его статичным, что в корне меняет саму специфику контента. Поэтому требуются особые подходы для сохранения социальных сетей.

Использование в исследованиях фрагментов веба и сведений из социальных сетей возможно и без использования веб-архивов. В этой ситуации возможен сбор отдельных сообщений в социальных сетях или даже целых аккаунтов. Некоторые медиа предоставляют открытый API для выкачивания необходимых постов, в первую очередь, следует сказать о платформе Twitter, открывшей API для исследовательских целей (подробнее об API см. далее). Однако этот сбор возможен лишь на тех материалах живого веба, которые не были удалены из сети.

Сохранение фрагментов информации и формирование отдельных наборов данных социальных медиа уже предпринималось ранее при проведении исследований. Сбор, организация и обработка данных социальных сетей проводится в рамках исследовательских проектов довольно широко как в области истории, так и политологии, социологии, культурологии, философии. Сохранение твитов, постов с Facebook или других социальных сетей стало важной частью практики историков, исследователей современности. Круг тем таких исследований довольно широк, и зачастую проекты носят междисциплинарный характер. Например, с использованием твитов как основного исторического источника было предпринято изучение последствий террористических атак в Ницце в 2015 г.<sup>94</sup>, проанализированы коммеморативные практики в отношении Первой мировой войны<sup>95</sup>, проведен сравнительный анализ твитов американских политиков во время президентства Дональда Трампа<sup>96</sup>, прослежена динамика постов с хештегами #blacklivesmatter и #alllivesmatter за несколько лет<sup>97</sup>, изучено использование платформы Twitter Д.А. Медведевым во время его президентства<sup>98</sup> и осуществлены другие исследования.

---

<sup>94</sup> Schafer V, Truc G, Badouard R, Castex L, Musiani F. Paris and Nice terrorist attacks: Exploring Twitter and web archives. *Media, War & Conflict*. 2019; 12(2):153–170. doi: 10.1177/1750635219839382. URL: <https://journals.sagepub.com/doi/pdf/10.1177/1750635219839382> (дата обращения: 01.05.2022).

<sup>95</sup> Clavert F. #ww1. The Great War on Twitter. In: *Proceedings of the conference DH2016*. URL: [https://halshs.archives-ouvertes.fr/halshs-01345568/file/dh2016\\_abstractCLAVERT.pdf](https://halshs.archives-ouvertes.fr/halshs-01345568/file/dh2016_abstractCLAVERT.pdf) (дата обращения: 01.05.2022).

<sup>96</sup> Schneider U. (2021) How Trump tweets: A comparative analysis of tweets by US politicians // *Research in Corpus Linguistics*, Vol. 9(2). Pp. 34–63.

<sup>97</sup> Anderson M. The hashtag #BlackLivesMatter emerges: Social activism on Twitter. *Social Media Conversations about Race*. Pew Research Center. Internet & Technology. August, 15, 2016. URL: <https://www.pewresearch.org/internet/2016/08/15/the-hashtag-blacklivesmatter-emerges-social-activism-on-twitter> (дата обращения: 01.05.2022).

<sup>98</sup> Иванова А.И. «Мы и Россия сегодня»: Лингвистический анализ Твиттер-платформы президента Дмитрия Медведева // *Политическая лингвистика*. 2011. 2(36). С. 104–107.

В большинстве из реализованных проектов твиты были получены из живого веба посредством сбора данных на основе тематических хештегов или анализа массива текстов отдельных аккаунтов (как в случае с Д. Трампом, например). Принимая во внимание ценность материалов и особенность социальных сетей, в которых удалить контент еще легче, чем создать его, можно утверждать, что подобные ресурсы в высшей степени уязвимы, подвержены исчезновению и требуют особых усилий по их документированию и архивированию. Поэтому частные репозитории не выглядят достаточными для долгосрочного сохранения этих источников, и необходимы систематические программы по архивированию контента социальных сетей. Кроме того, веб-архивирование существенно отличается от частных репозиториях необходимостью обеспечить доступ общественности к контенту, естественно с соблюдением требований законодательства, этики и прав правообладателей.

Архивирование такого типа веб-ресурсов как социальные сети началось не так давно ввиду особых требований к процессу. Одним из проектов является BeSocial<sup>99</sup>, направленный на сохранение социальных медиа Бельгии. Главной идеей проекта является создание и развитие устойчивой стратегии по веб-архивированию и сохранению социальных сетей и их контента, отражающих события современной истории страны. Основным фокусом стало архивирование тех аккаунтов в социальных сетях, которые связаны с различными важными событиями в Бельгии, общественными и культурными мероприятиями. Также в перечень для сохранения вошли новостные страницы.

По сравнению с существующими социальными сетями, более масштабные усилия по архивации контента предпринимает Twitter. С одной стороны, в этой социальной сети есть возможность архивирования контента личного аккаунта, что реализовано и на других платформах, как ВКонтакте, например. Но с другой стороны, Twitter создает собственный архив контента. В январе 2021 г. было объявлено о том, что указанный архив будет открыт общественности для изучения<sup>100</sup>.

---

<sup>99</sup> BESOCIAL – towards a sustainable social media archiving strategy for Belgium. URL: <https://www.ugent.be/mict/en/research/besocial-towards-a-sustainable-social-media-archiving-strategy-for-belgium> (дата обращения: 01.05.2022).

<sup>100</sup> Statt N. Twitter is opening up its full tweet archive to academic researchers for free. To help researchers better study online discourse and platform trends. 26 January, 2021. URL: <https://www.theverge.com/2021/1/26/22250203/twitter-academic-research-public-tweet-archive-free-access> (дата обращения: 01.05.2022).

В то же время, целый ряд крупных программ по веб-архивированию включает архивирование отдельных аккаунтов социальных сетей. Так, АНО «Информационная культура» осуществляет сбор и архивирование официальных аккаунтов российских государственных структур в Twitter и отдельные персональные аккаунты (например, Дмитрия Медведева, во время его пребывания на посту президента РФ)<sup>101</sup>. Датасеты этого проекта доступны для просмотра и скачивания в открытом цифровом репозитории<sup>102</sup>. Библиотека Конгресса США также осуществляет архивирование «значимых с исторической точки зрения» ресурсов сети Twitter<sup>103</sup>. В Интернет-архиве также представлены подборки материалов из социальных сетей (например, сохранены данные, связанные с выборами в России и Иране<sup>104</sup>).

Однако архивирование материалов социальных сетей возможно лишь в том случае, если позволяет архитектура самой социальной сети. Twitter имеет открытый API (application programming interface или интерфейс прикладного программирования)<sup>105</sup> – описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой. Таким образом, реализована сама возможность сбора данных. Другие социальные сети могут быть закрытыми в этом отношении. Так, невозможным является архивирование сторонних аккаунтов в Facebook, если нет авторизованного доступа. На значимость данных Facebook как одной из крупнейших в мире социальных сетей уже обращали внимание исследователи и отмечали, что архитектура этой сети принципиально не позволяет со-

---

<sup>101</sup> Национальный цифровой архив России. URL: <https://ruarhive.org> (дата обращения: 01.05.2022).

<sup>102</sup> Список всех аккаунтов в Twitter, подвергаемых регулярной архивации в Национальном цифровом архиве России. URL: <https://hubofdata.ru/dataset/govtwitters-archives/resource/312580af-725a-4feb-9665-663e14b573ea> (дата обращения: 01.05.2022).

<sup>103</sup> Zimmer M. The Twitter Archive at the Library of Congress: Challenges for information practice and information policy. First Monday, Volume 20, Number 7–6 July 2015. URL: <https://firstmonday.org/ojs/index.php/fm/article/download/5619/4653> (дата обращения: 01.05.2022).

<sup>104</sup> Twitter Elections Integrity Dataset. Internet Archive. URL: <https://archive.org/details/twittersdataarchive> (дата обращения: 01.05.2022).

<sup>105</sup> API содержит описание способов взаимодействия одной компьютерной программы с другой, если API является открытым, то это взаимодействие становится понятным и значит, возможным, то есть становится очевиден путь для извлечения данных.



хранять данные<sup>106</sup>. По своей сути отношение к данным со стороны Facebook представляет собой их монополизацию или иначе – колониализм данных (data colonialism)<sup>107</sup>. В этом случае ставшие традиционными подходы к сбору данных через API не являются реализуемыми и требуются другие методы для извлечения данных из подобных закрытых систем.

Существенным достоинством сохраненного в веб-архивах контента социальных медиа является создание постоянных ссылок, которые можно использовать для формирования библиографии и обращению к ресурсу в долгосрочной перспективе. В целом, это является характерной чертой любых веб-архивных ресурсов.

Априори социальные сети и их контент – значимый источник для изучения современной истории. Однако большим вызовом для современных исследователей является не только сбор и сохранение, но и организация данных для исследовательских целей и последующая интерпретация результатов, полученных при обработке этих данных. Социальные медиа имеют сложную структуру и неочевидную систему взаимосвязей отдельных фрагментов сети, а также сложное переплетение вторичной информации в виде реакций пользователей.

Неоднозначным является и воздействие социальных сетей на офлайн-жизнь: требуют ответа вопросы об их влиянии на современного человека и общество, каким образом это воздействие изменялось, какие факторы становились более актуальными во времени и многие другие. Поэтому изучение взаимодействия индивидов, сообществ, организаций и других акторов все еще требует пристального внимания, выработки подходящей методологии для анализа.

---

<sup>106</sup> Ben-David A. Counter-archiving Facebook // *European Journal of Communication*. 2020; 35(3):249-264. doi: 10.1177/0267323120922069. URL: <https://journals.sagepub.com/doi/full/10.1177/0267323120922069>.

<sup>107</sup> Couldry N. Mejias, UA (2019) Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media* 20(4): 336–349. URL: <https://journals.sagepub.com/doi/pdf/10.1177/1527476418796632>.

## Рекомендуемая литература

1. Иванова А. И. «Мы и Россия сегодня»: лингвистический анализ Твиттер-платформы президента Дмитрия Медведева // Политическая лингвистика. 2011. 2(36). С. 104–107.
2. Anderson M. The hashtag #BlackLivesMatter emerges: Social activism on Twitter. Social Media Conversations about Race. Pew Research Center. Internet & Technology. 15 August 2016. URL: <https://www.pewresearch.org/internet/2016/08/15/the-hashtag-blacklivesmatter-emerges-social-activism-on-twitter> (дата обращения: 01.05.2022).
3. Ayala B. R., Caragea C. (2015) Towards building a collection of web archiving research articles. Proceedings of the American Society for Information Science and Technology. URL: <https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.2014.14505101150> (дата обращения: 01.05.2022).
4. Ben-David A. (2020) Counter-archiving Facebook // European Journal of Communication. Vol. 35(3). Pp. 249–264. URL: <https://journals.sagepub.com/doi/full/10.1177/0267323120922069>, DOI: 10.1177/0267323120922069 (дата обращения: 01.05.2022).
5. Ben-David A., Huurdeman H. (2014) Web Archive Search as Research: Methodological and Theoretical Implications // Alexandria, Vol. 25(1). Pp. 93–111, URL: <https://journals.sagepub.com/doi/10.7227/ALX.0022>, DOI: 10.7227/ALX.0022 (дата обращения: 01.05.2022).
6. Chen K., Chen Y., Ting P. (2008) Developing National Taiwan University Web Archiving System. Proceedings of International Web Archiving Workshop. URL: <http://ntur.lib.ntu.edu.tw/retrieve/168049/> (дата обращения: 01.05.2022).
7. Clavert F. #ww1. The Great War on Twitter. In: Proceedings of the conference DH2016. URL: [https://halshs.archives-ouvertes.fr/halshs-01345568/file/dh2016\\_abstractCLAVERT.pdf](https://halshs.archives-ouvertes.fr/halshs-01345568/file/dh2016_abstractCLAVERT.pdf) (дата обращения: 01.05.2022).
8. Couldry N. Mejas, UA (2019) Data colonialism: Rethinking big data's relation to the contemporary subject // Television & New Media. Vol. 20(4). Pp. 336–349. URL: <https://journals.sagepub.com/doi/pdf/10.1177/1527476418796632> (дата обращения: 01.05.2022).
9. Hale S. A., Blank G., Alexander V. D. (2017) Live versus archive: Comparing a web archive to a population of web pages. In: Niels Brügger, Ralph Schroeder (Eds.) The Web as History. Using Web Archives to Understand the Past and the Present. Pp. 45–61.
10. Schafer V., Truc G., Badouard R., Castex L., Musiani F. (2019) Paris and Nice terrorist attacks: Exploring Twitter and web archives // Media, War & Conflict. Vol. 12(2). Pp. 153–170. DOI: 10.1177/1750635219839382. URL:

<https://journals.sagepub.com/doi/pdf/10.1177/1750635219839382> (дата обращения: 01.05.2022).

11. Schneider U. (2021) How Trump tweets: A comparative analysis of tweets by US politicians // *Research in Corpus Linguistics*, Vol. 9(2). Pp. 34–63. URL: [http://ricl.aelinco.es/first-view/145-Article%20Text-525-1-4-2020\\_0203.pdf](http://ricl.aelinco.es/first-view/145-Article%20Text-525-1-4-2020_0203.pdf) (дата обращения: 01.05.2022).

12. Statt N. Twitter is opening up its full tweet archive to academic researchers for free. To help researchers better study online discourse and platform trends. 26 January, 2021. URL: <https://www.theverge.com/2021/1/26/22250203/twitter-academic-research-public-tweet-archive-free-access> (дата обращения: 01.05.2022).

13. Thomas R., Demidova E., Dietze S., Peters W., Papailiou N., Doka K., Stavarakas Ya., Plachouras V., Senellart P., Carpentier F., Mantrach A., Cautis B., Siehndel P., Spiliotopoulos D. (2014) The ARCOMEM Architecture for Social- and Semantic-Driven Web Archiving // *Future Internet*. Vol. 6. Pp. 688–716. URL: <http://www.mdpi.com/1999-5903/6/4/688/htm> (дата обращения: 01.05.2022).

14. Zimmer M. The Twitter Archive at the Library of Congress: Challenges for information practice and information policy. *First Monday*, Volume 20, Number 7–6 July 2015. URL: <https://firstmonday.org/ojs/index.php/fm/article/download/5619/4653> (дата обращения: 01.05.2022).

## **Источники**

1. Национальный цифровой архив России. URL: <https://ruarxive.org> (дата обращения: 01.05.2022).

2. Панорамы экспозиций. Снапшот сайта Государственного музея изобразительных искусств им. А. С. Пушкина. URL: <http://web.archive.org/web/20001118051000/http://www.museum.ru/gmii/defrus.htm> (дата обращения: 01.05.2022).

3. Список всех аккаунтов в Twitter, подвергаемых регулярной архивации в Национальном цифровом архиве России. URL: <https://hubofdata.ru/dataset/govtwitters-archives/resource/312580af-725a-4feb-9665-663e14b573ea> (дата обращения: 01.05.2022).

4. Archive-It. URL: <https://archive-it.org/> (дата обращения: 01.05.2022).

5. AsianAvenue.com. Снапшоты ресурса доступны в Интернет-архиве. URL: [https://web.archive.org/web/\\*/asianavenue.com](https://web.archive.org/web/*/asianavenue.com) (дата обращения: 01.05.2022).

6. BESOCIAL – towards a sustainable social media archiving strategy for Belgium. URL: <https://www.ugent.be/mict/en/research/besocial-towards-a-sustainable-social-media-archiving-strategy-for-belgium> (дата обращения: 01.05.2022).

7. Celtic History Museum. 11 November 1991. Internet Archive. URL: [https://archive.org/details/hypercard\\_celtic-history-museum](https://archive.org/details/hypercard_celtic-history-museum) (дата обращения: 01.05.2022).
8. Conifer. Collect and revisit web pages. URL: <https://conifer.rhizome.org> (дата обращения: 01.05.2022).
9. Danish Web Archive. The Netarchive. URL: <http://netarkivet.dk> (дата обращения: 01.05.2022).
10. Descriptive Metadata for Web Archiving. Online Computer Library Center. URL: <https://www.oclc.org/research/publications/2018/oclcresearch-descriptive-metadata/recommendations.html> (дата обращения: 01.05.2022).
11. Hrvatski Arhiv Web. URL: <https://haw.nsk.hr> (дата обращения: 01.05.2022).
12. Internet Archive Metadata. Internet Archive. URL: <https://archive.org/services/docs/api/metadata-schema/index.html> (дата обращения: 01.05.2022).
13. New Zealand Web Archive. URL: <https://natlib.govt.nz/collections/a-z/new-zealand-web-archive> (дата обращения: 01.05.2022).
14. Number of social network users worldwide from 2017 to 2025(in billions). Statista. URL: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users> (дата обращения: 01.05.2022).
15. Online Computer Library Center (OCLC). URL: <https://www.oclc.org/research/about.html> (дата обращения: 01.05.2022).
16. SixDegrees.com. Снапшоты ресурса доступны в Интернет-архиве. URL: <https://web.archive.org/web/19971210214122/http://www.sixdegrees.com/> (дата обращения: 01.05.2022).
17. The Netarchive. URL: <http://netarkivet.dk/om-netarkivet> (дата обращения: 01.05.2022).
18. The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk/> (дата обращения: 01.05.2022).
19. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au> (дата обращения: 01.05.2022).
20. Twitter Elections Integrity Dataset. Internet Archive. URL: <https://archive.org/details/twittersdataarchive> (дата обращения: 01.05.2022).
21. Web Archive Singapore. URL: <https://eresources.nlb.gov.sg/webarchives/landing-page> (дата обращения: 01.05.2022).
22. WebArt: Web Archive Retrieval Tools. URL: <http://www.webarchiving.nl> (дата обращения: 01.05.2022).

## Вопросы для самопроверки

1. Как осуществляется сбор данных для веб-архивирования?
2. Перечислите основные подходы к сохранению веба.
3. Назовите уровни сохранения веб-контента.
4. Как определяется очередность и приоритеты по сохранению отдельных страниц и сайтов?
5. Какие существуют сервисы и инструменты для определения очередности и приоритетов сохранения?
6. Дайте определение термина снапшот.
7. В чем заключается различие между скриншотом и снапшотом веб-документа?
8. Какие компоненты метаданных можно выделить при веб-архивировании?
9. Какие существуют системы метаописания для веб-архивов?
10. Перечислите основные компоненты метаданных, которые могут экстрагироваться краулерами?
11. Сформулируйте уровни доступа к веб-архивной информации.
12. Перечислите примеры гибридного доступа к веб-архивам. Объясните их суть.
13. Какие существуют возможности поиска в веб-архивах?
14. Какие возможности предоставляет сервис WebArt?
15. Перечислите основные сложности, которые приходится решать при архивировании?
16. Раскройте проблему предвзятости данных. Приведите примеры.
17. Можно ли сохранить веб полностью? Раскройте проблему неполноты данных.
18. В чем заключается динамичность как особенность веб-контента?
19. Какие существуют проекты сохранения динамичного веба? Назовите особенности подхода датского веб-архива Netarchive.
20. Сформулируйте, какого рода ограничения вызваны закрытостью веб-ресурсов для их сохранения? В чем они проявляются?
21. Какие ограничения связаны с метаданными, их определением и экстрактированием?
22. Перечислите ограничения веб-архивирования технического характера.
23. Какие сложности связаны с устареванием программного обеспечения?
24. Какие возможности реализованы для воспроизведения устаревшего по формату контента?

25. Какое значение имеют социальные сети для историка?

26. Перечислите основные сложности, связанные с сохранением контента социальных сетей.

27. Назовите проекты в области изучения современной истории на основе использования материалов, извлеченных из социальных сетей.

28. Какая из социальных сетей является наиболее популярной для изучения и почему? Приведите примеры реализованных проектов.

## ГЛАВА 4. ВЕБ-ИСТОРИЯ СОЦИУМА И ЕГО ИНСТИТУТОВ: ОБЗОР НАПРАВЛЕНИЙ ИССЛЕДОВАНИЙ

Веб-история как дисциплина имеет очень широкое исследовательское поле, особенно принимая во внимание его междисциплинарность. В данной работе не ставится задача представить исчерпывающий обзор направлений исследований в области веб-истории или показать их разнообразие в полном объеме, основной целью главы является обозначить основные существующие тренды. Исследователи веба указывают на значительный диапазон исследовательской проблематики и разнообразие тем, которые могут быть проанализированы на основе материалов веб-архивов. Ученые отмечают богатство содержания этих ресурсов, незримость их границ и бесконечность возможностей для анализа<sup>108</sup>. Однако в данной главе обозначим лишь отдельные направления веб-истории, сделав акцент на социальные аспекты.

Исторические исследования на основе материалов веб-архивов проводятся в настоящий момент в разных форматах, различного уровня сложности используемого инструментария и масштабов привлеченных ресурсов. Постановка исследовательских задач в области веб-истории зависит от понимания значимости изучаемой тематики, что является ключевым для историка, и примеры реализованных исследований в области веб-истории способны помочь молодым ученым обозначить сферу собственного интереса и сформулировать задачи для анализа.

Веб-история может рассматриваться как история самого веб-пространства и отдельных его сегментов разного масштаба. В эту группу входят исследования, связанные непосредственно с историей веба и его развитием, в первую очередь, это исследования, посвященные национальным сегментам веба – национальным доменам и отдельным его частям. В качестве примера можно привести исследование британских ученых, посвященное национальному вебу Великобритании и изучению его эволюции (домен *.uk*), а также анализу развития университетского сегмента веба (субдомен *ac.uk*)<sup>109</sup>. Не менее масштабным является и изучение наци-

---

<sup>108</sup> Flórez J. A. M. Historia digital: la memoria en el archivo infinito // Historia crítica, ISSN 0121-1617, N°. 43, 2011. Pp. 82–103. URL: <https://revistas.uniandes.edu.co/doi/pdf/10.7440/histcrit43.2011.06> (дата обращения: 01.05.2022).

<sup>109</sup> Meyer E. T., Yasseri T., Hale, Scott A., Cowls J., Schroeder R., Margetts H. (2017) Analysing the UK web domain and exploring 15 years of UK universities on the web. In: Niels Brügger, Ralph

онального сегмента веба Дании – в серии работ ученые рассмотрели доменные имена и их изменение в ретроспективе<sup>110</sup>, проанализировали разрастание веб-ресурсов в домене *.dk*, определили возможности и ограничения использования ресурсов датского веб-архива *Netarkivet*<sup>111</sup>.

Изучение национального веба и динамики его развития действительно значимо, поскольку в результатах становится возможным отразить восприятие Всемирной паутины и интернет-технологий обществом, продемонстрировать их проникновение в жизнь людей и сообществ, общественных институтов и государственных структур. Подобный анализ способен также показать усиление интереса разных групп к созданию веб-ресурсов, их совершенствованию и приспособлению для нужд пользователей ввиду изменившихся представлений об оптимальной их структуре и содержании.

Во введении к сборнику “The Web as History” Р. Шредер и Н. Брюггер внесли в повестку обсуждений в академическом сообществе вопросы о том, как эволюция доменов национального веба отражает изменения в обществе<sup>112</sup>, каким образом веб меняется, наполняется, трансформируется, подстраивается под нужды общества, а также какая существует взаимосвязь между развитием веб-пространства с событиями внешнего и внутреннего характера – политическими, экономическими, культурными, общественными.

Вместе с тем веб-история трактуется существенно шире, не столько как история веб-пространства, сколько как история, которую мы можем реконструировать посредством веб-ресурсов (в первую очередь, заархивированных) как исторических источников. И основные направления исследований в рамках веб-истории включают политическую, экономическую, социальную и культурную сферы.

---

Schroeder (Hg.): *The Web as History. Using Web Archives to Understand the Past and the Present*. London: UCL Press 2017. Pp. 23–44. DOI: <https://doi.org/10.25969/mediarep/12515>

<sup>110</sup> Brügger N., Laursen D., Nielsen J. (2017) Exploring the domain names of the Danish web. In: In: Niels Brügger, Ralph Schroeder (Hg.): *The Web as History. Using Web Archives to Understand the Past and the Present*. London: UCL Press 2017. Pp. 62–80. DOI: <https://doi.org/10.25969/mediarep/12515>

<sup>111</sup> Brügger N., Nielsen J., Laursen D. (2020). Big data experiments with the archived Web: Methodological reflections on studying the development of a nation’s Web. *First Monday*, 25(3). <https://doi.org/10.5210/fm.v25i3.10384>

<sup>112</sup> Schroeder R., Brügger N. (2017) Introduction: The web as history. In: Niels Brügger, Ralph Schroeder (Hg.): *The Web as History. Using Web Archives to Understand the Past and the Present*. London: UCL Press 2017. Pp. 1–19. DOI: <https://doi.org/10.25969/mediarep/12515>



#### **4.1. Социальная веб-история: Общество и общественные объединения**

Реконструкция социальной истории на основе материалов веб-архивов – важная задача. Современная жизнь пронизана интернет- и веб-технологиями, которые способствуют кардинальному изменению общества. Всемирная паутина создала обширные возможности для производства, передачи и получения информации, стала новой средой, частью повседневности современного человека. Развитие веба привело к значительному социальному сдвигу<sup>113</sup>, природу и основные проявления которого еще предстоит изучить.

Ранее уже говорилось о значении социальных медиа в современном обществе (см. Гл. 3) и все расширяющихся возможностях распространения информации. Для понимания содержания медийных ресурсов необходимо уметь анализировать и выявлять, кто создает информацию для распространения по сетям, и для какой аудитории она предназначена. Речь идет не только о блогерах, колумнистах и ведущих популярных каналы в цифровых медиа, которые имеют широкую аудиторию в сотни тысяч и миллионы подписчиков. Вопрос шире – на какие демографические, социокультурные и иные категории можно разделить пользователей и производителей контента в сети. В социологических исследованиях показано, что распределение по возрастам пользователей сети является довольно четким – старшие возрастные категории не столь склонны публиковать контент, сколько другие, более молодые категории авторов. Также было отмечено, что гендерный состав аудитории производителей и пользователей информации более сложен для анализа<sup>114</sup>.

Социологические исследования интернет-аудитории весьма полезны для веб-историков, поскольку способны представить основные категории «производителей» и «потребителей» контента. Определенная трудность связана с тем, что в качестве источника данных социологи чаще используют «живой веб», который мог впоследствии исчезнуть или видоизме-

---

<sup>113</sup> Danyel J. (2012) Zeitgeschichte der Informationsgesellschaft, in: Zeithistorische Forschungen/Studies in Contemporary History, Online-Ausgabe, 9 (2012), H. 2, URL: <https://zeithistorische-forschungen.de/2-2012/4441>, DOI: <https://doi.org/10.14765/zzf.dok-1598> (дата обращения: 01.05.2022), Druckausgabe: S. 186–211; Change: 19 Key Essays on How the Internet Is Changing Our Lives. Handbook. 2014. URL: <https://www.bbvaopenmind.com/en/books/19-key-essays-on-how-internet-is-changing-our-lives> (дата обращения: 01.05.2022).

<sup>114</sup> Blank G. (2013) Who creates cocontent? // Information, Communication & Society, Vol. 16:4. Pp. 590–612.

ниться и не быть зафиксированным в веб-архивных материалах. Однако такие исследования позволяют понимать тенденции, преобладавшие в тот или иной период времени, даже если повторить или проверить эксперимент уже нет возможности. Такие срезы значимы при проведении исторических исследований, позволяют сравнивать ситуации и даже на уровне историографии можно получать своего рода снимоты происходящего.

Современная социальная история, несомненно, имеет свою специфику<sup>115</sup>, а веб-история социума как ее часть позволяет проанализировать информационное общество, его особенности и характеристики за последние четверть века.

Одной из существенных особенностей современной истории, ее феноменом, является создание **киберсообществ**, появление которых стало возможным только с развитием Интернета и веб-технологий. Киберсообщество – это объединение людей в виртуальной среде, чаще всего по интересам или роду деятельности, которые поддерживают общение посредством сайтов, мессенджеров, форумов или иных средств цифровой коммуникации.

Изучение киберсообществ позволяет получить представление о социальной структуре информационного общества, ее гибкости и динамичности. Киберсообщества могут быть различными исходя из интересов вовлеченных групп и индивидов. С точки зрения традиционного историка, сами киберсообщества, взаимодействие отдельных представителей внутри сообществ и соприкосновение с внешним контуром (другими виртуальными и «реальными» сообществами, общественными организациями, государственными структурами и др.) остаются невидимыми ввиду специфики цифровых каналов распространения информации. Прямым следствием этих процессов является тот факт, что источники для изучения феномена киберсообществ в исторической ретроспективе являются не вполне очевидными, что требует особых подходов к поиску сведений и формированию источниковой базы.

Начало широкого распространения виртуальных сообществ относят к активному развитию Всемирной паутины середины 1990-х гг. Естественно, что многие из киберсообществ уже прекратили свое существова-

---

<sup>115</sup> Зидер Р. Что такое социальная история? Разрывы и преемственность в освоении «Социального» = Was heisst Sozialgeschichte? Brüche und Kontinuitäten in der Aneignung des «Sozialen» // THESIS. 1993. № 1. С. 163-181. URL: [https://www.hse.ru/data/108/314/1234/1\\_4\\_2\\_Sied.pdf](https://www.hse.ru/data/108/314/1234/1_4_2_Sied.pdf) (дата обращения: 01.05.2022).

ние на настоящий момент, и реконструкция их истории может быть осуществлена преимущественно на основе материалов, сохранившихся в веб-архивах.

Исследования в области истории виртуальных сообществ интересны как постановкой вопроса, так и исследовательской методологией. Историку И. Миллигану удалось восстановить историю виртуальных сообществ на материалах платформы GeoCities, которая была популярна в 1990-е гг. и в первое десятилетие XXI в. прекратила свое существование<sup>116</sup>. Канадский исследователь собрал материалы снимков сайта GeoCities, заархивированных в Интернет-архиве (полный набор данных для анализа составил более 600 гигабайт), и проанализировал изменение численности пользователей на портале. И. Миллиган показал рост количества зарегистрированных аккаунтов с 10 тыс. в середине 1995 г. до 1 млн. два года спустя, выявил отдельные сообщества в рамках GeoCities и их лидеров, выделил основные интересы и темы для коммуникации на данном портале, проанализировав веб-архивные материалы на основе целого комплекса методов, включая дальнейшее чтение<sup>117</sup>.

Часто деятельность сообществ, осуществляющих коммуникацию преимущественно посредством виртуальных ресурсов, может быть направлена и на офлайновый мир. Такие **общественные движения** и коалиции могут быть как чисто онлайн-овыми, так и гибридными по своей сути, сочетая активность в физическом мире и цифровом. Распространение Всемирной паутины способствовало перенесению деятельности целого ряда сообществ из реальной среды в виртуальную. Например, к таким сообществам относятся фанатские движения в сфере культуры и искусства – фандомы, которые получили новый импульс своего развития с расширением веба. А. Хорбински в работе “Talking by letter: the hidden history of female media fans on the 1990s Internet”<sup>118</sup> проанализировала этот процесс на примере женских фанатских сообществ в ранний период существования Всемирной паутины.

---

<sup>116</sup> Milligan I. (2015) Finding Community in the Ruins of GeoCities: Distantly Reading a Web Archive // UWSpace. 2015. URL: <http://hdl.handle.net/10012/11650> (дата обращения: 01.05.2022).

<sup>117</sup> Jänicke S., Franzini G., Cheema M. F., Scheuermann G. (2015) On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges // Eurographics Conference on Visualization (EuroVis) – STARS (eds. Borgo R., Ganovelli F., Viola I.). URL: <http://www.informatik.uni-leipzig.de/~stjaenicke/Survey.pdf> (дата обращения: 01.05.2022).

<sup>118</sup> Horbinski A. (2018). Talking by letter: the hidden history of female media fans on the 1990s internet. Internet Histories. Pp. 1–17. DOI:10.1080/24701475.2018.1500794

Изучение истории онлайн-активизма и общественных движений становится возможным на основе тематических коллекций, сохраненных в веб-архивах. Такие подборки могут включать как сайты и материалы, создаваемые самими движениями для мобилизации, координации и других функций, так и ресурсы, освещающие их деятельность в реальном пространстве (например, новостные издания)<sup>119</sup>. Интернет-активизм предполагает различные варианты вовлеченности – от чисто информационного, когда доминирует получение информации с минимальным откликом на нее (преимущественно в виде репостов и лайков), до более активных вариантов участия, когда член движения или сочувствующий предпринимает конкретные шаги в виде финансовой поддержки или более глубокого вовлечения в деятельность организации.

Изучению онлайн-активизма и его эволюции посвящена работа М. Кланга и Н. Мэдисон “The domestication of online activism”<sup>120</sup>, в которой авторам удалось выявить факторы современности, определившие трансформацию разных форм онлайн-движений. Однако анализ интернет-активизма во всех его формах довольно затруднителен в исторической ретроспективе, поскольку требует наличия целого комплекса задокументированных источников, на основе которых можно реконструировать историю того или иного движения.

С развитием Всемирной паутины постепенно происходило приспособление общества и отдельных пользователей к открывшимся возможностям, формировались привычки, складывались правила (формальные и неформальные) и способы поведения в сети, со временем происходила их трансформация. Изучение этих процессов также является частью веб-истории, поскольку многое нашло свое отражение в архивированных веб-ресурсах. Примерами исследований в этом направлении являются статьи М. Найчика “Trolls and Trolling History: From Subculture to Mainstream

---

<sup>119</sup> См, например, Indianapolis Black Lives Matter (BLM) Movement. Collected by Indiana University School of Informatics and Computing. Archive-It. URL: <https://archive-it.org/collections/15192> (дата обращения: 01.05.2022); Local Activism collection. Collected by Michigan State University. Archive-It. URL: <https://archive-it.org/collections/11677> (дата обращения: 01.05.2022); Cuban 27N Movement Web Archive. Collected by Ivy Plus Libraries Confederation. Archive-It. URL: <https://archive-it.org/collections/16175> (дата обращения: 01.05.2022).

<sup>120</sup> Klang M., Madison N. The domestication of online activism. // First Monday, Vol. 21 (6), 6 June 2016. URL: <https://firstmonday.org/ojs/index.php/fm/article/download/6790/5520> (дата обращения: 01.05.2022).; doi: <http://dx.doi.org/10.5210/fm.v21i6.6790>

Practices”<sup>121</sup>, в которой автор проследил появление и развитие троллинга как социального явления, Э. Бельтрамини “Against technocratic authoritarianism. A short intellectual history of the cypherpunk movement”<sup>122</sup>, где рассмотрены вопросы анонимности как фактора активности социальных движений, Ф. Брантона “Spam”<sup>123</sup>, в которой проанализированы различные формы и проявления нежелательных сообщений и навязчивой рекламы и их развитие.

## 4.2. История государственных и общественных институтов

В рамках исследований по веб-истории могут найти отражение вопросы развития современных общественных институтов разного уровня – государственных (правительственных), образовательных, религиозных, экономических и семейных.

Электронные сервисы политических, экономических, социальных и культурных секторов приобретают особое значение в информационном обществе и являются частью глобальных процессов цифровой трансформации современного государства. Разработка таких сервисов позволяет изменить функционирование самих институтов, сделать их работу более слаженной и эффективной, способствовать развитию их транспарентности (открытости) для граждан. С точки зрения обывателя эти сервисы позволяют упростить многие повседневные процессы, сократить время получения услуг и т.д.

В исторической ретроспективе развитие виртуальных ресурсов общественных институтов можно проследить, через изменение функциональности сайтов и приложений посредством анализа снапшотов этих ресурсов, хранящихся в веб-архивах. Очевидно, что на примере «живого» веба можно оценить преимущественно текущую ситуацию, а веб-архивы позволяют (в разной степени глубины) отследить историю развития этих институтов. Официальные веб-сайты являются ключевыми источниками доступными в публичном пространстве, которые позволяют изучать ука-

---

<sup>121</sup> Nycyc M. (2019) Trolls and Trolling History: From Subculture to Mainstream Practices. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd, 2019. pp. 577–590.

<sup>122</sup> Beltramini E. (2020) Against technocratic authoritarianism. A short intellectual history of the cypherpunk movement // Internet Histories. 2020, DOI: 10.1080/24701475.2020.1731249

<sup>123</sup> Brunton F. (2019) Spam. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd. pp. 564–576.

занные процессы. Однако следует отметить, что для более полного анализа проблематики требуется привлечение и других типов источников, включая делопроизводственную документацию.

Изучение официальных сайтов и их содержания с позиций исторического исследования, с одной стороны, дает возможность оценить изменение представленной на ресурсах информации, а также выявить конкретные аудитории (граждане, сообщества, бизнес), для которых эти ресурсы предназначены. Кроме того, посредством такого анализа появляется возможность увидеть способы адаптации функционирования общественных институтов к современному укладу жизни посредством цифровых сервисов, представить становление информационного общества на практике.

В отношении политических институтов – государственного и муниципального управления – распространение специализированных сервисов и платформ, расширение их функционала и анализ этих процессов в исторической ретроспективе позволяют увидеть историю электронного государства и цифровой экономики и реализацию этих концепций на практике. Институт президентства, правительства в совокупности ветвей власти, муниципальные органы управления прошли через длительный этап эволюции, продолжают развиваться и формируют широкое поле для исследований, в том числе, в направлении веб-истории.

Ранее (см. Гл. 2–3) рассматривались различные программы сохранения веба и обращалось внимание на тот факт, что в случае отсутствия национальных программ по веб-архивированию ресурсы могут быть сохранены преимущественно в Интернет-архиве. Полнота такого сохранения в данном случае не является предметом обсуждения, но случайность как фактор, влияющий на полноту выборки, играет свою критическую роль и должен приниматься во внимание при планировании исследования.

В России изучение истории политических институтов современности возможно не только на материалах Интернет-архива. АНО «Информационная культура» осуществляет сохранение правительственных сайтов, а также перечня аккаунтов в социальных сетях (Национальный цифровой архив России – проекты «Консервированное государство» и «Специализированные коллекции»<sup>124</sup>). Регулярно пополняется репозиторий, содержимое которого может быть предоставлено инициаторами проекта исследователям для дальнейшего изучения.

---

<sup>124</sup> Национальный цифровой архив России. URL: <https://ruarxive.org> (дата обращения: 01.05.2022).

Следует привести примеры исследовательских проектов в области веб-истории общественных институтов. Например, в статье В. Шафер “From far away to a click away: The French state and public services in the 1990s”<sup>125</sup> представлен анализ французских публичных электронных сервисов и их развитие. В работе М. Хесля “Semantics of the internet: a political history”<sup>126</sup> рассматривается история политики Германии в отношении Интернета на основе анализа семантики веб-ресурсов.

Другие исследования раскрывают трансформацию институтов семьи и религии в связи с проникновением информационных технологий. Хорошим примером в данном направлении являются статьи П. Вебстера “Religion and Web History”<sup>127</sup> и “Technology, ethics and religious language: early Anglophone Christian reactions to «cyberspace»”<sup>128</sup>, посвященные трансформации религии как общественного института в его разных проявлениях и отражению этих процессов в англоязычных веб-ресурсах. Автору удалось показать изменение соотношения религиозных офлайн- и онлайн-процессов и процедур на основе материалов веб-архивов.

История экономических отношений также может быть проанализирована на основе веб-ресурсов. Примером подобных исследований может послужить работа М. Крэйна “A Critical Political Economy of Web Advertising History”<sup>129</sup>, в которой изучена история рекламы на просторах веба и рассмотрена эволюция ее форм.

Существуют работы не только национального, но и глобального характера. Они направлены на изучение современной истории гражданского общества и расширение общественного влияния на Интернет и управление сетями в общемировом масштабе. Таковым является исследование

---

<sup>125</sup> Schafer V. (2017) From far away to a click away: The French state and public services in the 1990s. In: Niels Brügger, Ralph Schroeder (Hg.): The Web as History. Using Web Archives to Understand the Past and the Present. London: UCL Press 2017. Pp. 117–133. DOI: <https://doi.org/10.25969/mediarep/12515>

<sup>126</sup> Hösl M. (2019) Semantics of the internet: a political history, Internet Histories, 3. Pp. 3–4, 275–292, DOI: 10.1080/24701475.2019.1656921

<sup>127</sup> Webster P. (2019) Religion and Web History. In: The SAGE Handbook of Web History. Eds. N. Brügger, I. Milligan. SAGE Publications Ltd. Pp. 479–490.

<sup>128</sup> Webster P. (2018) Technology, ethics and religious language: early Anglophone Christian reactions to ‘cyberspace’ // Internet Histories, Vol. 2, issue 3–4. Pp. 299–314, DOI: 10.1080/24701475.2018.1468976

<sup>129</sup> Crain M. (2019) A Critical Political Economy of Web Advertising History. In: The SAGE Handbook of Web History. Eds. N. Brügger, I. Milligan. SAGE Publications Ltd. Pp. 330–343.

Ш. Харистья “The efficacy of civil society in global internet governance”<sup>130</sup>, в котором проанализированы проблемы, с которыми сталкивается гражданское общество при расширении участия в управлении Интернетом.

#### **4.3. История культуры и эволюция способов ее репрезентации в сети**

Информационные ресурсы, созданные организациями сферы культуры, также могут являться предметом для изучения и реконструкции на основании материалов веб-архивов. Проникновение веб-технологий в культуру проявилось по-разному, и несомненно, оказало воздействие на все ее сферы. Учреждения начали активно создавать сайты и различные платформы для продвижения своей деятельности в сети и расширения возможностей собственного развития. Ранее обращалось внимание на изучение истории фанатских сообществ, что явно пересекается с историей современной культуры, и социальные факторы должны учитываться при анализе информационных ресурсов культурного сектора, где воздействие веб-технологий также является значительным.

Эволюция возможностей и способов репрезентации культурного наследия в разных формах также ставится в центр внимания, поскольку анализ трендов в исторической ретроспективе позволяет сформировать представление о будущем развитии. В этом направлении изучается представление как материальной культуры, так и нематериальной. В исследовании Дж. В. Морриса “Hearing the Past: The Sonic Web from MIDI to Music Streaming”<sup>131</sup> проанализирована эволюция способов публикации аудиоматериалов в веб-пространстве, форматов репрезентации и функционала стриминговых сервисов.

Ускорение процессов цифровизации в области культурного наследия, информатизация, интернетизация и компьютеризация учреждений хранения культурного наследия привели к трансформации функций организаций GLAM-индустрии (сокращ. от Galleries, Libraries, Archives, Museums). К GLAM-сектору относятся культурные институции, предназначение которых заключается в обеспечении доступа к культурному

---

<sup>130</sup> Haristya Sh. (2020) The efficacy of civil society in global internet governance // Internet Histories, 4:3. Pp. 252–270. URL: <https://www.tandfonline.com/doi/full/10.1080/24701475.2020.1769892> DOI: 10.1080/24701475.2020.1769892 (дата обращения: 01.05.2022).

<sup>131</sup> Morris J. W. (2019) Hearing the Past: The Sonic Web from MIDI to Music Streaming. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd, 2019. p. 491.



наследию, формированию и распространению знаний о нем. Воздействие веб-технологий на эту сферу и развитие информационного пространства находятся в центре внимания исследователей не только для анализа текущей ситуации и определения трендов, но и для формирования предложений и механизмов по усовершенствованию ресурсов.

Архивированные веб-сайты и другие ресурсы, созданные музеями и другими учреждениями культуры, позволяют реконструировать их веб-историю. В статье автора данного учебного пособия «Веб-архивы в реконструкции истории виртуальных музеев: потенциал и ограничения»<sup>132</sup> выявлен информационный потенциал веб-архивов для изучения музейной веб-истории, показаны ограничения веб-архивирования, которые препятствуют восстановлению веб-истории во всей полноте. В другой работе – “Digital History of Virtual Museums: The Transition from Analog to Internet Environment”<sup>133</sup> рассматривается процесс перехода виртуальных музеев из аналоговой среды в пространство Интернета, а также развитие первых виртуальных музеев во Всемирной паутине. Портал “Digital History of Virtual Museums”<sup>134</sup> содержит краткий анализ истории виртуальных музеев, выделение основных этапов эволюции ресурсов данного типа, а также несколько очерков по истории отдельных виртуальных музеев из России, США, Канады, Германии и Австралии.

История развития информационных ресурсов музеев, архивов и библиотек отражает эволюцию и изменение функций самих учреждений, а также их адаптивность и способность приспособить новые технологии к своей деятельности. В этом направлении важны исследования, посвященные как эволюции самих ресурсов, например, цифровых библиотек<sup>135</sup>, так и изменениям, связанным со взаимодействием человека и компьютера и

---

<sup>132</sup> Поврозник Н. Г. Веб-архивы в реконструкции истории виртуальных музеев: потенциал и ограничения // Вестник Пермского университета. Серия «История». 2020. № 4(51). С. 95–105.

<sup>133</sup> Povroznik N. (2020) Digital History of Virtual Museums: The Transition from Analog to Internet Environment. In: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference Riga, Latvia, October 21–23, 2020. pp. 125–136. URL: <http://ceur-ws.org/Vol-2612/paper9.pdf> (дата обращения: 01.05.2022).

<sup>134</sup> Digital History of Virtual Museums. URL: <http://virtualmuseumhistory.com> (дата обращения: 01.05.2022).

<sup>135</sup> См. например, Горный Е., Вигурский К. Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы // Интернет и российское общество / Под ред. И. Семенова; Моск. Центр Карнеги. – М.: Гендальф, 2002. – 279 с.; Candela L., Castelli D., Pagano P. History, Evolution, and Impact of Digital Libraries. In: E-Publishing and Digital Libraries: Legal and Organizational Issues. 2011. Pp. 1–30.

его интерфейса<sup>136</sup> (Human-computer interaction, HCI). Хотя эти направления по сравнению с веб-историей представляют намного более широкое поле прикладных информационных технологий, но, несомненно, пересекаются с ним в существенной мере.

В отношении веб-истории чрезвычайно важными являются вопросы развития информационных ресурсов и их репрезентации как цифрового культурного наследия. А. Щетвина в статье «Память о старых сайтах. Веб как часть истории в кураторских проектах» обращает внимание на процессы выбора веб-материалов для их репрезентации в электронной среде. В частности, исследовательница веба и координатор Клуба любителей Интернета<sup>137</sup> указывает: «Кураторские проекты приобретают особую роль: они создают нарративы, которые должны выглядеть правдоподобно с точки зрения современной исторической науки, но быть представлены понятным и интересным для зрителя способом. Это значит, что кураторские проекты используют разные стратегии репрезентации веба и по-своему определяют границы того, что стоит помнить. Кураторские проекты направляют пути публичного разговора об истории веба, определяют ее смысл и значение, устанавливают прецеденты того, каким образом и что именно стоит сохранять как важное»<sup>138</sup>. На примере двух выставок – “64bits”<sup>139</sup> и “One Terabyte of a Kilobyte Age”<sup>140</sup> – А. Щетвина рассматривает особенности представления веб-контента и его истории, в том числе специфику и разнообразие нарративов о вебе.

История современной художественной культуры и разных направлений внутри нее интересует многих исследователей. В качестве примера можно привести изучение истории развития современной датской музыки посредством анализа веба как главного исторического источника<sup>141</sup>.

---

<sup>136</sup> Hornecker E., Ciolfi C. Human-Computer Interactions in Museums. In: Synthesis Lectures on Human-Centered Informatics. 2019. 171 p.

<sup>137</sup> Клуб любителей интернета. URL: <http://clubforinternet.net/> (дата обращения: 01.05.2022).

<sup>138</sup> Щетвина А. Память о старых сайтах. Веб как часть истории в кураторских проектах // Новое литературное обозрение. 2020, № 2. С. 95–117. URL: [https://www.nlobooks.ru/magazines/neprikosnovennyy\\_zapas/130\\_nz\\_2\\_2020/article/22548/](https://www.nlobooks.ru/magazines/neprikosnovennyy_zapas/130_nz_2_2020/article/22548/) (дата обращения: 01.05.2022).

<sup>139</sup> 64 BITS – An Exhibition of the Webs Lost Past. URL: <https://64bits.co.uk/> (дата обращения: 01.05.2022).

<sup>140</sup> One Terabyte of Kilobyte Age Photo Op. URL: <https://oneterabyteofkilobyteage.tumblr.com/> (дата обращения: 01.05.2022).

<sup>141</sup> Smith-Sivertsen, H., Brügger N. (2015). The web archive as historical source: The case of recent music history. Paper præsenteret ved Web Archives as scholarly Sources: Issues, Practices and Perspectives, Aarhus, Danmark.

Широкое проникновение интернет-технологий в жизнь человека привело к складыванию нового явления – интернет-культуры или киберкультуры<sup>142</sup>, которое включает целый комплекс взаимосвязанных элементов современного информационного общества (в том числе, социально-культурных связей как человека с виртуальным миром, так и виртуальных сообществ и системы их отношений между собой и «внешним» миром). По мнению целого ряда исследователей, следует обсуждать множественность культур Интернета<sup>143</sup>. История интернет-культур(ы) раскрывает эволюцию этих связей и отношений, а также процессов «производства», распространения и «потребления» культуры, то есть развития массовой культуры во Всемирной паутине. Поэтому веб-ресурсы и веб-архивы играют ключевую роль в реконструкции указанных процессов<sup>144</sup>.

Цифровое искусство также является частью художественной культуры. История цифрового искусства как отдельное направление было институционализировано появлением и развитием факультетов, лабораторий и школ, специализированных образовательных курсов в университетах (одним из примеров является Кафедра искусства, истории искусств и визуальных исследований университета Дьюк, предлагающая магистерскую программу “Digital Art History”<sup>145</sup>), а также изданием научного журнала “International Journal on Digital Art History”<sup>146</sup>.

История цифрового искусства выходит далеко за пределы веб-истории, но в пересекающемся поле представляет существенный интерес для исследователей веба и истории веб-ресурсов. Перечислим только отдельные работы, опубликованные в этом направлении, в которых существенную роль играет анализ веба в качестве исторического источника – это, например, статьи: Т. Зауэрландер “A Short History of Self-

---

<sup>142</sup> Шерстобоева Е. А., Копалкина Т. Г., Лученко К. В., Новикова А. А., Санданов А. Б. Культура в Интернете. Интернет-культура. / Культура России. 2000-е годы. СПб.: Алетейя, 2012. С. 719–739.

<sup>143</sup> Kirmayer L. J., Raikhel E, Rahimi S. Cultures of the internet: identity, community and mental health // *Transcult Psychiatry*. 2013. Vol. 50(2). Pp. 165-91. DOI: 10.1177/1363461513490626. URL: <https://ssrn.com/abstract=2545596> (дата обращения: 01.05.2022).

<sup>144</sup> См. Turner F. (2017) Can we write a cultural history of the Internet? If so, how? // *Internet Histories*, Vol. 1: 1–2. Pp. 39–46, DOI: 10.1080/24701475.2017.1307540

<sup>145</sup> MA in Digital Art History. The Department of Art, Art History & Visual Studies. Trinity College of Arts and Sciences. Duke University. URL: <https://aahvs.duke.edu/graduate/ma-dah> (дата обращения: 01.05.2022).

<sup>146</sup> International Journal on Digital Art History. URL: <https://dahj.org/> (дата обращения: 01.05.2022).

Representation in Digital Art”<sup>147</sup>, посвященная истории саморепрезентации в цифровом искусстве, в том числе, в веб-пространстве; Д. Олдман, Д. Тэнасе и С. Сантски “The Problem of Distance in Digital Art History”<sup>148</sup>, в которой анализируется эволюция исторического нарратива вокруг творчества К. Хокусай.

Институционализация направления истории цифрового искусства способствует формированию исследовательских инфраструктур. Так, разные веб-архивные платформы часто ориентированы на создание тематических коллекций ресурсов в области искусства, в том числе, цифрового (такие коллекции есть в Интернет-архиве<sup>149</sup>, например). Кроме того, существуют специализированные программы по веб-архивированию веба в области цифрового искусства. Примерами могут служить веб-архив Digital Art Web Archive<sup>150</sup>, инициированный Библиотекой Корнельского университета (США), а также специализированный проект “New York Art Resources Consortium”<sup>151</sup>, реализуемый при сотрудничестве Справочной библиотеки по искусству Фрика, Библиотеки и архивов Бруклинского музея и Библиотеки Музея современного искусства (МоМА), посвященный истории искусства.

Следует также обратить внимание и на феномен Всемирной паутины, которая может сама по себе рассматриваться как часть цифровой культуры. Естественным образом, в WWW представлены различные направления цифровой культуры. Библиотека Конгресса США в рамках программы по веб-архивированию пополняет и развивает ресурс Web Cultures Web Archive<sup>152</sup>, в котором кроме подборок веб-сайтов по теме представлены и коллекции объектов цифрового искусства разных форма-

---

<sup>147</sup> Sauerlaender T. A Short History of Self-Representation in Digital Art // International Journal on Digital Art History. 2020. Vol. 5. URL: <https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/77407> (дата обращения: 01.05.2022).

<sup>148</sup> Oldman D., Tanase D., Santschi S. The Problem of Distance in Digital Art History: A ResearchSpace Case Study on Sequencing Hokusai Print Impressions to Form a Human Curated Network of Knowledge // International Journal for Digital Art History. 2019. Vol. 4:5.29–5.45. <https://doi.org/10.11588/dah.2019.4.72071>. URL: <https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/72071> (дата обращения: 01.05.2022).

<sup>149</sup> Internet Archive Art Projects. Internet Archive. URL: <https://archive.org/details/internet-archive-art-projects> (дата обращения: 01.05.2022).

<sup>150</sup> Digital Art Web Archive. Archive-It. URL: <https://archive-it.org/collections/4388> (дата обращения: 01.05.2022).

<sup>151</sup> New York Art Resources Consortium. URL: <https://nyarc.org/initiatives/web-archiving> (дата обращения: 01.05.2022).

<sup>152</sup> Web Cultures Web Archive. Library of Congress. URL: <https://www.loc.gov/collections/web-cultures-web-archive> (дата обращения: 01.05.2022).

тов (в т. ч. анимированные гифки<sup>153</sup>) и направлений (например, интернет-мемы<sup>154</sup>).

Другой составляющей социально-культурной веб-истории является репрезентация социальной памяти. Веб становится важнейшим ресурсом по сохранению человеческой памяти и воспоминаний, как приватной, личной, так и той, которой люди делятся публично на цифровых платформах разного уровня – локального, регионального, национального или международного. Как следствие, веб-архивные материалы могут являться источниками для **изучения памяти**, воспоминаний и коммеморативных практик современности.

Следует разделять социальную память и память историческую, поскольку они имеют свою специфику<sup>155</sup>. Историческая память направлена на фиксирование и сохранение памяти о прошлых событиях, тогда как социальная память предполагает ценностное и оценочное отношение, мировоззренческие аспекты. В контексте веб-истории социальная и историческая память могут иметь разные варианты рефлексии и определения предметного поля. Так, С. Кунце и Б. Пауэр посвятили свое исследование изучению социальной памяти на примере столетия Пасхального восстания 1916 г. в отражении ресурсов веб-архивов<sup>156</sup>. Другим примером является работа В. Бодуэн, З. Пехливан и П. Стирлинга “Exploring the Memory of the First World War Using Web Archives”<sup>157</sup>, в которой проанализированы разные аспекты памяти о Первой мировой войне на основе материалов французского национального веб-архива (см. подробнее о “legal deposit” в Гл. 2 данного пособия).

---

<sup>153</sup> Search Animated GIFs on the Web – Giphy. Library of Congress. URL: <https://www.loc.gov/item/lcwaN0010391> (дата обращения: 01.05.2022).

<sup>154</sup> Meme Generator. Library of Congress. URL: <https://www.loc.gov/item/lcwaN0010226> (дата обращения: 01.05.2022); Data Mining Memes in the Digital Culture Web Archive. The Signal. Blog. Library of Congress. URL: <https://blogs.loc.gov/thesignal/2018/10/data-mining-memes-in-the-digital-culture-web-archive> (дата обращения: 01.05.2022).

<sup>155</sup> Лойко О. Т., Вакурина Н. А. Социальная память в контексте историзма // Вестн. Том. гос. ун-та. Сер. История. 2011. № 1 (13). URL: [http://journals.tsu.ru/history/&journal\\_page=archive&id=419&article\\_id=7181](http://journals.tsu.ru/history/&journal_page=archive&id=419&article_id=7181) (дата обращения: 01.05.2022).

<sup>156</sup> Kunze S., Power B. (2018) Capturing commemoration: the 1916 Easter Rising web archive project, *Internet Histories*, 2:1-2, 202–217, DOI: 10.1080/24701475.2018.1446238

<sup>157</sup> Beaudouin V., Pehlivan Z., Stirling P. (2019) Exploring the Memory of the First World War Using Web Archives: Web Graphs Seen from Different Angles. *The SAGE Handbook of Web History*. Eds. N. Brugger, I. Milligan. In: *The SAGE Handbook of Web History*. SAGE Publications Ltd. Pp. 441–463.

#### 4.4. Веб-история как история развития медиaproстранства

Важным признаком информационного общества является изменение процессов коммуникации. Поэтому столь важно изучение ключевых акторов этих процессов, новых каналов распространения информации и анализа ее содержания. Частным случаем в этом направлении является исследование способов формирования общественного мнения посредством ресурсов Всемирной паутины. Формирование новостной повестки также является частью указанных процессов и связано с изменением среды для развития массовых коммуникаций, появлением новостных лент, формируемых по новым принципам, которые относятся к динамичности ресурсов и системе рекомендаций.

Фиксация и документирование событий и фактов – важная функция средств массовой информации, придающая им ценность исторических источников. А при должных подходах к архивированию медийных ресурсов веб-архив может стать незаменимым средством изучения коммуникаций, продвижения новых медиа как важной части Всемирной паутины.

Ключевыми компонентами в развитии цифрового медиaproстранства являются:

- появление новых каналов коммуникации, что многократно увеличило воздействие на общество (и, в том числе, расширило возможности для формирования общественного мнения);
- трансформация способов создания, обработки и передачи информации;
- виртуализация и появление новых типов средств массовой информации (и неуклонное снижение количества традиционных печатных СМИ);
- появление системы рекомендаций, индивидуализированных новостных лент и подборок, формируемых на основе технологий искусственного интеллекта (генерирование новостной повестки, специфичной для каждого пользователя согласно его интересам);
- разнообразие форматов продуктов цифровой журналистики и СМИ (гипертекстуальность, мультимедийность, интерактивность стали важными их характеристиками<sup>158</sup>) и др.

---

<sup>158</sup> Deuze M. The Web and its Journalisms: Considering the Consequences of Different Types of Newsmedia Online // New media & Society. 2003. Vol. 5(2). Pp. 203–230. URL: <https://journals.sagepub.com/doi/pdf/10.1177/1461444803005002004> (дата обращения: 01.05.2022).

Таким образом, возникают новые типы медиа-ресурсов, которые не только фиксируют происходящие события и содержат интерпретации фактов, но и создают информационное пространство нового типа – более динамичное и индивидуализированное.

Данные обстоятельства объясняют интерес исследователей к новому направлению – истории Интернет-журналистики и средств массовой коммуникации, которое непосредственно связано с веб-историей. Авторы, работающие в этом направлении, рассматривают появление, развитие и продвижение веб-ресурсов как новой, цифровой медиасреды. Так, в статье М. Вебера и Ф. Наполи “Journalism History, Web Archives, and New Methods for Understanding the Evolution of Digital Journalism”<sup>159</sup> рассматривается история цифровой журналистики и развитие медийных ресурсов на основании материалов веб-архивов. Исследователи анализируют новостные сайты США, представляют возможности изучения местной истории с использованием их заархивированных снапшотов, а также проводят параллели с демографическими и географическими характеристиками пользователей – потребителей контента и тематикой прочитанных ими страниц. В качестве одного из практических выводов, полученного американскими учеными, звучит тезис о необходимости формирования специализированной инфраструктуры для расширения возможностей анализа и более комплексного изучения медийной истории. Схожие идеи представлены во вводной статье Х. Бедкера “Journalism History and Digital Archives”<sup>160</sup> к специализированному выпуску журнала Digital Journalism, посвященному истории журналистики и значению цифровых архивов. Автор статьи и одновременно редактор этого номера журнала обратил внимание на расширение возможностей изучения медийной веб-истории на основе использования количественных методов, возможности для чего существенно расширяются благодаря накоплению больших объемов данных в веб-архивах.

Развитию отдельных СМИ и их веб-истории посвящена работа Э. Феррари “Bodies that matter, bodies that don’t: selective disembodiment in

---

<sup>159</sup> Matthew S. Weber & Philip M. Napoli () Journalism History, Web Archives, and New Methods for Understanding the Evolution of Digital Journalism // Digital Journalism, 2018. 6:9, 1186–1205, DOI: 10.1080/21670811.2018.1510293

<sup>160</sup> Bødker H. Journalism History and Digital Archives // Digital Journalism, 2018. Vol. 6(9). Pp. 1113–1120, DOI: 10.1080/21670811.2018.1516114. URL: <https://www.tandfonline.com/doi/full/10.1080/21670811.2018.1516114>

the early *Wired* magazine (1993–1997)»<sup>161</sup>. Автор рассматривает трансформацию журнала *Wired* во время транзита его выпусков из печатного формата в цифровой. Пример с журналом *Wired* интересен еще и в том смысле, что именно это издание одним из первых начало продвигать идею «Интернет-культуры» (Internet culture)<sup>162</sup>.

Эволюция медийных информационных ресурсов во многом определялась увеличением их социальной значимости, расширением их разнообразия и вовлеченностью аудиторий. Многие ресурсы стали активно привлекать мнения и опыт пользователей к созданию и наполнению информационных продуктов. На примере истории таких ресурсов можно увидеть как формируется мотивация для принятия решений под воздействием предпочтений других пользователей.

Исследование развития платформы TripAdvisor<sup>163</sup>, которая ориентирована на путешественников и позволяет спланировать туристическую поездку, продемонстрировало такой специфичный способ формирования общественного мнения. Для анализа были выбраны культурные локации Лондона и, соответственно, снимки страниц с ними на сайте TripAdvisor, сохраненные в веб-архиве. Исследование было предпринято для того, чтобы выявить представления пользователей о культурном пространстве британской столицы<sup>164</sup>. Одной из задач проекта было рассмотреть влияние платформы на известность и посещаемость культурных локаций, принадлежавших к популярной и «высокой» культурам, а также определить: каким образом транслируется и формируется общественное мнение по отношению к учреждениям и мероприятиям, проводимых в них, на основе оценки популярности разных типов аттракций. Ученые извлекли из веб-архива и сформировали большой набор данных, содержащий отзывы о точках притяжения и время их добавления. В результате исследования был показан процесс стирания различий между высокой и поп-культурами в представлении пользователей ресурса.

---

<sup>161</sup> Ferrari E. (2020) Bodies that matter, bodies that don't: selective disembodiment in the early *Wired* magazine (1993–1997) // *Internet Histories*, Vol. 4:3. Pp. 333–348, DOI: 10.1080/24701475.2020.1769891

<sup>162</sup> Internet Culture. *Wired*. URL: <https://www.wired.com/tag/internet-culture> (дата обращения: 01.05.2022).

<sup>163</sup> TripAdvisor. URL: <https://www.tripadvisor.ru> (дата обращения: 01.05.2022).

<sup>164</sup> Scott A. Hale, Grant Blank and Victoria D. Alexander (2017) Live versus archive: Comparing a web archive to a population of web pages. In: Niels Brügger, Ralph Schroeder (Eds.) *The Web as History. Using Web Archives to Understand the Past and the Present*.



Реализация данного проекта способствовала лучшему пониманию эволюции интернет-ресурсов, в частности, была установлена зависимость от нарастания количества отзывов и появления новых точек притяжения. Кроме того, было проведено сравнение полноты представленной информации в веб-архиве об определенных аттракциях и отзывах пользователей о них. В результате были определены перекосы в полноте сохраненной информации и сделан вывод о том, что в архивном вебе представлено существенно меньше сведений о менее популярных локациях. На основе результатов исследования, было высказано предположение, что Интернет-архив склонен сохранять наиболее популярные и известные веб-страницы с высоким рейтингом. Также было проведено сравнение информации, сохраненной в веб-архиве с той, что представлена на сайте. Срезы проводились на определенный момент времени, и было показано, что репрезентативность данных, сохраненных в веб-архиве, существенно ниже по сравнению с живым вебом. Авторы пришли к выводу о том, что на основании данных Интернет-архива невозможно проводить исследования, связанные со сравнением абсолютных величин ввиду сильного искажения выборки, которую довольно сложно оценить.

Процессы коммуникации и их эволюция в содержательном плане также являются частью медийной веб-истории. История развития информационных ресурсов, и, в частности, медиапространства показывает, что веб следует рассматривать как культурный феномен, влияющий на формирование общественного мнения и предпочтений, на выбор пользователя в разных сферах и на представления о культуре. Вместе с тем интернет-технологии и Всемирная паутина стали ключевыми факторами трансформации культуры.

Таким образом, веб-история как дисциплина и ее исследовательское поле продолжают складываться; отдельные направления более проработаны, другие представляют скорее разнообразные эксперименты с темами, методами и исследовательским инструментарием. Современные возможности и перспективы проведения исследований в области веб-истории напрямую связаны с работой веб-архивов, подходами к сохранению веба, совершенствованием их информационной среды, а также постепенной трансформацией веб-архивов в инфраструктуру, приспособленную для проведения научных исследований. Оценка информационного потенциала веб-архивов, разработка исследовательских подходов, методов и методик, а также становление историографии направления являются ключевыми факторами для продвижения исследований в области веб-истории.

## Рекомендуемая литература

1. Володин А. Ю. «Цифровая история»: ремесло историка в цифровую эпоху // Электронный научно-образовательный журнал «История». 2015, вып. 6 (8). [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001228-9-1> (дата обращения: 01.05.2022).
2. Горный Е., Вигурский К. Развитие электронных библиотек: мировой и российский опыт, проблемы, перспективы // Интернет и российское общество / Под ред. И. Семенова; Моск. Центр Карнеги. – М.: Гендальф, 2002. 279 с.
3. Зидер Р. Что такое социальная история? Разрывы и преемственность в освоении «Социального» = Was heisst Sozialgeschichte? Brüche und Kontinuitäten in der Aneignung des “Sozialen” // THESIS. 1993. № 1. С. 163–181. URL: [https://www.hse.ru/data/108/314/1234/1\\_4\\_2Sied.pdf](https://www.hse.ru/data/108/314/1234/1_4_2Sied.pdf) (дата обращения: 01.05.2022).
4. Лойко О. Т., Вакурина Н. А. Социальная память в контексте историзма // Вестн. Том. гос. ун-та. Сер. История. 2011. № 1 (13). URL: [http://journals.tsu.ru/history/&journal\\_page=archive&id=419&article\\_id=7181](http://journals.tsu.ru/history/&journal_page=archive&id=419&article_id=7181) (дата обращения: 01.05.2022).
5. Поврозник Н. Г. Веб-архивы в реконструкции истории виртуальных музеев: потенциал и ограничения // Вестник Пермского университета. Серия «История». 2020. № 4(51). С. 95–105.
6. Шерстобоева Е. А., Копалкина Т. Г., Лученко К. В., Новикова А. А., Санданов А. Б. Культура в Интернете. Интернет-культура. / Культура России. 2000-е годы. СПб.: Алетейя, 2012. С. 719–739.
7. Щетвина А. Память о старых сайтах. Веб как часть истории в кураторских проектах // Новое литературное обозрение. 2020, № 2. С. 95–117. URL: [https://www.nlobooks.ru/magazines/neprikosnovennyu\\_zapas/130\\_nz\\_2\\_2020/article/22548/](https://www.nlobooks.ru/magazines/neprikosnovennyu_zapas/130_nz_2_2020/article/22548/) (дата обращения: 01.05.2022).
8. Beaudouin V., Pehlivan Z., Stirling P. Exploring the Memory of the First World War Using Web Archives: Web Graphs Seen from Different Angles. The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd, 2019. Pp. 441–463.
9. Beltramini E. (2020) Against technocratic authoritarianism. A short intellectual history of the cypherpunk movement // Internet Histories. 2020, DOI: 10.1080/24701475.2020.1731249.
10. Blank G. (2013) Who creates coontent? // Information, Communication & Society, Vol. 16:4. Pp. 590–612.

11. Bødker H. Journalism History and Digital Archives // Digital Journalism, 2018. Vol. 6(9). Pp. 1113–1120, DOI: 10.1080/21670811.2018.1516114. URL: <https://www.tandfonline.com/doi/full/10.1080/21670811.2018.1516114> (дата обращения: 01.05.2022).
12. Brügger N., Laursen D., Nielsen J. (2017) Exploring the domain names of the Danish web. In: Niels Brügger, Ralph Schroeder (Hg.): The Web as History. Using Web Archives to Understand the Past and the Present. London: UCL Press 2017. Pp. 62–80. DOI: <https://doi.org/10.25969/mediarep/12515> (дата обращения: 01.05.2022).
13. Brügger N., Nielsen J., Laursen D. (2020). Big data experiments with the archived Web: Methodological reflections on studying the development of a nation's Web. First Monday, 25(3). <https://doi.org/10.5210/fm.v25i3.10384> (дата обращения: 01.05.2022).
14. Brunton F. (2019) Spam. In: The SAGE Handbook of Web History. Eds. N. Brügger, I. Milligan. SAGE Publications Ltd. Pp. 564–576.
15. Candela L., Castelli D., Pagano P. History, Evolution, and Impact of Digital Libraries. In: E-Publishing and Digital Libraries: Legal and Organizational Issues. 2011. Pp. 1–30.
16. Crain M. (2019) A Critical Political Economy of Web Advertising History. In: The SAGE Handbook of Web History. Eds. N. Brügger, I. Milligan. SAGE Publications Ltd. Pp. 330–343.
17. Danyel J. (2012) Zeitgeschichte der Informationsgesellschaft, in: Zeithistorische Forschungen/Studies in Contemporary History, Online-Ausgabe, 9 (2012), H. 2, URL: <https://zeithistorische-forschungen.de/2-2012/4441>, DOI: <https://doi.org/10.14765/zzf.dok-1598> (дата обращения: 01.05.2022).
18. Druckausgabe: Change: 19 Key Essays on How the Internet Is Changing Our Lives. Handbook. 2014. Pp. 186–211; URL: <https://www.bbvaopenmind.com/en/books/19-key-essays-on-how-internet-is-changing-our-lives> (дата обращения: 01.05.2022).
19. Deuze M. The Web and its Journalisms: Considering the Consequences of Different Types of Newsmedia Online // New media & Society. 2003. Vol. 5(2). Pp. 203–230. URL: <https://journals.sagepub.com/doi/pdf/10.1177/1461444803005002004> (дата обращения: 01.05.2022).
20. Ferrari E. (2020) Bodies that matter, bodies that don't: selective disembodiment in the early Wired magazine (1993–1997) // Internet Histories, Vol. 4:3. Pp. 333–348, DOI: 10.1080/24701475.2020.1769891
21. Flórez J.A.M. Historia digital: la memoria en el archivo infinito // Historia crítica, ISSN 0121-1617, № 43, 2011. Pp. 82–103. URL: <https://revistas.uniandes.edu.co/doi/pdf/10.7440/histcrit43.2011.06> (дата обращения: 01.05.2022).

22. Haristya Sh. (2020) The efficacy of civil society in global internet governance // *Internet Histories*, 4:3. Pp. 252–270. URL: <https://www.tandfonline.com/doi/full/10.1080/24701475.2020.1769892> DOI: 10.1080/24701475.2020.1769892 (дата обращения: 01.05.2022).
23. Horbinski, A. (2018). Talking by letter: the hidden history of female media fans on the 1990s internet. *Internet Histories*. Pp. 1–17. DOI: 10.1080/24701475.2018.1500794.
24. Hornecker E., Ciolfi C. Human-Computer Interactions in Museums. In: *Synthesis Lectures on Human-Centered Informatics*. 2019. P. 171.
25. Hösl M. (2019) Semantics of the internet: a political history, *Internet Histories*, 3:3-4. Pp. 275–292, DOI: 10.1080/24701475.2019.1656921.
26. Jänicke S., Franzini G., Cheema M. F., Scheuermann G. (2015) On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges // *Eurographics Conference on Visualization (EuroVis) – STARs* (eds. Borgo R., Ganovelli F., Viola I.). URL: <http://www.informatik.uni-leipzig.de/~stjaenicke/Survey.pdf> (дата обращения: 01.05.2022).
27. Kirmayer L. J., Raikhel E, Rahimi S. (2013) Cultures of the internet: identity, community and mental health // *Transcult Psychiatry*, Vol. 50(2). Pp. 165–91. DOI: 10.1177/1363461513490626. URL: <https://ssrn.com/abstract=2545596> (дата обращения: 01.05.2022).
28. Klang M., Madison N. The domestication of online activism. // *First Monday*, Vol. 21 (6), 6 June 2016. URL: <https://firstmonday.org/ojs/index.php/fm/article/download/6790/5520> (дата обращения: 01.05.2022).; DOI: <http://dx.doi.org/10.5210/fm.v21i6.6790>
29. Kunze S., Power B. (2018) Capturing commemoration: the 1916 Easter Rising web archive project, *Internet Histories*, 2:1-2. Pp. 202–217, DOI: 10.1080/24701475.2018.1446238
30. Weber M. S., Napoli Ph. M. (2018) Journalism History, Web Archives, and New Methods for Understanding the Evolution of Digital Journalism // *Digital Journalism*, 2018. 6:9. Pp. 1186–1205, DOI: 10.1080/21670811.2018.1510293
31. Meyer, Eric T. Yasseri, Taha; Hale, Scott A.; Josh Cows; Schroeder, Ralph; Margetts, Helen (2017) Analysing the UK web domain and exploring 15 years of UK universities on the web. In: Niels Brügger, Schroeder R. (2017): *The Web as History. Using Web Archives to Understand the Past and the Present*. London: UCL Press 2017. Pp. 23–44. DOI: <https://doi.org/10.25969/mediarep/12515>
32. Milligan I. (2015) Finding Community in the Ruins of GeoCities: Distantly Reading a Web Archive // *UWSpace*. 2015. URL: <http://hdl.handle.net/10012/11650> (дата обращения: 01.05.2022).

33. Morris J. W. (2019) Hearing the Past: The Sonic Web from MIDI to Music Streaming. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd, 2019. P. 491.
34. Nycyc M. (2019) Trolls and Trolling History: From Subculture to Mainstream Practices. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd, 2019. Pp. 577–590.
35. Oldman D., Tanase D., Santschi S. (2019) The Problem of Distance in Digital Art History: A ResearchSpace Case Study on Sequencing Hokusai Print Impressions to Form a Human Curated Network of Knowledge // International Journal for Digital Art History, Vol. 4:5.29-5.45. <https://doi.org/10.11588/dah.2019.4.72071>. URL: <https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/72071> (дата обращения: 01.05.2022).
36. Povroznik N. (2020) Digital History of Virtual Museums: The Transition from Analog to Internet Environment. In: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference Riga, Latvia, October 21–23, 2020. Pp. 125–136. URL: <http://ceur-ws.org/Vol-2612/paper9.pdf> (дата обращения: 01.05.2022).
37. Sauerlaender T. A Short History of Self-Representation in Digital Art // International Journal on Digital Art History. 2020. Vol. 5. URL: <https://journals.ub.uni-heidelberg.de/index.php/dah/article/view/77407> (дата обращения: 01.05.2022).
38. Schafer V. (2017) From far away to a click away: The French state and public services in the 1990s. In: Niels Brügger, Ralph Schroeder (Hg.): The Web as History. Using Web Archives to Understand the Past and the Present. London: UCL Press 2017. Pp. 117–133. DOI: <https://doi.org/10.25969/mediarep/12515>
39. Schroeder R., Brügger N. (2017) Introduction: The web as history. In: Niels Brügger, Ralph Schroeder (Hg.): The Web as History. Using Web Archives to Understand the Past and the Present. London: UCL Press 2017. Pp. 1–19. DOI: <https://doi.org/10.25969/mediarep/12515>
40. Scott A. Hale, Grant Blank and Victoria D. Alexander (2017) Live versus archive: Comparing a web archive to a population of web pages. In: Niels Brügger, Ralph Schroeder (Eds.) The Web as History. Using Web Archives to Understand the Past and the Present.
41. Smith-Sivertsen H., Brügger, N. (2015). The web archive as historical source: The case of recent music history. Paper præsenteret ved Web Archives as scholarly Sources: Issues, Practices and Perspectives, Aarhus, Danmark. P. 6.
42. Turner F. (2017) Can we write a cultural history of the Internet? If so, how? // Internet Histories, Vol. 1:1-2. Pp. 39–46, DOI: 10.1080/24701475.2017.1307540

43. Webster P. (2018) Technology, ethics and religious language: early Anglo-phone Christian reactions to 'cyberspace' // Internet Histories, Vol. 2, issue 3-4. Pp. 299–314, DOI: 10.1080/24701475.2018.1468976

44. Webster P. (2019) Religion and Web History. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE Publications Ltd. Pp. 479–490.

## **Источники**

1. Клуб любителей интернета. URL: <http://clubforinternet.net/> (дата обращения: 01.05.2022).

2. Национальный цифровой архив России. URL: <https://ruarxive.org> (дата обращения: 01.05.2022).

3. 64 BITS – An Exhibition of the Webs Lost Past. URL: <https://64bits.co.uk/> (дата обращения: 01.05.2022).

4. Cuban 27N Movement Web Archive. Collected by Ivy Plus Libraries Confederation. Archive-It. URL: <https://archive-it.org/collections/16175> (дата обращения: 01.05.2022).

5. Digital Art Web Archive. Archive-It. URL: <https://archive-it.org/collections/4388> (дата обращения: 01.05.2022).

6. Digital History of Virtual Museums. URL: <http://virtualmuseumhistory.com> (дата обращения: 01.05.2022).

7. Indianapolis Black Lives Matter (BLM) Movement. Collected by Indiana University School of Informatics and Computing. Archive-It. URL: <https://archive-it.org/collections/15192> (дата обращения: 01.05.2022).

8. International Journal on Digital Art History. URL: <https://dahj.org/> (дата обращения: 01.05.2022).

9. Internet Archive Art Projects. Internet Archive. URL: <https://archive.org/details/internet-archive-art-projects> (дата обращения: 01.05.2022).

10. Internet Culture. Wired. URL: <https://www.wired.com/tag/internet-culture> (дата обращения: 01.05.2022).

11. Local Activism collection. Collected by Michigan State University. Archive-It. URL: <https://archive-it.org/collections/11677> (дата обращения: 01.05.2022).

12. MA in Digital Art History. The Department of Art, Art History & Visual Studies. Trinity College of Arts and Sciences. Duke University. URL: <https://aahvs.duke.edu/graduate/ma-dah> (дата обращения: 01.05.2022).

13. Meme Generator. Library of Congress. URL: <https://www.loc.gov/item/lcwaN0010226> (дата обращения: 01.05.2022);

14. Data Mining Memes in the Digital Culture Web Archive. The Signal. Blog. Library of Congress. URL: <https://blogs.loc.gov/thesignal/2018/10/data-mining-memes-in-the-digital-culture-web-archive> (дата обращения: 01.05.2022).
15. New York Art Resources Consortium. URL: <https://nyarc.org/initiatives/web-archiving> (дата обращения: 01.05.2022).
16. One Terabyte of Kilobyte Age Photo Op. URL: <https://oneterabyteofkilobyteage.tumblr.com/> (дата обращения: 01.05.2022).
17. Search Animated GIFs on the Web – Giphy. Library of Congress. URL: <https://www.loc.gov/item/lcwaN0010391> (дата обращения: 01.05.2022).
18. Web Cultures Web Archive. Library of Congress. URL: <https://www.loc.gov/collections/web-cultures-web-archive> (дата обращения: 01.05.2022).

### **Вопросы для самопроверки**

1. Перечислите направления, в которых проводятся исследования в области веб-истории.
2. В чем заключается значение изучение национального веба на основании анализа доменов? Приведите примеры подобных исследований.
3. Как повлияло появление и развитие веба на общество?
4. Сформулируйте определение киберсообществ.
5. Как можно изучать историю киберсообществ на основании веб-архивов? Приведите примеры исследовательских примеров в сфере изучения киберсообществ.
6. Что такое фандомы? В чем заключается особенность этих сообществ? Приведите примеры исследовательских проектов в области истории фандомов?
7. На основе каких ресурсов возможно изучение истории онлайн-активизма и общественных движений?
8. Приведите примеры изучения истории адаптации общества к веб-пространству, формирования привычек и правил поведения.
9. Какие общественные институты возможно изучать на основе материалов веб-архивов.
10. Что является основой для изучения электронных сервисов государства?
11. В чем заключается значение исследования сайтов органов государственной власти и их истории?
12. На какой основе возможно проведение исследований истории политических институтов?
13. Какой проект формирует источниковую базу для изучения истории правительственных сайтов в России? В чем его назначение?

14. Приведите примеры исследовательских проектов в области изучения общественных институтов. В чем заключается их значение?
15. Какие исследовательские проекты в области веб-истории направлены на изучение истории культуры?
16. Назовите ресурс, посвященный истории виртуальных музеев. В чем заключается его содержательная специфика?
17. К какому направлению относятся исследования, связанные с взаимодействием человека и компьютера и эволюции интерфейса?
18. В чем заключается значимость изучения ранней истории сайтов учреждений культуры? Приведите примеры исследовательских проектов в этой области.
19. Что такое киберкультура?
20. В чем заключается значимость изучения истории киберкультуры?
21. Приведите примеры исследовательских проектов в сфере истории цифрового искусства.
22. Какие существуют исследовательские инфраструктуры для изучения истории искусства, в том числе цифрового?
23. В чем заключается значение изучения памяти и ее репрезентации посредством заархивированных веб-ресурсов?
24. Сформулируйте значимость исследования истории медиапространства.
25. Какие ключевые компоненты развития медийных ресурсов можно выделить?
26. Приведите примеры исследовательских проектов в области истории журналистики, реализованные на основе использования ресурсов веб-архивов.
27. Какие результаты были получены при изучении развития платформы TripAdvisor на примере лондонских культурных локаций?
28. В чем заключается значение выявления информационного потенциала веб-архивов для изучения социальной истории?



## ГЛАВА 5. ПРЕДМЕТ И ИСТОЧНИКОВАЯ БАЗА ИССЛЕДОВАНИЯ

### 5.1. Выбор и формулировка исследовательского вопроса

Современная история, несомненно, отличается от всех других направлений в широком предметном поле истории специфическим характером источниковой основы для ее изучения. Особая сложность заключается в том, что современность, чрезвычайная близость шкалы изучаемого времени к настоящему диктуют свои правила работы с источниками, предполагая массу разных особенностей, в том числе связанных с этикой, копирайтом (см. Гл. 2), особенностями критики источников, рожденных в цифровом формате и т.д. Всю эту специфику необходимо принимать во внимание исследователям при формулировке целей и задач, планировании и проведении научных проектов.

Исследовательский процесс в области веб-истории, как и в любой другой, требует постановки центрального вопроса для изучения, определения цели и задач. Предмет исследования в области веб-истории может формулироваться с нескольких позиций: выбора объекта исследования, определенного типа ресурса, имеющейся источниковой базы, либо исходить из конкретной исследовательской проблемы. Рассмотренные в Гл. 4 данного пособия направления исследований убедительно показывают, что ресурсы веб-архивов позволяют изучать широкий спектр вопросов социальной, политической и культурной направленности.

Умение ставить исследовательский вопрос в рамках современной истории имеет свою специфику, связанную с различными цифровыми аспектами, в т.ч. особенностями источников, их организацией и анализом. Вместе с тем эта способность имеет общие основания с любыми научными направлениями.

Итак, формулировка исследовательского вопроса может базироваться на **тех проблемах, которые не попали в фокус внимания других авторов**, либо были рассмотрены недостаточно. Анализ историографии – это неотъемлемая часть любого исследования, которая необходима для выявления лакун, определения дискуссионных вопросов, иногда даже поллярных точек зрения на изучаемый предмет.

Ученые указывают на недостаточный уровень развития историографии в области веб-истории. Причиной этого является тот факт, что исследовательское поле дисциплины продолжает свое становление. Поэтому

анализ историографии также требует предварительного поиска релевантной литературы в смежных областях. Помочь исследователю в формировании списка такой научной литературы, а также в поиске отдельных публикаций могут сборники статей в области веб-истории, такие как “The Web as History. Using Web Archives to Understand the Past and the Present”<sup>165</sup>, “Handbook on Web History”<sup>166</sup>, а также специализированные периодические издания, например, журнал “Internet Histories. Digital Technology, Culture and Society”<sup>167</sup>, электронный журнал «Историческая Информатика»<sup>168</sup>, Journal of Digital History<sup>169</sup> и т.д.

Ранее уже рассматривалась веб-история в контексте междисциплинарности (см. Гл. 1), поэтому релевантные публикации могут быть напечатаны в сборниках и других научных изданиях в областях истории, цифровой истории, цифровой гуманитаристики, медиаархеологии, internet и media studies, а также могут быть отражены в публикациях профессиональных сообществ:

- историков, использующих информационные технологии в исследованиях (например, членов Российской ассоциации «История и Компьютер»<sup>170</sup>);
- цифровых гуманитариев (членов национальных и межнациональных ветвей в структуре Альянса цифровых гуманитарных организаций (Alliance of Digital Humanities Organizations)<sup>171</sup>);
- веб-архивистов и исследователей веба (в т.ч., членов RESAW (Research Infrastructure for the Study of Archived Web Materials)<sup>172</sup> и International Internet Preservation Consortium<sup>173</sup>);

---

<sup>165</sup> The Web as History. Using Web Archives to Understand the Past and the Present. Eds. Niels Brügger, Ralph Schroeder. London, UCL Press, 2017. DOI: <https://doi.org/10.14324/111.9781911307563>.

<sup>166</sup> The SAGE Handbook of Web History. Eds. N. Brügger, I. Milligan. SAGE reference, Los Angeles. 2019. 630 p. DOI: <http://dx.doi.org/10.4135/9781526470546>.

<sup>167</sup> Internet Histories. Digital Technology, Culture and Society. URL: <https://www.tandfonline.com/toc/rint20/current> (дата обращения: 01.05.2022).

<sup>168</sup> Историческая Информатика. URL: <https://e-notabene.ru/istinf> (дата обращения: 01.05.2022).

<sup>169</sup> Journal of Digital History. URL: <https://journalofdigitalhistory.org/> (дата обращения: 01.05.2022).

<sup>170</sup> Российская ассоциация «История и Компьютер». URL: <http://aik-sng.ru/> (дата обращения: 01.05.2022).

<sup>171</sup> Alliance of Digital Humanities Organizations. URL: <https://adho.org> (дата обращения: 01.05.2022).

<sup>172</sup> RESAW, a Research Infrastructure for the Study of Archived Web Materials. URL: <https://resaw.eu/> (дата обращения: 01.05.2022).

- исследователей Интернета (например, членов Клуба любителей Интернета и общества<sup>174</sup>) и др.

**Выявление дискуссионных вопросов, различных точек зрения на них и даже полярных мнений**, анализ аргументации, поиск и формулировка новых аргументов способны помочь в определении исследовательского вопроса и затем могут внести вклад в развитие существующих дискуссий. При этом исследователь может присоединиться к тому или иному «лагерю», либо попытаться сформулировать некий альтернативный или компромиссный вариант на основе проведенного изыскания.

Близким к указанному подходу по формулировке и выбору исследовательского вопроса является **«разоблачение мифов»**, поиск и расследование «заблуждений» или же изучение тех вопросов, которые принимаются на веру, не имея обоснованной доказательной базы. Этот подход весьма сложен из-за необходимости выявить эти аксиоматические заявления, а также найти, подобрать, сформулировать весомые аргументы, которые опровергнут или докажут (частично или полностью) высказанную идею.

Формулировка слишком широкого исследовательского вопроса или отсутствие проблематизации в его постановке – это распространенная ошибка, с которой сталкиваются многие молодые исследователи. Поэтому одной из стратегий в ясном изложении вопроса является **сужение и конкретизация исследовательского поля**, рассмотрение отдельных сторон избранного предмета изучения.

Следующим этапом проведения исследования является формирование источниковой базы, анализ полноты которой неизбежно приводит к корректировке исследовательского вопроса и (или) конкретизации цели и задач исследования.

---

<sup>173</sup> International Internet Preservation Consortium. URL: <https://netpreserve.org/> (дата обращения: 01.05.2022).

<sup>174</sup> Клуб любителей интернета и общества. URL: <http://clubforinternet.net/> (дата обращения: 01.05.2022).

## 5.2. Формирование источниковой базы исследования

Зачастую источниковая база, которой располагает ученый, определяет формулировку исследовательского вопроса.

Веб-архивы позволяют реконструировать историю веба различной степени полноты по отношению к разным его сегментам. Частной версией веб-истории является история сайтов (в англоязычной традиции – это поле называется “site history”). Если историк работает со снимками отдельных веб-страниц или целых сайтов, то он работает с задокументированным экраном (screencast documentary)<sup>175</sup>.

Можно выделить микро- и макроуровни исследований согласно объему анализируемых данных, способам их сбора и организации. К макроуровню относятся исследования, связанные с большими данными, работа с которыми требует значительной подготовки, специфических навыков и инструментария, а также коллаборации со специалистами из разных предметных областей. Макроуровень предполагает принципиально более высокий способ организации источниковой базы, в том числе в технологическом плане.

Образовательный процесс требует адаптации существующих методов и выработки более простых подходов для познания основ и принципов использования веб-архивов в исторических исследованиях. Они должны способствовать развитию умения формулировать цель и задачи исследования и решать их в ходе ограниченной по времени (часто семестровой) индивидуальной или групповой работы. Аналогичным образом, формирование источниковой основы исследования в рамках учебного процесса, создание соответствующего набора данных и подбор инструментов для его обработки должны соответствовать образовательным задачам и уровню подготовки студента.

Существует несколько основных подходов к формированию источниковой базы при изучении веб-истории – можно обратиться к поиску по отдельным ресурсам веб-архивов и создать свою подборку снимков для анализа, воспользоваться готовой коллекцией ресурсов (см. в Гл. 1 о тематических коллекциях более подробно) или создать собственную коллекцию ресурсов, предназначенную для решения определенных исследовательских задач или реализации конкретного проекта.

---

<sup>175</sup> Термин был предложен Нильсом Брюггером. Brugger, N. (2008). The archived website and website philology: A new type of historical document? // Nordicom Review, Vol. 29(2). Pp. 151–171.

Большинство веб-архивов имеют выделенные коллекции, классифицирующие сохраненные ресурсы по теме или хронологии для удобства работы исследователей. Тематические коллекции могут быть сфокусированы на разных вопросах политического, экономического, культурного или иного характера. При формировании коллекций ключевыми характеристиками могут являться событие (что имеет существенное значение для историка), персона, тема, жанр, формат контента или тип медиаконтента. Форматы контента могут быть различными – текстовый, изобразительный, аудио-, видео-, мультимедиа и некоторые другие. Интернет-архив, в частности, сканирует контент с платформы YouTube и формирует коллекцию видео<sup>176</sup>. Также в указанном веб-архиве представлены подборки по разным типам цифрового контента<sup>177</sup>.

Иногда подходы к определению специфики коллекции могут быть гибридными и сочетать, например, тему и формат контента. Так, например, в коллекции Indigenous Australians национального австралийского веб-архива TROVE представлены три субподборки, в которых выделены видеофайлы, новости и «Национальные извинения», тематически связанные с аборигенным населением<sup>178</sup>.

Успешное проведение исследований может потребовать формирования специализированных коллекций для решения определенных целей и задач. Подобные исследования по своим масштабам часто принадлежат к макроуровню и могут быть связаны с большими данными, которые, как уже указывалось, требуют комплексного подхода. Достоинство крупномасштабных проектов заключается в том, что они дают более широкий обзор изучаемой части действительности<sup>179</sup>. Такие исследования могут базироваться как на использовании материалов веб-архивов, так и на сохранении «живого» веба, сканировании определенных сайтов или доменов, формировании новых веб-архивных коллекций и наборов данных. Такой подход требует особых навыков, более существенных временных и иных

---

<sup>176</sup> Youtube Videos. Internet Archive. URL: <https://archive.org/details/youtubecrawl> (дата обращения: 01.05.2022).

<sup>177</sup> Internet Archive Blogs. Internet Archive. URL: <http://blog.archive.org/category/video-archive> (дата обращения: 01.05.2022).

<sup>178</sup> Archived Webpage Collections. Indigenous Australians. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au/collection/15012> (дата обращения: 01.05.2022).

<sup>179</sup> Halevy A., Norvig P., Pereira F. (2009) The Unreasonable Effectiveness of Data In: IEEE Intelligent Systems, vol. 24, № 2. Pp. 8–12, March-April 2009, DOI: 10.1109/MIS.2009.36. URL: <https://ieeexplore.ieee.org/document/4804817> (дата обращения: 01.05.2022).

затрат на поиск и интеллектуальный анализ данных (data mining) и имеет более сложную технологическую основу.

В качестве примера таких исследований можно привести проект, посвященный анализу дискуссий об абортах в Австралии в 2005–2015 гг.<sup>180</sup>. Ученые сначала сформировали список сайтов для изучения на основе релевантных теме поисковых запросов, отобрали примеры из «живого» веба, основываясь на запросах к поисковому сервису Google. Был создан список ключевых слов, при обнаружении которых на том или ином ресурсе запускался краулер для сканирования сайта и архивирования материалов. Автоматизация процедур сбора ресурсов, использование поискового движка и краулера позволило исследователям создать не просто случайную выборку сайтов, а сделать ее более репрезентативной и менее зависимой от субъективных факторов и возможного вмешательства человека в процесс подбора. Основываясь на анализе выявленной ссылочной массы и тех взаимосвязях, которые эти ссылки формируют, и анализируя контент, ученые рассмотрели структуру аргументации и апеллирование к противоположным точкам зрения.

Исследование проводилось в два этапа – в 2005 и 2015 гг. Было сформировано два набора данных для использования компаративных методов при анализе материалов, оба из которых были сохранены в репозиториях. Следует подчеркнуть, что авторы основывались на сведениях, которые удалось собрать в эти годы при использовании одинаковой методики сбора данных.

В результате реализации проекта австралийские исследователи выявили три основных дискурса, связанных с абортами – это индустриальный и проактивный дискурсы, относящиеся к инвестициям и бизнесу, научно-исследовательский и образовательный, а также социальный дискурс. Последний из перечисленных основывался на апелляциях к риску для жизни и здоровья и необходимости продолжения дискуссий на политическом уровне.

К микроуровню исследований относится работа с отдельными записями из веб-архивов или создание небольших наборов данных. Наиболее распространенными методами анализа в данном случае являются преимущественно качественные методы изучения ресурсов и методы «ближ-

---

<sup>180</sup> Ackland R., Evans A. (2017) Using the web to examine the evolution of the abortion debate in Australia, 2005–2015. In: Niels Brügger, Ralph Schroeder (Eds.) *The Web as History. Using Web Archives to Understand the Past and the Present*. Pp. 159–189.

него чтения» (см. подробнее о «ближнем» и «дальнем чтении» Гл. 6). Простота данного подхода позволяет сделать его базовым для использования в образовательном курсе.

Доступность ресурсов веб-архивов в сети является другим ключевым фактором их успешного применения в образовательном процессе. На данный момент существует обширный список веб-архивных инициатив<sup>181</sup>, часть из которых предоставляет открытый доступ к ресурсам для просмотра и изучения. Ко многим другим можно обратиться только через локальные сети ответственных за веб-архивирование организаций, что обосновывается различными ограничениями (в т.ч. по авторскому праву). Среди доступных для удаленного использования проектов есть глобальные, как Интернет-Архив<sup>182</sup>. Некоторые крупные проекты по сохранению национального веба, отдельные региональные и тематические проекты также предоставляют доступ к ресурсам посредством открытых сетей. Нет смысла перечислять и приводить примеры этих инициатив, но следует подчеркнуть, что на данный момент функционирует целый ряд веб-архивных платформ с открытым доступом к контенту, которые могут успешно использоваться в образовательном и исследовательском процессе.

У открытых веб-архивов существует целый спектр ограничений в функциональном плане, которые негативно влияют на возможности работы исследователей с ними. Анализируя современные информационные системы веб-архивов, С. Беловари определила целый список проблем, с которыми может столкнуться потенциальный историк будущего. В указанном списке эти проблемы и ограничения дифференцированы по категориям и связаны с поиском и обнаружением, содержанием, объемом и оценкой, просмотром, контекстуализацией, происхождением, предусмотренными функциями и некоторыми другими процессами<sup>183</sup>.

Частичное преодоление этих проблем возможно на основе разработанных специализированных сервисов и дополнительного инструментария для работы с веб-архивами, поиском и анализом данных.

---

<sup>181</sup> List of Web archiving initiatives. Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives) (дата обращения: 01.05.2022).

<sup>182</sup> Internet Archive. URL: <https://archive.org> (дата обращения: 01.05.2022).

<sup>183</sup> Belovari S. (2017) Historians and Web Archives // Archivaria, Vol. 83. P. 61. URL: <https://archivaria.ca/index.php/archivaria/article/view/13600/14985> (дата обращения: 01.05.2022).

## 5.3. Сервисы для формирования источниковой базы

### 5.3.1. Wayback Machine

В Гл. 2 был предложен обзор веб-архивных инициатив, в том числе тех проектов, которые публикуют заархивированные веб-материалы в открытом доступе. В данном подразделе обратим внимание на возможности работы с Интернет-архивом для поиска ресурсов и формирования источниковой базы исследования.

Интернет-архив<sup>184</sup> является крупнейшим веб-архивом современности. Сервис Wayback Machine<sup>185</sup> Интернет-архива создан для поиска по архивному вебу и предоставляет возможность реализации запросов по названию ресурса или URL-адресу, где в определенный момент времени располагался ресурс. Представление информации по каждому заархивированному сайту организовано на основе следующих разделов сервиса, представляющих разные сведения и имеющих специфичный функционал:

**CALENDAR** – открывает доступ к снимкам сайта. В разделе представлена интерактивная шкала времени, где можно выбрать год, а на отобразившемся календаре за этот год определить месяц, день и время созданного снимка для просмотра сохраненного сайта (или его отдельных страниц). Временная шкала также позволяет пользователю ориентироваться в количестве сканов за месяц или целый год. В календаре дни месяца, когда осуществлялось сканирование сайта или его отдельных страниц краулером Интернет-архива, обозначены кружками разных цветов и размеров. Цвета кружков вокруг того или иного дня характеризуют качество сканирования сайта. Естественно, если день календаря не отмечен цветом, то в это время сканирование краулером веб-архива не осуществлялось, и значит, что снимки за этот день отсутствуют.

**Голубой** цвет обозначает хорошее качество захвата, отсутствие ошибок при сканировании, и снимки, сделанные в данные дни, имеют более высокое качество и рекомендуются к использованию. **Зеленый** цвет указывает на те снимки, во время сканирования которых были обнаружены редиректы (перенаправления) с одной страницы на другую, что означает, что не все страницы были доступны для сканирования и поэтому не отображены в снимках сайта за данный период. **Оранжевый** и **крас-**

---

<sup>184</sup> Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).

<sup>185</sup> Wayback Machine. Internet Archive. URL: <https://archive.org/web/> (дата обращения: 01.05.2022).



**ный** цвета обозначают существенные ошибки технического характера, как например, отсутствие ответа клиента сервера при сканировании или другие сложности. Такие снимки не рекомендуются для использования ввиду низкого качества сохраненной информации и большой вероятности отсутствия существенной части контента веб-страницы в архиве.

Размер кружка вокруг даты позволяет ориентироваться в количестве предпринятых попыток сканирования страницы. Соответственно, если было реализовано несколько сканов, то во всплывающем при наведении курсором на день календаря окне из списка интерактивных ссылок можно выбрать время сканирования с большей точностью или с наилучшим качеством сохранения (цвет ссылок подскажет о качестве сохраненного веб-контента). Возникающие при выборе времени снимка подсказки также сообщают подробности об обстоятельствах сканирования страницы краулером (например, указывают на то, что сканирование было предпринято командой Интернет-архива, или страница была сохранена по рекомендации пользователей с помощью кнопки “Save Page Now”). Помимо обстоятельств появления снимка в архиве отражаются сведения о том, кто сохранил страницу в Интернет-архив и в какой коллекции она находится.

Следует обратить внимание на важное обстоятельство: сайт в совокупности всех его страниц не всегда может быть сохранен полностью в один день. При переходе с одного снимка на другой по ссылке в структуре сайта сервис Wayback Machine автоматически направляет на снимок, созданный в ближайшую дату к выбранному снимку.

При переходе на снимок сайта появляется меню навигации Wayback Machine, которое посредством шкалы времени позволяет отслеживать плотность сканов страницы краулером и перемещаться по версиям страницы, сохраненным в тот или иной момент времени за весь период существования сайта по данному URL-адресу. Это меню также дает важные сведения о дате создания снимка и данные о структуре страницы и ее отдельных элементах (гиперссылках, медиа-файлах и т.д.), которые можно посмотреть в разделе “About this capture”. Кроме того, в меню предусмотрена кнопка сохранения снимка в индивидуальной коллекции в личном кабинете пользователя.

**COLLECTION** – представляет информацию о коллекциях, в которые были сохранены страницы. По табличной форме удобно ориентироваться по коллекциям и периодам, когда сохранялась информация. При необходимости можно сравнить коллекции за разные годы, что дает интересные

данные о причинах сохранения страниц. Например, сайт Музеев Пермского университета в 2019 г. не только сохранялся целенаправленно краулером Alexa Интернет-архива, но захватывался после перехода по ссылке с сайта Википедии<sup>186</sup>.

**CHANGES** – позволяет визуализировать обновления страницы во времени. Сервис представляет календарь в виде диаграммы, в которой дни месяца обозначены ячейками с цветом. Белый цвет обозначает, что сканирование в данный момент времени не проводилось. Серый цвет ячейки указывает на то, что страница не изменялась со времени последнего сканирования. Диапазон от желтого к синему показывает динамику уровня изменений страницы, где желтый тон указывает на низкий уровень, а синим цветом выделены дни, в которые краулер при сканировании страницы определил наиболее существенные изменения. Нажатие на цветную ячейку загружает сервис визуального сравнения двух снимков страниц, находящихся в разное время по одному адресу, и подсвечивает цветом изменения этих вариантов сохраненных снимков.

**SUMMARY** – показывает информацию для всего домена (в отличие от предыдущих сервисов, предоставляющих данные по отдельным страницам согласно определенному URL-адресу) и отражает форматы информации, размещенной в данном домене. С помощью интерактивной круговой диаграммы сервис визуализирует основные цифровые форматы данных (текст, изображение и т.д.) и изменение их соотношения по годам. Другие визуализации позволяют оценить появление новых страниц в домене в сравнении с существующими и количеством сканов сайта краулером Интернет-архива в динамике за всю историю сайта в сети.

**SITE MAP** также предоставляет информацию по всему домену и отображает карту сайта в динамике по годам. В центре кольцевой диаграммы находится корневой каталог, от которого на разных уровнях в зависимости от иерархии сайта располагаются разделы сайта и отдельные его страницы. Данные визуализации позволяют оценивать изменение комплексности и структуры сайта во времени, анализировать подходы к организации информации и их изменение.

Сервисы Интернет-архива не только позволяют просматривать заархивированные веб-ресурсы, но и предоставляют возможность пользовате-

---

<sup>186</sup> См. Музеи Пермского университета. Wayback Machine. Internet Archive. URL: [https://web.archive.org/web/collections/20190715000000\\*/http://museum.psu.ru/](https://web.archive.org/web/collections/20190715000000*/http://museum.psu.ru/) (дата обращения: 01.05.2022).

лям сохранять страницы из «живого» веба на своей платформе. Функция “Save Page Now” расположена на главной странице платформы WayBack Machine и является полезной при проведении исследований и написании научных работ. Она позволяет гарантированно заархивировать нужную веб-страницу для последующего доступа к ее снимку в Интернет-архиве и создать постоянную ссылку для библиографического описания. При сохранении страницы можно выбрать необходимые опции и размещать контент в персональной коллекции.

Личный кабинет пользователя Интернет-архива обеспечивает разные функциональные возможности для работы с ресурсами, в том числе создание пользовательских подборок ресурсов платформы и быстрый доступ к ним, загрузка собственных цифровых материалов на платформу, пополнение личной библиотеки на основе избранных материалов Интернет-архива. Также предлагаются инструменты для организации персонального веб-архивного контента, размещенного в Интернет-архиве с помощью функции “Save Page Now”.

Интернет-архив имеет открытый доступ к API, что определяет возможности для широкого использования данных веб-архива и реализации масштабных проектов. Эта функция популярна у исследователей веб-истории, хотя и требует существенных технических навыков для полноценного применения.

Необходимо обратить внимание на некоторые ограничения платформы WayBack Machine и, в частности, на ее поисковые возможности – идентификация веб-ресурсов по ключевым словам реализуется только по названию и метаданным сайтов, и отсутствует поиск по их содержанию. Кроме того, платформа WayBack Machine не предусматривает возможностей сортировки и фильтрации материалов по всему массиву найденных ресурсов.

### 5.3.2. Archive-It

Некоторые расширенные поисковые возможности по контенту Интернет-архива воплощены в сервисе Archive-It<sup>187</sup>. Содержательно он выполняется организациями, ведущими веб-архивные инициативы, и предлагает партнерам поддержку в создании, хранении и обеспечении доступа к создаваемым веб-архивным коллекциям. Для всех пользователей этот

---

<sup>187</sup> Archive-It. URL: <https://archive-it.org> (дата обращения: 01.05.2022).

сервис открывает возможности поиска по созданным тематическим коллекциям проекта. По состоянию на май 2022 г. на платформе выложено 9 620<sup>188</sup> тематических коллекций, созданных 1 069<sup>189</sup> партнерскими организациями, что составляет 1 467 641<sup>190</sup> сайт.

Сервис Archive-It предоставляет возможность просмотра полного списка созданных коллекций или организаций, осуществляющих проекты по архивированию веба. Поисковая система позволяет сортировать данные по следующим характеристикам:

- 1) тип организации (в т.ч. академические и образовательные организации высшей ступени, музеи, архивы, библиотеки, некоммерческие организации и т.д.);
- 2) предмет (в т.ч. такие разделы как общество и культура, искусство и гуманитарные науки, блоги и социальные медиа и т.д.),
- 3) создатель коллекции (указана организация, инициировавшая архивирование веба);
- 4) издатель (организация, публикующая материалы);
- 5) контрибьютор (включает организации, внесшие вклад в создание коллекции);
- 6) дата (по годам),
- 7) тип контента (коллекция, сайт, веб-страница и их совокупности, интерактивный ресурс и др.),
- 8) формат контента (текст, изображение, видео-, аудио- и др.);
- 9) язык (указывается основной язык созданной коллекции);
- 10) зона охвата (например, географические характеристики территории, тематически связанной с созданной коллекцией ресурсов);
- 11) организация-составитель коллекции.

Поиск по коллекциям возможен как по названиям коллекций, так и по их содержанию. Контекстный поиск может быть реализован по фразам, по определенному формату контента или задан временными рамками по дате сканирования и архивирования ресурса. Расширенные возможности поиска по коллекциям проекта Archive.It и создают очевидные удобства

---

<sup>188</sup> Collections. Explore All Archives. Archive-It. URL: <https://archive-it.org/explore/?show=Collections> (дата обращения: 01.05.2022).

<sup>189</sup> Organizations. Explore All Archives. Archive-It. URL: <https://archive-it.org/explore/?show=Organizations> (дата обращения: 01.05.2022).

<sup>190</sup> Sites. Explore All Archives. Archive-It. URL: <https://archive-it.org/explore/?show=Sites> (дата обращения: 01.05.2022).

пользователям для реализации более точных запросов и обнаружения искомого контента.

Обратите внимание, что при работе с любым веб-архивом и любой платформой, представляющей коллекции архивного веба, следует учитывать особенности процесса веб-архивирования и организации коллекций. Для использования источников необходимо выявить предвзятости, связанные с партнерскими организациями и приоритетами в формировании коллекций, о которых говорилось ранее (см. Гл. 2-3). В отношении платформы Archive-It следует отметить, что среди партнеров проекта преобладают организации из США, что сказывается на выборке и тематике представленных на платформе ресурсов.

### 5.3.3. Memento

Развитие веб-архивных инициатив в глобальном масштабе привело к тому, что порции веба стали сохраняться параллельно в нескольких веб-архивных проектах. Чтобы объединить возможности поиска по разным веб-архивам был создан сервис Memento<sup>191</sup>. В режиме одного окна сервис Memento позволяет осуществлять поиск сведений по ресурсам целого спектра веб-архивных проектов: Archive.today<sup>192</sup>, Archive-It, Arquivo.pt: the Portuguese Web Archive, Bibliotheca Alexandrina Web Archive, DBpedia archive, DBpedia Triple Pattern Fragments archive, Canadian Government Web Archive, Croatian Web Archive, Estonian Web Archive, Icelandic web archive, Internet Archive, Library of Congress Web Archive, NARA Web Archive, National Library of Ireland Web Archive, National Records of Scotland, perma.cc, PRONI Web Archive, Slovenian Web Archive, Stanford Web Archive, UK Government Web Archive, UK Parliament's Web Archive, UK Web Archive, Web Archive Singapore, WebCite, Bayerische Staatsbibliothek. Дополнительным удобством работы с сервисом является наличие расширений для браузеров (Chrome и Firefox), которые удобны для осуществления поиска по сохраненным в веб-архивах снимкам сайтов, не покидая страницы «живого» веба.

Частью проекта Memento являются сервисы Time Travel и Memento Tracer. Инструменты Time Travel<sup>193</sup> позволяют находить версии веб-сайтов

---

<sup>191</sup> The Memento Project. URL: <http://mementoweb.org> (дата обращения: 01.05.2022).

<sup>192</sup> Archive.today. URL: <http://archive.today> (дата обращения: 01.05.2022).

<sup>193</sup> Time Travel. Memento. URL: <https://timetravel.mementoweb.org/> (дата обращения: 01.05.2022).

по заданному URL-адресу, наиболее близко сохраненные в конкретное время определенной даты, из полного списка веб-архивных проектов, упомянутых выше, а также реконструировать внешний вид и содержание искомых страниц.

Memento Tracer<sup>194</sup> предназначен для сбора отдельных веб-публикаций для их последующего архивирования. Подход к сохранению веб-страниц существенно отличается от сплошного сканирования, являющегося одним из основных подходов в сохранении веб-сайтов, и заключается в том, чтобы были заархивированы только те страницы, которые являются значимыми для куратора, сохраняющего нужные данные. При этом сервис способен отслеживать действия куратора и определенным образом реагировать на его активность, предлагая контент для архива.

#### **5.3.4. Webrecorder**

Webrecorder.io (или Conifer)<sup>195</sup> является сервисом с открытым программным кодом, который направлен на создание коллекций веб-ресурсов, сохраненных в формате архивного веб-контента WARC. В основе подхода, реализованного в Webrecorder, лежит идея, близкая Memento Tracer, когда сохраняются только страницы, определенные куратором. Webrecorder отслеживает переходы пользователя, сохраняющего контент, и архивирует указанные им данные. Куратор данных управляет всем процессом с помощью веб-браузера. Сервис не только позволяет сохранить веб-страницы, но и делает это таким образом, чтобы заархивированный контент можно было воспроизвести в аутентичном виде.

#### **5.3.5. GLAM Workbench**

Инструментарий проекта GLAM Workbench<sup>196</sup> посвящен сектору культурных учреждений, таких как галереи, музеи, архивы и библиотеки, и предназначен для изучения их ресурсов, сохраненных в веб-архивах. Несмотря на то, что данный ресурс сфокусирован в основном на организациях в Австралии и Новой Зеландии, с его помощью можно отследить

---

<sup>194</sup> Memento Tracer. Memento. <http://tracer.mementoweb.org/> (дата обращения: 01.05.2022).

<sup>195</sup> Webrecorder.io. URL: <https://webrecorder.net/>; см. также Conifer. URL: <https://conifer.rhizome.org/> (дата обращения: 01.05.2022).

<sup>196</sup> GLAM Workbench. URLs: <https://glam-workbench.github.io/>; <https://glam-workbench.net/> (дата обращения: 01.05.2022).

информацию и о веб-ресурсах, созданных в других странах и принадлежащих к разным национальным доменам.

Разработка коллекции Jupyter notebooks для работы с данными веб-архивов предусмотрена в рамках проекта. Jupyter notebooks – это приложение, позволяющее представить искомые данные из веб-архивов на основе изменения тех или иных параметров в строках готового кода. Достоинство приложения заключается в том, что оно позволяет строить визуализации для анализа данных без существенных навыков программирования, что доступно новичкам.

Проект GLAM Workbench представляет собой исследовательскую информационную инфраструктуру, с помощью которой можно находить и анализировать данные, связанные с заархивированными ресурсами музеев, архивов и библиотек.

### 5.3.6. Другие сервисы

Разработаны также и другие инструменты для работы с веб-архивными ресурсами и формирования источниковой базы исследования. В частности, расширение Mink<sup>197</sup> для браузера Chrome позволяет отслеживать сколько раз и в каких архивах сохранялась та или иная веб-страница, что облегчает поиски необходимой архивной информации.

Альтернативным браузерным расширением для целей веб-архивирования и создания персональных коллекций архивного веба является WARCreate<sup>198</sup> (работает в браузере Chrome). Расширение разработано для создания веб-архивных файлов с расширением WARC (от Web Archive), являющимся стандартным форматом для данного типа ресурсов. Расширение WARCreate сохраняет страницу в совокупности со всем контентом, метаданные, а также сведения о структуре и организации документа для последующего воспроизведения заархивированного веб-контента.

Воспроизведение файлов в формате WARC может осуществляться с помощью приложения WAIL (Web Archiving Integration Layer)<sup>199</sup>. Приложение требует установки на персональный компьютер пользователя, является свободно распространяемым программным обеспечением. Альтерна-

---

<sup>197</sup> Mink – Integrate Live & Archived Web +Memento. URL: <https://chrome.google.com/webstore/detail/mink-integrate-live-archi/jemoalkmipibchioofomhkgimhofbbem> (дата обращения: 01.05.2022).

<sup>198</sup> WARCreate. URL: <https://warcreate.com/> (дата обращения: 01.05.2022).

<sup>199</sup> Web Archiving Integration Layer (WAIL). URL: <http://machawk1.github.io/wail/> (дата обращения: 01.05.2022).

тивным вариантом программ с подобным назначением является Webrecorder Player<sup>200</sup>.

#### 5.4. Внешняя и внутренняя критика источников

Изучение веб-архивов для последующего использования их ресурсов в исторических исследованиях пересекается с предметным полем цифровой истории, поскольку отсылает к производству исторического знания на основе цифровых материалов<sup>201</sup>. Современные исследователи выделяют **дигитальность** как ключевой признак современной истории и неотъемлемую характеристику веб-истории<sup>202</sup>, которая проявляется в расширении процессов оцифровки (перевода источников из аналогового в электронный формат) и цифровизации в области истории, более широкого тренда, в том числе, связанного с сохранением цифрового культурного наследия и производством знания на основе цифровых методов, публикацией в открытой информационной среде наборов данных и полученных результатов. Дигитальность непосредственно влияет на все процессы и процедуры проведения исторического исследования – от формулировки исследовательского вопроса и формирования источниковой базы до ее анализа, интерпретации и публикации результатов. Следовательно, требует своего внимания и специфика источников, созданных изначально в цифровом формате (born-digital resources), к каким относятся веб-ресурсы (см. также Гл. 1–2).

Использование материалов веб-архивов в исторических исследованиях требует проведения внешней и внутренней критики в отношении цифровых источников. Источниковедческая критика связана с анализом способов сохранения источника, что относится к изучению особенностей формирования веб-архивных материалов (см. Гл. 2), а также с исследованием самого источника и информации о нем.

---

<sup>200</sup> Webrecorder Player. URL: <https://github.com/webrecorder/webrecorder-player> (дата обращения: 01.05.2022).

<sup>201</sup> См. более подробно о цифровой истории Володин А.Ю. «Цифровая история: ремесло историка в цифровую эпоху» // Электронный научно-образовательный журнал «История». 2015, 6 (8). [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001228-9-1> (дата обращения: 01.05.2022); Salmi H. What is Digital History? Medford: Polity Press, 2020. 130 p.

<sup>202</sup> Salmi H. What is Digital History? Medford: Polity Press, 2020. P. 1.



А. Фикерс формулирует пять основных вопросов, которые следует задать по отношению к источнику, осуществляя источниковедческую критику<sup>203</sup>:

1. Кто создал / произвел источник (автор)?
2. К какому типу этот документ относится (жанр и особенности использования языка)?
3. Где он был создан и распространялся (распространение и аудитория)?
4. Когда он был создан (дата и период)?
5. Зачем он был создан (намерение)?

Важным компонентом источниковедческой критики является рассмотрение способов и условий формирования электронных коллекций, наличие возможностей оценки полноты данных и определения репрезентативности выборки. Х. Салми в книге “What is Digital History?”<sup>204</sup> выделяет следующие факторы, которые должны учитываться при работе с цифровыми источниками и являются справедливыми и для исследований в области веб-истории:

1) **партнерства**, которые могут быть частными, общественными, государственными, смешанными и др. и определяют организации, участвующие в формировании цифровых коллекций, в том числе, финансирующие проекты по сохранению веба. Партнерства неизбежно оказывают влияние на полноту выборки, а также на содержание сохраняемого контента. Петер Вебстер обращает внимание на предварительные этапы анализа источников и указывает на то, что перед тем, как интерпретировать источник, нужно узнать, каким образом он попал в архив<sup>205</sup>, с какой целью архив создан и чьим интересам служит. Это вопросы довольно сложны для детального анализа, поскольку не вся информация может быть доступна и отражена на ресурсах, либо эти сведения довольно сложно обнаружить. Организация, создающая веб-архив, может не сопровождать процесс сканирования генерированием метаданных об этом и публикацией

---

<sup>203</sup> Fickers A. (2012) Towards a New Digital Historicism? Doing History In The Age Of Abundance // Journal of European History and Culture, Vol. 1(1). P. 7. URL: <https://orbilu.uni.lu/bitstream/10993/7615/1/4-4-1-PB.pdf> (дата обращения: 01.05.2022).

<sup>204</sup> См. параграф “Biases of the Digital Past”. Salmi H. What is Digital History? Medford: Polity Press, 2020. Pp. 17-19.

<sup>205</sup> Webster P. Users, technologies, organisations: Towards a cultural history of world webarchiving. Essay. URL: [https://peterwebster.me/wp-content/uploads/2017/08/webster\\_historyofwebarchiving\\_postprint.pdf](https://peterwebster.me/wp-content/uploads/2017/08/webster_historyofwebarchiving_postprint.pdf) (дата обращения: 01.05.2022).

комментариев с подобного рода сведениями. Поэтому при анализе источников и наборов данных должны рассматриваться партнерства и потенциальные предвзятости в формировании коллекций, которые могут иметь политический, экономический, социальный или иной характер;

2) **временной ритм** проявляется в сохранении информации отдельными фрагментами в зависимости от интереса к тем или иным событиям или временному пулу. В результате получается выборка, лимитированная определенным периодом, в которую не попадает информация, выходящая за установленные границы, что сужает аналитические возможности. Вариацией временного фактора является зависимость сохранения и репрезентации источников в веб-архивах от технологического прогресса и адаптивности технологий. Различная полнота сохраненного контента, возможности захвата и публикации разных форматов цифрового контента отражаются на состоянии коллекций веб-архивов в разные временные отрезки, различия в сложности и возможностях технологий обуславливают разную глубину и качество сохранения контента;

3) **масштаб оцифровки** напрямую влияет на создаваемые коллекции материалов. Естественным образом программы оцифровки точечного, регионального, национального и транснационального уровней определяют приоритеты и подробности сохраняемого материала. Необходимо также обратить внимание на неравномерное распределение информации в транснациональных проектах, какими, например, являются Europeana<sup>206</sup>, в которой представлены данные из культурных учреждений разных стран мира, преимущественно европейские, и Google Arts&Culture<sup>207</sup>, по отношению к которому ученые указывали на проблемы цифрового колониализма<sup>208</sup>. Для веб-истории вопросу о масштабах оцифровки сопутствует проблема определения полноты информации и ее репрезентации в веб-архивах в разных доменных зонах и тематических коллекциях. Поэтому необходимо принимать во внимание постоянную неполноту выборки. Нильс О. Финнеман предостерегает исследователей веб-архивных коллек-

---

<sup>206</sup> Europeana. URL: <https://www.europeana.eu> (дата обращения: 01.05.2022).

<sup>207</sup> Google Arts&Culture. URL: <https://artsandculture.google.com> (дата обращения: 01.05.2022).

<sup>208</sup> Кижнер И.А., Румянцев М.В., Хохлова В.Н., Рудов И.Н., Афанасьева Ю.Ю. Цифровые коллекции в области культурного наследия и развитие цифровой инфраструктуры для изучения культуры (на примерах цифровой платформы Google Arts and Culture и цифровой библиотеки «Европеана») // Сибирский антропологический журнал. 2020. № 2 (06). С. 103–110.

ций от распространения выводов на все подобные ресурсы ввиду изначальной предвзятости при формировании выборки<sup>209</sup>;

4) **правовые нормы и правила**, которые не только регулируют возможности публикации источников, но могут создавать существенные ограничения их использования. Авторские права на цифровой контент, необходимость получения разрешений, согласование условий и другие факторы сдерживают общий прогресс в использовании цифровых ресурсов. Определенные решения этих сложностей предлагают подходы открытой науки<sup>210</sup>, представление о всеобщем достоянии (Public domain), нулевые лицензии в отношении прав на цифровые ресурсы (Creative Commons Zero<sup>211</sup>). Однако эти подходы только начинают распространяться и еще не используются большинством организаций, хотя их применение уже показывает существенные результаты. Кроме того, веб-архивирование реализуется преимущественно в открытой зоне, доступной для работы краулеров, поэтому большинство ресурсов, находящихся в закрытых или коммерческих доменах лишены возможности быть сохраненными. Указанный фактор препятствует широкому архивированию ресурсов, а значит, ставит под угрозу возможность доступа к этим материалам в будущем, и также должен приниматься во внимание при анализе источников;

5) **коммерциализация** процессов оцифровки и обеспечения доступа к цифровому историко-культурному наследию имеет последствия, связанные, например, с наличием платных подписок на создаваемые коммерческими структурами ресурсы. Следствием коммерциализации становится неравенство возможностей доступа к создаваемому цифровому контенту. В отношении веб-архивов вопросы коммерциализации не столь очевидны, как в случае с оцифрованным наследием, однако на коммерческой основе могут создаваться отдельные коллекции, и ввиду возможных предвзятостей в их формировании это обстоятельство следует принимать во внимание.

---

<sup>209</sup> Finnemann N. O. (2018) Web archives and Knowledge organization. In: Birger Hjørland (Ed.) ISKO Encyclopedia of Knowledge Organization & Knowledge Organization. URL: [https://www.isko.org/cyclo/web\\_archive](https://www.isko.org/cyclo/web_archive) (дата обращения: 01.05.2022).

<sup>210</sup> См. Open Science. UNESCO. <https://en.unesco.org/science-sustainable-future/open-science> (дата обращения: 01.05.2022); What is Open Science? Introduction. FOSTER. URL: <https://www.fosteropenscience.eu/content/what-open-science-introduction> (дата обращения: 01.05.2022).

<sup>211</sup> CC0 1.0 Универсальная (CC0 1.0). Передача в общественное достояние. URL: <https://creativecommons.org/publicdomain/zero/1.0/deed.ru> (дата обращения: 01.05.2022).

При проведении источниковедческой критики также следует учитывать особенности данных, рожденных в цифровом формате. Одной из таких характеристик является динамичность веб-контента, его постоянная обновляемость и наращивание информационных слоев. Это свойство весьма затруднительно сохранить и представить в веб-архиве. Наполнение веб-ресурсов – это динамичный процесс, гибкий, неравномерный во времени, трудно подверженный оценке этих изменений. Постоянное обновление и «наращивание» разных информационных слоев сайтов создают ощущение присутствия в данный момент времени как «постоянное ощущение момента сейчас»<sup>212</sup>. Архив фиксирует текущий момент в определенной точке и не позволяет представить динамику. Эта особенность должна приниматься во внимание как при формулировке исследовательского вопроса, так и при сборе и организации данных и анализе источниковой базы.

Критика веб-архивных источников предполагает также проверку источников на **надежность и достоверность**. Как следствие, важным навыком является умение отличать ресурсы, которые заслуживают доверия, от тех, которые могут содержать фальсификации. В этом отношении следует понимать разницу между фактом и лжесвидетельством, фейком, а также необходимо знать критерии проверки ресурсов на достоверность и уметь их применять при анализе ресурсов.

Веб-ресурсы могут относиться к базам знаний, отражать факты, документировать события, либо являться недостоверными или теми, чью природу верифицировать невозможно ввиду отсутствия информации о создателях, степени того, насколько доступ к представленной информации ее источникам является авторизованным. Важным показателем является прозрачность происхождения информации. Кроме того, для цифровых источников серьезной проблемой является установление подлинности опубликованных сведений, определение границ и различий между оригиналом и архивной копией, а также поиск ресурса-первоисточника тех или иных данных.

В эру постправды<sup>213</sup> определение истинности, правдивости цифрового источника становится проблематичным делом ввиду вариативности интерпретаций, калейдоскопа мнений, а также разнообразия типового состава информационных ресурсов (в том числе, социальных медиа, различных платформ и площадок для коммуникации). Гибкость информационных ре-

---

<sup>212</sup> Belovari S. (2017) Historians and Web Archives // *Archivaria*, Vol. 83. Pp. 59–79.

<sup>213</sup> Якимов А. Е. Постправда и повседневность. К проблеме определения понятия «постправда» // *Философия и культура*. 2020. № 9. С. 1–8. DOI 10.7256/2454-0757.2020.9.33801.

сурсов и отсутствие строгой онтологии осложняют определение типов ресурсов и их границ, а наличие различных смежных и пограничных вариантов затрудняют процесс разностороннего источниковедческого анализа этих ресурсов.

Кроме того, для born-digital источников существенной проблемой является еще и разделение первичных и вторичных источников, что во многом связано с гибкостью и размытостью типов информационных ресурсов. Идентификация тех или иных веб-ресурсов как первичных (вторичных) источников восходит к эпистемологии, источниковедческой критике и связана с необходимостью прохождения всех этапов получения научного знания при проведении научного исследования. Важным компонентом в этом направлении является определение подлинности источника, а также выделение аргументов о том, является ли тот или иной веб-ресурс первичным источником или представляет собой копию. Отличить копию от оригинала в цифровой среде особенно сложно, поскольку тексты подвержены копирейту, рерайту и другим манипуляциям с текстовыми источниками, характерным для современности. Аналогичные проблемы могут быть присущи и другим, нетекстовым, форматам источников.

Критика источников подразумевает проверку ресурсов на их авторитетность и авторизованный характер по отношению к тому материалу, который та или иная платформа презентует. Например, сайты музеев, архивов и библиотек являются авторизованными площадками для репрезентации историко-культурного наследия и памяти. Назначение информационных ресурсов этих организаций напрямую связано с выполнением их функций, а цифровой контент может быть при необходимости верифицирован, то есть проверен на соответствие оригиналу. Подобным образом, авторизованными являются научные и образовательные учреждения, информационные ресурсы которых также могут быть отнесены к базам знаний. В частности, если ресурс создан библиотекой, архивом или, например, академической организацией, то это априори предполагает высокую степень доверия такому ресурсу. В то же время, определение создателя того или иного веб-ресурса, анализ его авторитетности и оценка степени доверия осуществляются намного сложнее, если ресурс не содержит отсылок к организационной принадлежности. В случае указания персоналий создателей без аффилиации, требуется определение их позиций, рода деятельности и других процедур для определения предвзятостей и границ

личной ответственности за создаваемый ресурс и информацию, которая на нем размещена.

При анализе веб-ресурсов необходимо понимать аудиторию, для которой они предназначены, инициаторов ресурсов и производителей контента, а также уметь определять цели и задачи функционирования этих ресурсов. Можно провести аналогию с изучением газет в исторической ретроспективе: чтобы интерпретировать информацию, опубликованную в определенном издании, и понять, какие свидетельства о прошлом оно несет, нужно знать – кто его финансировал, какие журналисты, с каким опытом, персональной позицией по разным вопросам (в первую очередь политическим) и бэкграундом писали материалы, на какие аудитории читателей было ориентировано издание и т.д.

Контекстуализация источников является другим важным компонентом источниковедческого анализа веб-архивных материалов при проведении исторического исследования. Сам по себе цифровой источник и его наличие в веб-архиве не говорят о его значимости. Понимание ценности заархивированного веб-ресурса возможно только на основе помещения его в контекст событий, явлений и процессов, которые происходили в тот момент времени, когда источник был создан. В этом отношении необходимы оценка и анализ исследуемого объекта в отношении естественной среды его бытования, происхождения, условий и обстоятельств, в которых он создавался и функционировал. В веб-архиве сайт находится в отрыве от контекста использования, решения определенных задач, что подразумевает необходимость реконструкции контекста существования ресурса.

\*\*\*

Таким образом, несмотря на представленные сложности работы с веб-архивами и необходимость тщательной источниковедческой критики, с этими ограничениями можно работать. В целом, веб-архивы и комплекс сервисов формируют благоприятную среду как для использования в образовательном процессе в рамках курса для историков, так и при реализации исследовательских проектов в области веб-истории различного масштаба.

Обратите внимание на оформление стандартного библиографического описания при цитировании веб-архивных материалов:

1) для цитирования снимка сохраненного сайта в целом: название сайта, дата сохранения в веб-архиве, веб-архив, ссылка, дата обращения. Например: Русский музей. Снимок от 01.03.2016. Internet Archive. URL:

<https://web.archive.org/web/20160301072341/http://rusmuseum.ru/>  
(дата обращения: 01.05.2022).

2) для цитирования отдельных страниц: заглавие страницы, название сайта, дата сохранения в веб-архиве, веб-архив, ссылка, дата обращения. Например: Древнерусское искусство XII–XVII вв. Русский музей. Снапшот от 12.03.2016. Internet Archive. URL: <https://web.archive.org/web/20160312162339/http://www.rusmuseum.ru/mikhailovsky-palace/exhibitions/ancient-art-of-the-xii-xvii-centuries/> (дата обращения: 01.05.2022).

### Рекомендуемая литература

1. Володин А. Ю. «Цифровая история»: ремесло историка в цифровую эпоху // Электронный научно-образовательный журнал «История». 2015, вып. 6(8). [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001228-9-1> (дата обращения: 01.05.2022).

2. Кижнер И. А., Румянцев М. В., Хохлова В. Н., Рудов И. Н., Афанасьева Ю. Ю. Цифровые коллекции в области культурного наследия и развитие цифровой инфраструктуры для изучения культуры (на примерах цифровой платформы Google Arts and Culture и цифровой библиотеки «Европеана») // Сибирский антропологический журнал. 2020. № 2 (06). С. 103–110.

3. Якимов А. Е. Постправда и повседневность. К проблеме определения понятия «постправда» // Философия и культура. 2020. № 9. С. 1–8. DOI: 10.7256/2454-0757.2020.9.33801.

4. Ackland R., Evans A. (2017) Using the web to examine the evolution of the abortion debate in Australia, 2005–2015. In: Niels Brügger, Ralph Schroeder (Eds.) The Web as History. Using Web Archives to Understand the Past and the Present. Pp. 159–189.

5. Belovari S. (2017) Historians and Web Archives // Archivaria. Vol. 83. Pp. 59–79. URL: <https://archivaria.ca/index.php/archivaria/article/view/13600/14985> (дата обращения: 01.05.2022).

6. Brugger N. (2008). The archived website and website philology: A new type of historical document? // Nordicom Review. Vol. 29(2). Pp. 151–171. URL: <https://www.nordicom.gu.se/en/tidskrifter/nordicom-information-42008/archived-website-and-website-philology-new-type-historical> (дата обращения: 01.05.2022).

7. Fickers A. (2012) Towards a New Digital Historicism? DOIing History in the Age of Abundance // Journal of European History and Culture. Vol. 1(1), p. 7. URL: <https://orbilu.uni.lu/bitstream/10993/7615/1/4-4-1-PB.pdf> (дата обращения: 01.05.2022).

8. Finnemann N. O. (2018) Web archives and Knowledge organization. In: Birger Hjørland (Ed.) ISKO Encyclopedia of Knowledge Organization & Knowledge Organization. URL: [https://www.isko.org/cyclo/web\\_archive](https://www.isko.org/cyclo/web_archive) (дата обращения: 01.05.2022).

9. Halevy A., Norvig P., Pereira F. (2009) The Unreasonable Effectiveness of Data In: IEEE Intelligent Systems. Vol. 24, № 2. Pp. 8–12, March-April 2009, DOI: 10.1109/MIS.2009.36. URL: <https://ieeexplore.ieee.org/document/4804817> (дата обращения: 01.05.2022).

10. Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).

11. Salmi H. What is Digital History? Medford: Polity Press, 2020. 130 p.

12. The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. SAGE reference, Los Angeles. 2019. 630 p. DOI:<http://dx.doi.org/10.4135/9781526470546>.

13. The Web as History. Using Web Archives to Understand the Past and the Present. Eds. Niels Brügger, Ralph Schroeder. London, UCL Press, 2017. <https://doi.org/10.14324/111.9781911307563>.

14. Webster P. Users, technologies, organisations: Towards a cultural history of world webarchiving. Essay. URL: [https://peterwebster.me/wp-content/uploads/2017/08/webster\\_historyofwebarchiving\\_postprint.pdf](https://peterwebster.me/wp-content/uploads/2017/08/webster_historyofwebarchiving_postprint.pdf) (дата обращения: 01.05.2022).

## Источники

1. Историческая Информатика. URL: <https://e-notabene.ru/istinf> (дата обращения: 01.05.2022).

2. Клуб любителей Интернета и общества. URL: <http://clubforinternet.net/> (дата обращения: 01.05.2022).

3. Российская ассоциация «История и Компьютер». URL: <http://aik-sng.ru/> (дата обращения: 01.05.2022).

4. Archive.today. URL: <http://archive.today/> (дата обращения: 01.05.2022).

5. Archived Webpage Collections. Indigenous Australians. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au/collection/15012> (дата обращения: 01.05.2022).

6. Archive-It. URL: <https://archive-it.org> (дата обращения: 01.05.2022).

7. CC0 1.0 Универсальная (CC0 1.0). Передача в общественное достояние. URL: <https://creativecommons.org/publicdomain/zero/1.0/deed.ru> (дата обращения: 01.05.2022).

8. Collections. Explore All Archives. Archive-It. URL: <https://archive-it.org/explore/?show=Collections> (дата обращения: 01.05.2022).

9. Conifer. URL: <https://conifer.rhizome.org/> (дата обращения: 01.05.2022).



10. Europeana. URL: <https://www.europeana.eu/> (дата обращения: 01.05.2022).
11. GLAM Workbench. URLs: <https://glam-workbench.github.io>; <https://glam-workbench.net/> (дата обращения: 01.05.2022).
12. International Internet Preservation Consortium. URL: <https://netpreserve.org/> (дата обращения: 01.05.2022).
13. Internet Archive Blogs. Internet Archive. URL: <http://blog.archive.org/category/video-archive> (дата обращения: 01.05.2022).
14. Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).
15. Internet Histories. Digital Technology, Culture and Society. URL: <https://www.tandfonline.com/toc/rint20/current> (дата обращения: 01.05.2022).
16. Journal of Digital History. URL: <https://journalofdigitalhistory.org/> (дата обращения: 01.05.2022).
17. List of Web archiving initiatives. Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives) (дата обращения: 01.05.2022).
18. Memento Tracer. Memento. <http://tracer.mementoweb.org/> (дата обращения: 01.05.2022).
19. Mink – Integrate Live & Archived Web +Memento. URL: <https://chrome.google.com/webstore/detail/mink-integrate-live-archi/jemoa1kmiplibchioofomhkgimhofbbem> (дата обращения: 01.05.2022).
20. Open Science. UNESCO. <https://en.unesco.org/science-sustainable-future/open-science> (дата обращения: 01.05.2022).
21. Organizations. Explore All Archives. Archive-It. URL: <https://archive-it.org/explore/?show=Organizations> (дата обращения: 01.05.2022).
22. RESAW, a Research Infrastructure for the Study of Archived Web Materials. URL: <https://resaw.eu/> (дата обращения: 01.05.2022).
23. Sites. Explore All Archives. Archive-It. URL: <https://archive-it.org/explore/?show=Sites> (дата обращения: 01.05.2022).
24. The Memento Project. URL: <http://mementoweb.org> (дата обращения: 01.05.2022).
25. Time Travel. Memento. URL: <https://timetravel.mementoweb.org/> (дата обращения: 01.05.2022).
26. WARCreate. URL: <https://warcreate.com/> (дата обращения: 01.05.2022).
27. Wayback Machine. Internet Archive. URL: <https://archive.org/web/> (дата обращения: 01.05.2022).
28. Web Archiving Integration Layer (WAIL). URL: <http://machawk1.github.io/wail/> (дата обращения: 01.05.2022).

29. Webrecorder Player. URL: <https://github.com/webrecorder/webrecorder-player> (дата обращения: 01.05.2022).

30. Webrecorder.io. URL: <https://webrecorder.net/> (дата обращения: 01.05.2022).

31. What is Open Science? Introduction. FOSTER. URL: <https://www.fosteropenscience.eu/content/what-open-science-introduction> (дата обращения: 01.05.2022).

32. Youtube Videos. Internet Archive. URL: <https://archive.org/details/youtubecrawl> (дата обращения: 01.05.2022).

### **Вопросы для самопроверки**

1. В чем заключаются основные особенности современной истории, существенные для ее изучения?

2. На какой основе может быть сформулирован исследовательский вопрос в области веб-истории?

3. Перечислите основные источники для изучения историографии веб-истории и отдельных ее направлений.

4. В чем заключается подход "разоблачение мифов" для постановки исследовательского вопроса?

5. Какие уровни исследований можно выделить согласно объему используемых источников?

6. Какие существуют способы формирования источниковой базы для проведения исследования в области веб-истории?

7. В чем заключается значение тематических коллекций для проведения исследований в области веб-истории?

8. Приведите примеры тематических коллекций веб-архивов и коллекций, созданных на основе разных форматов веб-контента.

9. В чем заключается особенность проведения крупномасштабных исследовательских проектов в области веб-истории?

10. Какой метод формирования источниковой базы был использован при изучении дискуссий об абортах в Австралии?

11. В чем заключается специфика микроуровня исследовательских проектов по отношению к источниковой базе и способам ее изучения?

12. Перечислите достоинства и ограничения открытых веб-архивов для формирования источниковой базы исследования.

13. Опишите назначение сервиса Wayback Machine.

14. Какой функционал имеют отдельные разделы сервиса Wayback Machine?

15. В чем заключаются поисковые ограничения сервиса Wayback Machine?
16. Какое назначение имеет ресурс Archive-It?
17. Какие возможности имеет поисковая система сервиса Archive-It?
18. В чем заключается назначение сервиса Memento?
19. Какие функциональные возможности имеют сервисы Time Travel и Memento Tracer?
20. Какую функцию выполняет Webrecorder.io (или Conifer) для формирования источниковой базы?
21. Перечислите основные особенности инструментария GLAM Workbench?
22. Назовите примеры цифровых инструментов для создания персональных коллекций архивного веба и их воспроизведения.
23. В чем заключается дигитальность как ключевой признак современной истории?
24. Перечислите основные факторы, которые должны учитываться при работе с цифровыми источниками, согласно книге Х. Салми “What is Digital History?”.
25. В чем заключается влияние партнерств на формирование коллекций архивного веба?
26. Какие существуют правовые нормы и правила, регулирующие возможности публикации источников?
27. Какое влияние оказывает коммерциализация процессов оцифровки на доступность цифрового культурного наследия?
28. Какая особенность архивированного веба должна приниматься во внимание при проведении источниковедческой критики?
29. В чем заключается необходимость проверки источников на надежность и достоверность при формировании источниковой базы исследования в области веб-истории?
30. В чем заключается проблема разграничения первичных и вторичных источников по отношению к архивному вебу?
31. Какое значение имеет проверка ресурсов на их авторитетность?
32. Какое влияние оказывает аудитория ресурса для источниковедческого анализа?
33. Перечислите основные компоненты библиографического описания для цитирования снимка сохраненного сайта в целом и его отдельных страниц.

## ГЛАВА 6. МЕТОДЫ И ИНСТРУМЕНТАРИЙ ИССЛЕДОВАНИЯ

Реализация проектов по веб-истории осуществляется на основе методов, которые широко используются в исторических исследованиях в целом<sup>214</sup>. В данном разделе обратим внимание лишь на некоторые моменты, связанные с применением информационных методов к изучению веб-истории. Естественно, что большинство проектов реализуются на основе комбинации разных методов, специально подобранных для решения исследовательских задач, поэтому в тексте пособия приведены отдельные примеры, которые иллюстрируют возможности применения тех или иных методов и исследовательского инструментария для поиска ответов на поставленные вопросы.

Исследовательская инфраструктура веб-архивов и разработанный дополнительный инструментарий для работы с веб-архивными материалами создают условия проведения исследований в направлении веб-истории, в том числе, для изучения истории социума и его институтов. Наиболее распространенным случаем в академической практике является то, что методы исследования подбираются исходя из поставленного вопроса и источниковой базы.

В отчете о будущем развитии веб-архивов, опубликованном в 2011 г., обращалось внимание на необходимость реализации «интероперабельности» исследовательских методов с цифровой инфраструктурой веб-архивов<sup>215</sup>. Использование встроенного функционала веб-архивов является базовой возможностью изучения ресурсов и имеет высокий потенциал применения в исследовательской деятельности. Такой базовый инструментарий позволяет как извлекать отдельные архивные записи, так и формировать датасеты небольших размеров. Работа же с большими объемами данных предполагает более сложные формы работы, инструментарий и аналитические средства.

---

<sup>214</sup> Ковальченко И. Д. Методы исторического исследования. Отделение историко-филологических наук. 2-е изд., доп. М.: Наука, 2003. 486 с. URL: [http://www.hist.msu.ru/departments/8819/research/publications/detail.php?ELEMENT\\_ID=43145](http://www.hist.msu.ru/departments/8819/research/publications/detail.php?ELEMENT_ID=43145) (дата обращения: 01.05.2022).

<sup>215</sup> Meyer E.T., Thomas A., Schroeder R. Web Archives: The Future(s). June 30, 2011. URL: <https://ssrn.com/abstract=1830025> (дата обращения: 01.05.2022).

Исследования в области веб-истории реализуются как на основе методов ближнего, так и дальнего чтения<sup>216</sup>, что является естественным процессом для проектов в области истории<sup>217</sup> и других гуманитарных наук. Методы ближнего чтения основаны на внимательном изучении, т.н. «медленном» чтении источника. Тогда как дальнее чтение (его принципы сформулированы Ф. Моретти<sup>218</sup>) дистанцирует исследователя непосредственно от содержания источника и связано с формализацией текста, выявлением и анализом его отдельных структур. Довольно частыми являются случаи комбинации приемов ближнего и дальнего чтения.

Методы дальнего чтения необходимы для извлечения отдельных пластов обобщенной информации при изучении больших объемов данных. Примером их применения является проект Archives Unleashed Project<sup>219</sup>, предназначенный для визуализации петабайтов информации, хранящихся в веб-архивах. Разрабатываемый инструментарий предназначен для использования исследователями из разных областей наук, в том числе истории. В инструкции на сайте проекта указано, что инструменты могут применяться пользователями с различным уровнем подготовки и опытом программирования. Необходимым условием является обладание доступом к аккаунту Archive-It, который может быть зарегистрирован только партнерской организацией и вестись трудоустроенным в этой организации специалистом, например библиотекарем. Это требование существенно снижает возможности использования инструмента в образовательной деятельности, несмотря на высокую педагогическую и исследовательскую ценность.

Далее перечислим исследовательские методы и примеры их применения в проектах по веб-истории.

---

<sup>216</sup> Jänicke S., Franzini G., Cheema M. F., Scheuermann G. (2015) On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges // The proceedings of the Eurographics Conference on Visualization (EuroVis). Eds.: R. Borgo, F. Ganovelli, and I. Viola. 2015. URL: <https://www.informatik.uni-leipzig.de/~stjaenicke/Survey.pdf> (дата обращения: 01.05.2022).

<sup>217</sup> Володин А. Ю. «Цифровая история»: ремесло историка в цифровую эпоху // Электронный научно-образовательный журнал «История», 2015. Т. 6, вып. 8 (41). DOI: 10.18254/S0001228-9-1. [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001228-9-1> (дата обращения: 01.05.2022).

<sup>218</sup> Моретти Ф. Дальнее чтение / Франко Моретти ; пер. с англ. Алексея Вдовина [и др.] / под науч. ред. Инны Кушнareвой. М.: Изд-во Института Гайдара, 2016. 342 с.

<sup>219</sup> Archives Unleashed Project. URL: <https://archivesunleashed.org> (дата обращения: 01.05.2022).

## 6.1. Моделирование данных

Основой для изучения веб-архивных источников может являться моделирование данных, успешно применяемое в гуманитарных исследованиях в целом<sup>220</sup>, и в исторических науках в частности<sup>221</sup>. Моделирование не является новым подходом к анализу данных, а появилось довольно давно и получило импульс для широкого применения благодаря развитию информационных технологий и цифрового поворота в гуманитарных науках<sup>222</sup>.

Моделирование данных основано на формализации характеристик и выделении сущностных элементов в составе и содержании изучаемого объекта и их анализе. В случае с веб-архивами в структурном плане такими элементами могут служить веб-страницы и их фрагменты, совокупности страниц (разделы каталога, например), а также веб-сайты полностью, а в содержательном – сведения, отраженные в этих элементах.

На базовом уровне модель может быть представлена в виде схемы, отражающей ключевые элементы и их взаимосвязи. Сущностью при моделировании может выступать любой объект, предмет или их совокупность, а также процесс или явление. При этом необходимо помнить, что модель представляет реальность в упрощенном, хотя и структурированном виде. Необходимым шагом при использовании методов моделирования является т.н. валидация, то есть сопоставление полученной модели и оригинала – реального объекта, явления или процесса.

Метод моделирования данных при анализе материалов веб-архивов был успешно применен при изучении цифровой истории виртуальных музеев<sup>223</sup>. В частности, на примерах музейных сайтов из США, Канады, Гер-

---

<sup>220</sup> Ciula A., Eide Ø. (2017) Modelling in digital humanities: Signs in context. In: Digital Scholarship in the Humanities, Volume 32, Issue suppl\_1. Pp. i33–i46.

<sup>221</sup> Бородкин Л. И. Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив. СПб.: Алетейя, 2017. 303 с.

<sup>222</sup> См. примеры создания моделей разных сущностей в публикациях: Кирьянов И. К., Корниенко С. И., Сенина А. В. – Депутатский корпус Государственной думы поздней имперской России: от текста источника к статистическим компьютеризованным методам анализа // Историческая информатика. 2019. № 1. С. 118–127. DOI: 10.7256/2585-7797.2019.1.29148 URL: [https://nbpublish.com/library\\_read\\_article.php?id=29148](https://nbpublish.com/library_read_article.php?id=29148) (дата обращения: 01.05.2022); Гагарина Д. А., Корниенко С. И., Поврозник Н. Г. Информационные системы в цифровой среде исторической науки // Электронный научно-образовательный журнал «История». 2016. Т. 7, вып. 7 (51) [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001638-0-1>. DOI: 10.18254/S0001638-0-1 (дата обращения: 01.05.2022).

<sup>223</sup> Povrozniuk N. (2020) Digital History of Virtual Museums: The Transition from Analog to Internet Environment. In: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference

мании, России и Австралии были построены модели виртуального музея как специфического информационного ресурса. Модели отражают развитие на музейных сайтах цифрового контента разных типов, обеспечивающего возможность удаленного использования электронными ресурсами музея и придающего сайтам стиль виртуального музея<sup>224</sup>, а также выделение аудиторий, для которых был разработан тот или иной контент. Модели были сформированы с учетом их динамики, проявившейся в изменении функционала и появлении тех или иных элементов с течением времени<sup>225</sup>. Анализ моделей позволил выйти на формулировку концепции цифровой идентичности музеев.

Кроме самостоятельного значения для анализа отдельных аспектов изучаемого предмета, моделирование данных открывает дальнейшие перспективы для продолжения исследования. Например, сформированные модели данных могут служить основой для выработки концептуальной и логической моделей с целью их последующей реализации при создании баз данных или информационных систем и формирования датасетов для проведения количественного анализа, например.

## 6.2. Методы количественного анализа

Сохранение больших объемов данных в веб-архивах открывает широкие возможности для применения количественных методов исследования, построения различных математических моделей и статистического анализа. Количественные методы позволяют проводить целый ряд манипуляций с данными, в том числе, выявлять тренды и динамику процессов и явлений. Именно количественные методы чаще всего служат основой для анализа больших объемов данных.

Ранее уже обращалось внимание на масштабирование проектов – микро- и макроуровни исследований. Соответственно, выбор тех или иных

---

Riga, Latvia, October 21-23, 2020. Pp. 125–136. URL: <http://ceur-ws.org/Vol-2612/paper9.pdf> (дата обращения: 01.05.2022).

<sup>224</sup> Виртуальный музей понимается как «информационная система, содержащая концептуально единую электронную коллекцию или совокупность коллекций предметов (экспонатов) с метаданными, имеющая характеристики музея и позволяющая осуществлять научную, просветительскую, экспозиционную и экскурсионную деятельность в виртуальном пространстве». См. Поврозник Н.Г. Виртуальный музей: сохранение и репрезентация историко-культурного наследия // Вестник Пермского университета. Сер. «История». 2015. № 4(31). С. 214.

<sup>225</sup> См. раздел Digital Initiatives на странице Infographics. Digital History of Virtual Museums. URL: <http://virtualmuseumhistory.com/infographics/> (дата обращения: 01.05.2022).

количественных методов, а также уровня их сложности зависит от размеров наборов данных, способов их организации, форматов данных и т.д.

Веб-сайт представляет собой комплексный по своему характеру ресурс, в котором могут быть объединены разные форматы представления данных. Поэтому возможности обработки массивов могут быть связаны с выделением отдельных форматов для изучения – например, построение текстовых наборов данных (как небольшого размера, так и создание корпусов разного типа и тематики), выделение массивов изображений, извлечение ссылочной массы и т.д. Соответственно, каждый из форматов предполагает те или иные методы обработки и анализа данных, в том числе, например, методы компьютерной лингвистики предусматривают целый спектр возможностей для изучения текстов. Применение методов количественного анализа позволяет выявлять содержание текстов, определять основные тренды в них и демонстрировать динамику, взаимосвязи отдельных маркеров, определять эмоциональную окраску и реализовывать другие возможности, которые широко используются для анализа исторических текстов.

Методы количественного анализа могут быть применены к разным информационным слоям архивного веба: контента страниц (в различных форматах), совокупностей страниц в структуре домена и их взаимосвязей с внешними сайтами, а также в отношении метаданных. Особенность веб-архива как информационного ресурса заключается в том, что он может рассматриваться как единый крупномасштабный корпус документов, к которому могут быть применены средства анализа больших объемов данных – в том числе, интеллектуальный (глубинный) анализ данных (text mining и data mining), а также разнообразные возможности анализа и визуализации (в т.ч. построение сетей).

Количественные методы существенно расширяют аналитические возможности, но существуют и принципиальные ограничения в их использовании по отношению к данным веб-архивов. С одной стороны, у архивированных веб-материалов есть очевидное преимущество в виде изначально электронного формата существования, что облегчает извлечение и обработку данных. Но с другой стороны, построение датасетов и формирование корпусов для анализа часто требуют решения существенных проблем. Среди них, помимо обозначенных в Гл. 3 данного пособия сложностей общего характера, относящихся к функционированию веб-



архивов, можно выделить следующие трудности, характерные для самих данных:

1. **неполнота данных**, которая зачастую препятствует полноценной реконструкции целых пластов истории ввиду того, что информация с сайтов могла быть не сохранена в веб-архиве в нужном объеме, а в отдельные периоды может вообще отсутствовать. С данной проблемой, например, столкнулись исследователи при реконструкции истории сайта университета Болоньи, когда существенная часть сведений не была сохранена в Интернет-архиве, а также сайт не был полностью архивирован как часть национального домена из-за нерегулярности предпринимаемых действий в рамках итальянской программы по сохранению веба<sup>226</sup>;

2. **гибкость структуры и содержания источников и динамизм ресурсов**, проявляющиеся в частой трансформации структуры сайтов и постоянного изменения, дополнения, редактирования их контента. Эта проблема оказывает прямое воздействие на оценку выборки при построении наборов данных, и определение полноты итогового датасета по сравнению со сведениями первичного источника. Вариативность структуры и содержания существенно осложняет возможности применения источниково-ориентированного подхода во всей его полноте<sup>227</sup>;

3. **дублирование информации**, которое неизбежно при попадании сайта в веб-архив, особенно при регулярном его сохранении, что сказывается на формировании наборов данных, в связи с чем необходимо очищать датасеты от повторяющихся сведений. В частности, эта проблема связана с организацией веб-архивов и хранением снапшотов сайтов, когда сайт или отдельные его разделы сохраняются полностью и затем дублируются в последующих версиях заархивированного контента. Кроме того, у исследователя отсутствует возможность отследить приращение информации на сайте или ее исчезновение, и требуются дополнительные манипуляции для выявления этой информации и подготовки итоговых наборов данных для изучения;

---

<sup>226</sup> Brügger N., Milligan I., Ben-David A., Gebeil S., Nanni F., Rogers R., Turkel W. J., Weber M. S., Webster P. Internet histories and computational methods: a “round-doc” discussion // *Internet Histories*, 2019, 3:3-4, p. 210. DOI: 10.1080/24701475.2019.1639352

<sup>227</sup> Гагарина Д. А., Корниенко С. И., Поврозник Н. Г. Информационные системы в цифровой среде исторической науки // *Электронный научно-образовательный журнал «История»*. 2016. Т. 7, вып. 7 (51) [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001638-0-1>. DOI: 10.18254/S0001638-0-1 (дата обращения: 01.05.2022)

4. сложность в организации наборов данных, взятых из разных веб-архивов и репозиториях, ввиду отличающихся **метаданных**. Исследователи неоднократно указывали на различия в структуре метаданных, используемых в веб-архивах, и необходимость стандартизации метаданных, чтобы расширить возможности создания обобщенных датасетов на основе материалов разных веб-архивов<sup>228</sup>. Вариативность метаданных, различная степень их совместимости препятствуют качественному объединению наборов данных в один датасет для последующего анализа, требует перестроения структуры метаданных, и, в конечном итоге, приводит к невозможности использовать большую часть сведений, которую сложно организовать по неким общим правилам;

5. необходимость **поиска подходящих методов и инструментария** для организации и анализа больших данных. Сложности работы с большими данными<sup>229</sup> относятся к более широкой проблеме – изобилию цифровых данных, которое в рамках веб-истории приобретает колоссальные масштабы<sup>230</sup>.

В качестве примера такого исторического исследования, выполненного на основе сбора и анализа материалов веб-архивов с применением количественных методов, можно привести проект, посвященный дискуссиям об абортах в Австралии и их изменению, уже упомянутый ранее.<sup>231</sup> На основе методов математической статистики ученые изучали видимость и посещаемость сайтов, на которых были опубликованы материалы в пользу или против поддержки жизни, рассматривали изменение взаимосвязей сайтов, содержащих противоположные позиции, с помощью социально-сетевого анализа, а также выявляли свидетельства о том, насколько сторонники тех или иных подходов апеллировали к аргументации друг друга с использованием методов анализа массива гиперссылок. Количественный и качественный анализ выделенных с сайтов дискуссионных текстов позволил авторам ответить на вопросы об изменениях артикуля-

---

<sup>228</sup> Brügger N., Milligan I., Ben-David A., Gebeil S., Nanni F., Rogers R., Turkel W. J., Weber M. S., Webster P. Internet histories and computational methods: a “round-doc” discussion // *Internet Histories*, 2019, 3:3-4, p. 211. DOI: 10.1080/24701475.2019.1639352

<sup>229</sup> Tufekci Z. Big Data: Pitfalls, Methods and Concepts for an Emergent Field // SSRN. March 7, 2013. URL: <https://ssrn.com/abstract=2229952> (дата обращения: 01.05.2022)

<sup>230</sup> Milligan I. *History in the Age of Abundance? How the Web is Transforming Historical Research*. McGill-Queen’s University Press, 2019.

<sup>231</sup> Robert Ackland, Ann Evans (2017) Using the web to examine the evolution of the abortion debate in Australia, 2005–2015. In: Niels Brügger, Ralph Schroeder (Eds.) *The Web as History. Using Web Archives to Understand the Past and the Present*.

ции проблемного поля с течением времени. Важным результатом работы стали выводы об эволюции способов коммуникации и выборе пользователями тех или иных каналов, в том числе социальных медиа.

### 6.3. Сетевой анализ данных

Сетевой анализ также широко применяется в исторических исследованиях<sup>232</sup>, а в области веб-истории открываются дополнительные возможности ввиду особенностей веб-ресурсов, взаимосвязи которых уже реализованы на техническом уровне. Так, на основе сетевого анализа может быть проанализирована ссылочная масса того или иного ресурса или комплекса ресурсов и выявлены связи с внешними сайтами.

Важным подходом к выявлению и изучению сетей во Всемирной паутине является анализ сегментов веба по доменной принадлежности и изучение внешних взаимосвязей разных национальных доменов. Интересным исследованием в этом направлении является проект, посвященный анализу югославского национального домена .yu, реализованный с использованием методов сетевого моделирования<sup>233</sup>. Угроза утраты цифрового наследия и данных сайтов национального югославского домена назрела к 2008 г., когда было объявлено о необходимости перевода ресурсов в другой домен и их перерегистрации на новом (сербском) домене ввиду разделения страны. В ходе этого проекта по сохранению и анализу югославского веба было отсканировано 17 460<sup>234</sup> веб-сайтов, большинство из которых оказались уникальными, то есть не были заархивированы в других веб-архивах (был проведен сравнительный анализ наличия снапшотов сайтов в Интернет-архиве). В результате сетевого анализа выделенной гиперссылочной массы удалось реконструировать динамику в развитии сайтов национального югославского домена по годам и продемонстрировать влияние политических факторов на эти процессы, связанные с режимом С. Милошевича и военными действиями в Косово. Кроме того,

---

<sup>232</sup> Бородкин Л. И. Сетевой анализ в исторических исследованиях: микро- макроподходы // Историческая информатика. 2017. № 1. С. 110–124. DOI: 10.7256/2306-0891.2017.1.22842 URL: [https://nbpublish.com/library\\_read\\_article.php?id=22842](https://nbpublish.com/library_read_article.php?id=22842) (дата обращения: 01.05.2022).

<sup>233</sup> Ben-David A. (2016) What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain // *New Media & Society*. Vol. 18(7). Pp. 1103–1119. DOI: 10.1177/1461444816643790 (дата обращения: 01.05.2022).

<sup>234</sup> Там же. Р. 1110.

на основе визуализации данных были показаны внешние связи домена .uu с другими доменами.

На основе методов построения и анализа сетей могут быть изучены и особенности развития социума. В частности, исследовательский вопрос может относиться к изучению персональных сетей и выявлению личных предпочтений индивидов, в том числе касающихся потребления товаров и услуг, определению особенностей проведения досуга, осуществления покупок, выбора материалов для чтения и т.д. Другим направлением применения сетевых методов является анализ социально-сетевых структур, общественных движений, виртуальных сообществ на основе общности интересов, критически важным средством обеспечения и поддержания деятельности которых является медиакommunikация.

Создание наборов данных для проведения социально-сетевого анализа активно используется на примерах «живого» веба, а некоторые социальные медиа имеют инструментарий для извлечения данных и их визуализации<sup>235</sup>. Обращение исключительно к материалам веб-архивов существенно ограничивает исследовательские возможности, поскольку многие ресурсы исчезают, не попадая в архив. Это происходит по разным причинам: недостаточности предпринимаемых мер по сохранению материалов социальных сетей как цифрового наследия, архивирования преимущественно открытой части веба и т.д. Поэтому изучение веб-истории общества и использование методов социально-сетевого моделирования чаще реализуется на уровне организаций, общественных объединений и сообществ, информация о которых есть в открытом вебе и доступна для размещения в веб-архивах.

Выделение гиперссылок в тех или иных наборах данных и изучение их взаимосвязей позволяет построить визуализации сетей, определить силу связей на основе частоты появления гиперссылки, связанной с определенным веб-ресурсом (страницей). Ключевым аспектом может являться определение центра и периферии по отношению к изучаемому объекту, визуализации данных связей позволит их интерпретировать. Анализ коэффициентов центральности, измерение силы связи узлов, определение плотности сети и дистанций между узлами также могут дать информацию о позиционировании рассматриваемых элементов сети и их роль.

---

<sup>235</sup> Батура Т. В. Модели и методы анализа компьютерных социальных сетей // Программные продукты и системы. 2013. № 3. URL: <https://cyberleninka.ru/article/n/modeli-i-metody-analiza-kompyuternyh-sotsialnyh-setey> (дата обращения: 01.05.2022)

М. Стевенсон и А. Бен-Давид выделяют следующие основные направления применения методов сетевого анализа по отношению в данным веб-архивов в исторических исследованиях<sup>236</sup>:

- 1) изучение эволюции сетей во времени с фокусом на изменение значения содержания ресурсов;
- 2) анализ развития веба как медиума, где ключевое значение имеет технологический прогресс, выявление изменений в структуре и свойствах веб-ресурсов;
- 3) источниковедческий анализ веба как первичного исторического источника.

Пристальное изучение веба при условии достаточности источниковой базы способно выявить дополнительные слои информации, визуализировать их, дополнить аргументацию и способствовать появлению новых аргументов и знаний при интерпретации полученных данных. К таким слоям относятся данные о взаимосвязях между сайтами, классификация ссылочной массы, ее изменение и т.д.

Создание сетей и их визуализация при изучении дискуссий способны выявить и продемонстрировать взаимосвязи на уровне идей, позволяют моделировать движение потоков информации, что становится возможным при указании ссылки на другого автора, сторонника или оппонента.

Проиллюстрировать применение методов сетевого анализа можно на примере исследований, посвященных изучению новостей. Так, в статье М.С. Вебера на основе построения сетей сайтов показана эволюция новостного пространства в США<sup>237</sup>. Данные для анализа были получены на основе материалов Интернет-архива и проекта ArchiveHub. Из созданного датасета была извлечена гиперссылочная масса, визуализация взаимосвязей которой позволила продемонстрировать потоки информации между центральными крупными новостными платформами и местными новостными изданиями. На примере отдельных сайтов было показано заимствование визуального материала (фотографий и других иллюстраций) более мелкими платформами у более крупных. Для периодизации новостной истории исследователи применили методы сетевого анализа. Были постро-

---

<sup>236</sup> Stevenson M., Ben-David A. Network Analysis for Web History. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. 2019. Pp. 125–138.

<sup>237</sup> Weber M. S. The tumultuous history of news on the web. In: The Web as History. Using Web Archives to Understand the Past and the Present. Eds. Niels Brügger, Ralph Schroeder. London, UCL Press, 2017. Pp. 83–100. DOI: <https://doi.org/10.14324/111.9781911307563>.

ны визуализации на основе данных за 1999 и 2005 гг. в виде сетевых графов с демонстрацией взаимосвязей новостных сайтов с другими сайтами и проанализирована плотность этих связей, а также динамика количества узлов сети. Результаты сетевого анализа показали существенное увеличение плотности сети, значительный рост количества новостных сайтов в целом и их взаимосвязей. Анализируя связи федеральных и местных новостных платформ, авторы пришли к выводу о постепенном «выдавливании» мелких игроков с рынка и кризисе местных новостных компаний.

#### **6.4. Методы качественного анализа**

Методы качественного анализа основаны на выявлении и изучении значимых характеристик изучаемого объекта для дальнейшей их интерпретации. При этом особую ценность имеет уже само наличие данных характеристик, которые могут указать на уникальные черты объекта, находящегося в фокусе исследования. Одним из методов качественного анализа выступает контент-анализ, который может сочетать качественно-количественные элементы.

Метод контент-анализа уже давно используется в исторических исследованиях. И. М. Гарскова указывает на тот факт, что контент-анализ является не просто методом исследования, а может выступать системой междисциплинарных методов<sup>238</sup>.

Контент-анализ основан на изучении формализованных структур текстов. Такими структурами могут служить отдельные части слов, слова и словоформы, комбинации слов, смысловые категории и т.д. Достоинством метода является то, что он способствует выявлению скрытого содержания текста, обнаружение которого традиционным путем, методами ближнего чтения, весьма затруднительно или часто не представляется возможным.

Для проведения контент-анализа необходимо собрать массив данных, для чего требуется особый инструментарий. Корпус текстов на основе веб-архивных материалов может быть сформирован, например, с помощью сервиса Sketch Engine<sup>239</sup>. Данный сервис представляет различные

---

<sup>238</sup> Гарскова И. М. Новые тенденции в компьютеризованном анализе текстов: концепции, методы, технологии // Электронный научно-образовательный журнал «История». 2015. Т. 6, вып. 8 (41). URL: <https://history.jes.su/s207987840001255-9-1/> (дата обращения: 01.05.2022).

<sup>239</sup> Sketch Engine. URL: <https://www.sketchengine.eu/guide/create-a-corpus-from-the-web/> (дата обращения: 01.05.2022).

способы формирования наборов данных, в частности, позволяет в автоматическом режиме загружать указанные пользователем веб-страницы (посредством списка URL-адресов), подбирать для загрузки страницы согласно списку ключевых слов (т.н. seed words), либо скачивать сайты целиком. При этом сервис очищает загружаемые данные от иных, нетекстовых форматов, формируя корпус текстов, пригодный для дальнейшей обработки.

По отношению к веб-архивным материалам контент-анализ может применяться как для изучения контента веб-страниц и сайтов в целом, так и для исследования метаданных, текстов ссылочной массы и даже программного кода.

В Гл. 5 данного учебного пособия при рассмотрении особенностей формирования исследовательских датасетов уже упоминался проект, посвященный анализу дискуссий об абортах, разворачивавшихся в австралийском онлайн-пространстве<sup>240</sup>. Сетевой анализ гиперссылок и контент-анализ текстов были выбраны в качестве центральных исследовательских методов этого проекта. Предварительно сайты были разделены на категории, отражающие основную позицию по отношению к абортам (за сохранение жизни, за возможность выбора, нейтральное и некоторые другие), и по типам ресурсов (такие как, правительственные сайты, блоги, сайты религиозных организаций, академических учреждений и т.д.). Создание сетей сайтов позволило визуализировать ресурсы, отражающие разные точки зрения на обсуждаемые проблемы, и апеллирование к альтернативным взглядам. Контент-анализ дополнил результаты сетевого анализа и значительно углубил взгляд на содержание дискуссий. Проведение исследования потребовало формирования двух текстовых корпусов – один представил метаданные (ключевые слова, заголовки и аннотации), второй был сформирован непосредственно на базе основного содержания веб-страниц. Контент-анализ проводился на основе анализа отдельных слов, а словосочетания практически не использовались. На основании двух корпусов были выявлены основные маркеры, которые затем сравнивались для разных типов ресурсов, изучалась риторика высказывания разных групп, были выявлены и визуализированы различия в языке двух основных «лагерей» в рамках общественной дискуссии. Контент-анализ на разных

---

<sup>240</sup> Ackland R., Evans A. Using the web to examine the evolution of the abortion debate in Australia, 2005–2015. In: Brügger N., Schroeder R. (Eds.) *The Web as History. Using Web Archives to Understand the Past and the Present*. 2017. Pp. 159–189.

уровнях – метаданных и основного содержания – позволил углубить представление о значимости категорий и их месте в обсуждениях.

Проект по изучению процессов зарождения и развития блогосферы в США может служить другим ярким примером применения методов качественного анализа в области веб-истории<sup>241</sup>, а тема проекта является особенно значимой для понимания истории современного общества и развития медиакommunikации. Автор проекта использует разные методы качественного анализа, в том числе, интервью с разными акторами, включенными в блогосферу, включая создателей контента, пользователей, веб-разработчиков и т.д., а также содержательный анализ блогов, сохраненных в веб-архивах.

Следует также обратить внимание на возможности применения качественных методов исследования для проведения источниковедческого анализа веб-ресурсов. Ранее уже рассматривались характерные черты источников, изначально созданных в цифровом формате, и особенности их источниковедческой критики (см. Гл. 5). Качественные методы позволяют выявлять изменения, которые были привнесены на сайт, и изучать ресурсы в сравнительной перспективе. Это бывает особенно важно в тех ситуациях, когда сайты, их содержание и даже программный код модифицировались без авторизованного доступа к ресурсу. Такие случаи связаны с хакерскими атаками и могут рассматриваться не только в контексте киберпреступности, но и с точки зрения необходимости сохранения следов такого несанкционированного доступа к веб-архивам и их документирования. Этим проблемам посвящен исследовательский проект М. Курцмаера<sup>242</sup>, в котором особое внимание уделяется деятельности хакеров по преобразованию структуры и содержания веб-ресурсов и возможностям фиксации этих изменений в веб-архивах.

---

<sup>241</sup> Siles I. Blogs. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. 2019. Pp. 359–371.

<sup>242</sup> Kurzmeier M. (2020) Towards a concept for archiving hacked websites // NPPSH Reflections, 2020, Vol. 3. Pp. 35–58. URL: <http://mural.maynoothuniversity.ie/13352/> (дата обращения: 01.05.2022)



## 6.5. Визуализация данных

Визуализация является способом представления организованных данных в виде графиков, графов, диаграмм, карт и различных схем<sup>243</sup>. Этот процесс имеет существенное значение при проведении исследования для интерпретации результатов как возможность увидеть тренды и тенденции, что особенно важно в отношении больших объемов данных. Подавляющее большинство методов исследования предполагают необходимость построения визуализации для выявления значимой информации.

Визуализация данных выполняет следующие основные функции:

- 1) мониторинг данных с целью отследить отраженные в наборе данных процессы;
- 2) оценка трендов и выявление тенденций и закономерностей (анализ данных);
- 3) прояснение текущей ситуации, представление структуры и отдельных фрагментов данных для понимания содержания датасетов и получения представления о дальнейшем направлении в работе с данными;
- 4) представление данных как демонстрация полученных результатов.

Визуализации могут быть двух-, трех- или многомерными, последние из которых требуют упрощения для облегчения восприятия человеком. Исторические данные, сформированные на материалах веб-архивов, также могут быть представлены в виде разных вариантов визуализаций<sup>244</sup>.

Ряд веб-архивов предлагает инфраструктурные средства анализа и визуализации накопленной информации, а также специальный инструментарий для изучения массивов данных. Например, в структуре британского веб-архива UKWA была реализована возможность построения N-gram по массиву данных, находящихся в открытом контуре веб-архива. Инструмент N-gram Search<sup>245</sup> был создан для выявления тех или иных терминов в содержании заархивированных веб-материалов и их визуализации помещенно.

---

<sup>243</sup> См., например, Knight A. (2019) Where Data Meets Design: Visualization in the Digital Humanities. In: Doing More Digital Humanities. Open Approaches to Creation, Growth, and Development. Eds.: Crompton C., Lane R.J., Siemens R. London. Routledge. Pp. 141–155.

<sup>244</sup> Joque J. Visualizing Historical Web Data. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. 2019. Pp. 168–185.

<sup>245</sup> N-gram Search. UKWA. Internet Archive. Снапшот от 01.09.2014. URL: <https://web.archive.org/web/20140901210533/http://www.webarchive.org.uk/ukwa/ngram/> (дата обращения: 01.05.2022).

Для изучения контента нидерландского веб-архива был разработан инструмент WebArt<sup>246</sup>, позволяющий выявлять и визуализировать данные контекстного поиска по отдельным терминам. Практическое использование данного инструмента было показано на коллекции новостных документов веб-архива при выявлении динамики упоминания глав ряда государств на протяжении нескольких лет<sup>247</sup>.

В последнее время начинают появляться средства визуализации веб-архивного контента, не относящиеся напрямую к инфраструктуре веб-архивов. Так, при поддержке ИПС разрабатывается проект LinkGate, в рамках которого создаются специализированные программные средства для визуализации и документирования ресурсов веб-архивов – link-serv, link-indexer, and link-viz<sup>248</sup>. Разработанные инструменты позволяют визуализировать ресурсы веб-архивов в виде графа с узлами и исходящими связями. По результатам проекта разработанные средства публикуются в открытом репозитории на GitHub<sup>249</sup> и доступны для использования.

### Рекомендуемая литература

1. Батура Т. В. Модели и методы анализа компьютерных социальных сетей // Программные продукты и системы. 2013. № 3. URL: <https://cyberleninka.ru/article/n/modeli-i-metody-analiza-kompyuternyh-sotsialnyh-setey> (дата обращения: 01.05.2022)

2. Бородин Л. И. Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив. СПб.: Алетей, 2017. 303 с.

3. Бородин Л. И. Сетевой анализ в исторических исследованиях: микро-макроподходы // Историческая информатика. 2017. № 1. С. 110–124. DOI: 10.7256/2306-0891.2017.1.22842. URL: [https://nbpublish.com/library\\_read\\_article.php?id=22842](https://nbpublish.com/library_read_article.php?id=22842) (дата обращения: 01.05.2022).

4. Володин А. Ю. «Цифровая история»: ремесло историка в цифровую эпоху // Электронный научно-образовательный журнал «История». 2015, вып. 6

---

<sup>246</sup> WebArt: Web Archive Retrieval Tools. URL: <http://www.webarchiving.nl> (дата обращения: 01.05.2022).

<sup>247</sup> Searching the Archives. Digital Methods Initiative Winterschool, 22–25 January 2013. Digital Methods Initiative. URL: <https://wiki.digitalmethods.net/Dmi/Winter13SearchingTheArchive> (дата обращения: 01.05.2022).

<sup>248</sup> LinkGate: Core Functionality and Future Use Cases. ИПС. URL: <https://netpreserve.org/projects/linkgate> (дата обращения: 01.05.2022).

<sup>249</sup> BA-WA. bibalex.org – Web Archiving. GitHub. URL: <https://github.com/arcalex> (дата обращения: 01.05.2022).

(8). [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <http://history.jes.su/s207987840001228-9-1> (дата обращения: 01.05.2022).

5. Гагарина Д. А., Корниенко С. И., Поврозник Н. Г. Информационные системы в цифровой среде исторической науки // Электронный научно-образовательный журнал «История». 2016. Т. 7, вып. 7 (51) [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <https://history.jes.su/s207987840001638-0-1> (дата обращения: 01.05.2022).

6. Гарскова И. М. Новые тенденции в компьютеризованном анализе текстов: концепции, методы, технологии // Электронный научно-образовательный журнал «История». 2015. Т. 6, вып. 8 (41). [Электронный ресурс]. Доступ для зарегистрированных пользователей. URL: <https://history.jes.su/s207987840001255-9-1> (дата обращения: 01.05.2022).

7. Кирьянов И. К., Корниенко С. И., Сенина А. В. Депутатский корпус Государственной Думы позднеимперской России: от текста источника к статистическим компьютеризованным методам анализа // Историческая информатика. 2019. № 1. С. 118–127. DOI: 10.7256/2585-7797.2019.1.29148. URL: [https://nbpublish.com/library\\_read\\_article.php?id=29148](https://nbpublish.com/library_read_article.php?id=29148) (дата обращения: 01.05.2022).

8. Ковальченко И. Д. Методы исторического исследования. Отделение историко-филологических наук. 2-е изд., доп. М.: Наука, 2003. 486 с. URL: [http://www.hist.msu.ru/departments/8819/research/publications/detail.php?ELEMENT\\_ID=43145](http://www.hist.msu.ru/departments/8819/research/publications/detail.php?ELEMENT_ID=43145) (дата обращения: 01.05.2022).

9. Моретти Ф. Дальнее чтение / Франко Моретти; пер. с англ. Алексея Вдовина [и др.] / под науч. ред. Инны Кушнareвой. М.: Изд-во Института Гайдара, 2016. 342 с.

10. Ackland R., Evans A. Using the web to examine the evolution of the abortion debate in Australia, 2005–2015. In: Brügger N., Schroeder R. (Eds.) The Web as History. Using Web Archives to Understand the Past and the Present. 2017. Pp. 159–189.

11. Ben-David A. (2016) What does the Web remember of its deleted past? An archival reconstruction of the former Yugoslav top-level domain // New Media & Society. Vol. 18(7). Pp. 1103–1119. DOI: 10.1177/1461444816643790 (дата обращения: 01.05.2022).

12. Brügger N., Milligan I., Ben-David A., Gebeil S., Nanni F., Rogers R., Turkel W. J., Weber M. S., Webster P. Internet histories and computational methods: a “round-doc” discussion // Internet Histories, 2019, 3:3-4, p. 210. DOI: 10.1080/24701475.2019.1639352

13. Ciula A., Eide Ø. (2017) Modelling in digital humanities: Signs in context. In: Digital Scholarship in the Humanities, Volume 32, Issue suppl\_1. Pp. i33–i46.

14. Jänicke S., Franzini G., Cheema M. F., Scheuermann G. (2015) On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges // The proceedings of the Eurographics Conference on Visualization (EuroVis). Eds.: R. Borgo, F. Ganovelli, and I. Viola. 2015. URL: <https://www.informatik.uni-leipzig.de/~stjaenicke/Survey.pdf> (дата обращения: 01.05.2022).
15. Joque J. Visualizing Historical Web Data. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. 2019. Pp. 168–185.
16. Knight A. (2019) Where Data Meets Design: Visualization in the Digital Humanities. In: DOIng More Digital Humanities. Open Approaches to Creation, Growth, and Development. Eds.: Crompton C., Lane R. J., Siemens R. London. Routledge. Pp. 141–155.
17. Kurzmeier M. (2020) Towards a concept for archiving hacked websites // NPPSH Reflections, 2020, Vol. 3. Pp. 35–58. URL: <http://mural.maynoothuniversity.ie/13352/> (дата обращения: 01.05.2022)
18. Meyer E. T., Thomas A., Schroeder R. Web Archives: The Future(s). June 30, 2011. URL: <https://ssrn.com/abstract=1830025> (дата обращения: 01.05.2022).
19. Milligan I. History in the Age of Abundance? How the Web is Transforming Historical Research. McGill-Queen's University Press, 2019.
20. Povroznik N. (2020) Digital History of Virtual Museums: The Transition from Analog to Internet Environment. In: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference Riga, Latvia, October 21-23, 2020. Pp. 125–136. URL: <http://ceur-ws.org/Vol-2612/paper9.pdf> (дата обращения: 01.05.2022).
21. Siles I. Blogs. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. 2019. Pp. 359–371.
22. Stevenson M., Ben-David A. Network Analysis for Web History. In: The SAGE Handbook of Web History. Eds. N. Brugger, I. Milligan. 2019. Pp. 125–138.
23. Tufekci Z. Big Data: Pitfalls, Methods and Concepts for an Emergent Field // SSRN. March 7, 2013. URL: <https://ssrn.com/abstract=2229952> (дата обращения: 01.05.2022)
24. Weber M. S. The tumultuous history of news on the web. In: The Web as History. Using Web Archives to Understand the Past and the Present. Eds. Niels Brügger, Ralph Schroeder. London, UCL Press, 2017. Pp. 83–100. DOI: <https://doi.org/10.14324/111.9781911307563>.

## **Источники**

1. Archives Unleashed Project. URL: <https://archivesunleashed.org> (дата обращения: 01.05.2022).
2. BA-WA. bibalex.org – Web Archiving. GitHub. URL: <https://github.com/arcalex> (дата обращения: 01.05.2022).
3. Digital History of Virtual Museums. URL: <http://virtualmuseumhistory.com/> (дата обращения: 01.05.2022).
4. LinkGate: Core Functionality and Future Use Cases. IIPC. URL: <https://netpreserve.org/projects/linkgate> (дата обращения: 01.05.2022).
5. N-gram Search. UKWA. Internet Archive. Снапшот от 01.09.2014. URL: <https://web.archive.org/web/20140901210533/http://www.webarchive.org.uk/ukwa/ngram/> (дата обращения: 01.05.2022).
6. Searching the Archives. Digital Methods Initiative Winterschool, 22-25 January 2013. Digital Methods Initiative. URL: <https://wiki.digitalmethods.net/Dmi/Winter13SearchingTheArchive> (дата обращения: 01.05.2022).
7. Sketch Engine. URL: <https://www.sketchengine.eu/guide/create-a-corpus-from-the-web/> (дата обращения: 01.05.2022).
8. WebArt: Web Archive Retrieval Tools. URL: <http://www.webarchiving.nl> (дата обращения: 01.05.2022).

## **Вопросы для самопроверки**

1. Сформулируйте отличия подходов ближнего и дальнего чтения.
2. Назовите ученого, который является теоретиком подходов дальнего чтения.
3. Приведите пример крупного проекта, в котором использован подход дальнего чтения по отношению к веб-архивам. В чем его специфика?
4. В чем заключается суть метода моделирования данных?
5. Какие структурные единицы веба могут служить основой для моделирования данных заархивированного веба?
6. На основе каких характеристик музеев были построены модели виртуальных музеев в проекте “Digital History of Virtual Museums”?
7. В чем заключается преимущество методов моделирования для продолжения исследования в плане разработки информационных ресурсов?
8. Перечислите методы количественного анализа, которые вам известны.
9. Сформулируйте особенности масштабирования проектов в области веб-истории.

10. Какие форматы данных могут быть проанализированы на основе количественных методов? Приведите примеры.
11. В чем заключаются возможности применения количественных методов? Для чего они используются?
12. К каким форматам данных из материалов веб-архивов могут быть применены количественные методы?
13. Перечислите, какие существуют ограничения в применении количественных методов, которые необходимо принимать во внимание, работая с данными веб-архивов?
14. Приведите примеры использования количественных методов в исследовательских проектах по веб-истории?
15. В чем заключается особенность применения методов сетевого анализа по отношению к архивным веб-ресурсам?
16. Приведите пример исследовательского проекта, основанного на изучении связей национального веб-домена с другими доменами. Почему он важен как проект в области веб-истории?
17. Какие существуют возможности для изучения социальных связей на основе методов сетевого анализа?
18. Какое значение имеет визуализация сетей для исследования?
19. Перечислите основные направления применения методов сетевого анализа данных веб-архивов?
20. Приведите пример использования сетевого анализа в исследовательском проекте в области веб-истории. Какое значение имел этот метод для исследования?
21. Какие методы качественного анализа вы знаете?
22. Для каких целей применяются методы качественного анализа? В чем заключается их смысл?
23. В чем заключается процедура контент-анализа?
24. Сформулируйте преимущество контент-анализа для изучения текстов?
25. Какие существуют сервисы для формирования корпусов на основе веб-ресурсов? Опишите принципы работы сервиса SketchEngine?
26. Для изучения каких слоев информации веб-ресурсов может быть применен контент-анализ?
27. Приведите примеры исследовательских проектов в области веб-истории, выполненных с применением методов контент-анализа. Какие задачи контент-анализ был призван решить?
28. Каким образом методы качественного анализа могут быть применены в источниковедческих исследованиях веба?

29. Какие функции выполняет визуализация данных в исследованиях по веб-истории?

30. Приведите примеры инфраструктурных решений веб-архивов для визуализации данных.

31. Какие существуют программные инструменты визуализации данных веб-архивов? В чем заключается принцип их работы?

## ЗАКЛЮЧЕНИЕ

### Рекомендации преподавателю

Курс по «Веб-история социума и его институтов» разработан для образовательных магистерских программ в области истории и направлен на подготовку специалистов-историков и других гуманитариев, использующих ресурсы веб-архивов в профессиональной деятельности.

Знакомство с курсом рекомендуется начинать с изучения основ веб-архивирования и устройства веб-архивов. При рассмотрении ресурсов веб-архивов в образовательном процессе студент должен научиться выявлять их информационный потенциал для проведения исторического исследования и иметь представление о том, что доступные в онлайн-пространстве веб-архивы формируют достаточно широкое источниковое поле для проектов разного масштаба. Умение критически относиться к ресурсам веб-архивов является важным навыком, поскольку необходимо знать ограничения веб-архивов, их влияние на сохранение ресурсов и формирование выборки. В результате освоения курса студент должен уметь применять эти знания на практике при реализации исторических исследований.

Рассмотрение возможностей разных веб-архивов в ходе практических занятий может быть осуществлено на основе сравнения функционала и содержания открытых ресурсов, опробовано с помощью заданий на поиск контента по определенной теме и других операций. Рекомендованные ресурсы (см. списки литературы и источников к главам, а также Прил. 2) могут быть встроены в структуру образовательного курса, позволят студентам лучше ориентироваться в разных веб-архивах, понимать их специфику и эффективно реализовывать поисковые запросы к их контенту. Одним из заданий для практикума может быть работа типа «детективная история», которая предполагает поиск в веб-архивах и анализ информации, удаленной из «живого» веба. Такие истории могут относиться, например, к изменениям политического курса в связи с переизбранием президента в той или иной стране и последующей корректировкой веб-контента (см., например, Гл. 3 данного пособия о действиях Д. Трампа на посту президента США), либо преднамеренным искажением и подменой информации на веб-сайтах. Детективные истории, предложенные преподавателем, позволят обратить внимание студентов на манипулирование



данными, цензуру, предвзятости разного рода и необходимость источниковедческой критики цифровых источников, которые обсуждаются на занятиях.

Современная инфраструктура веб-архивов постепенно меняется, сотрудничество специалистов из разных отраслей – архивистов, библиотекарей, цифровых гуманитариев и других экспертов – приводит к совершенствованию способов сбора, хранения, поиска материалов в веб-архивах, что, несомненно, положительно сказывается на возможностях применения этих ресурсов в исследованиях и образовании. Созданные и создаваемые сервисы последовательно формируют фундамент для более глубокого использования материалов веб-архивов менее подготовленными в технологическом плане пользователями, какими зачастую являются студенты-гуманитарии. Естественным образом современные требования к цифровой грамотности гуманитариев меняются: дисциплины, связанные с прикладной информатикой, программированием и более глубоким использованием количественных методов вводятся в программы подготовки для повышения профессиональной цифровой грамотности выпускников гуманитарных специальностей. Однако этот тренд еще не стал ведущим в глобальном масштабе. Поэтому в рамках курса для освоения основ работы с веб-архивами и проведения исследований в области веб-истории преподавателем должен быть предложен широкий круг инструментов от самых базовых, не требующих специальных знаний и навыков использования сложных методов и программирования, до более продвинутых, чтобы расширить горизонты представлений о возможностях технологий и ресурсов в обработке данных веб-архивов, их визуализации и анализа.

Курс разработан как проектно-ориентированный, чтобы студенты обучались на применении веб-архивов на практике, что предполагает в итоге освоения курса реализацию индивидуального проекта, исследования, выполненного студентом самостоятельно.

Первые попытки использования веб-архивных материалов должны быть историями успеха в решении определенной, пусть и небольшой по масштабу, задачи. В результате освоения заданий базового уровня студент будет обладать набором знаний для проведения самостоятельного исследования с использованием ресурсов веб-архивов методами ближнего чтения и уметь формулировать задачи в области веб-истории для их решения с помощью более сложного инструментария, включая дальнейшее чтение.

В ходе освоения курса предполагается поэтапная подготовка к проведению самостоятельного исследования. Начальным этапом является рассмотрение кейсов уже реализованных исторических проектов, в которых получены результаты и тщательно описана методика и методология исследования. Рассмотрение теории и практики через призму выполненных проектов – ответственный шаг для подготовки студентов к самостоятельному проведению научного изыскания на следующих этапах курса. Естественно, в рамках краткосрочного образовательного курса невозможно повторить аналогичные представленным в кейсах процедуры анализа ввиду сложности используемой методологии, подходов к созданию наборов данных, а также учесть всю специфику поставленных исследовательских задач и условия выбора инструментария. Подробное рассмотрение предложенных преподавателем кейсов необходимо для того, чтобы сформировать у студента представление о научном исследовании в области веб-истории, о постановке центрального вопроса, о подборе инструментария, о процедурах сбора, организации, анализа данных, их визуализации и интерпретации. На конкретных примерах должны быть представлены и проблемы, связанные с критикой цифровых источников, как важной составляющей исследования.

По своему масштабу научные проекты в области веб-истории, представленные в научных публикациях, часто относятся к макроуровню и реализуются на основе значительного объема данных или предполагают организацию и обработку больших данных. Выполнение таких проектов требует серьезной подготовки, сотрудничества специалистов из разных областей знаний и использования (или даже создания) сложных исследовательских инструментов. Анализ крупномасштабных проектов способен сформировать у обучающихся представления о возможностях работы с веб-архивами и о проведении перспективных исследований. Рассмотрение результатов таких проектов имеет явную ценность не только для понимания всей цепочки реализации исследовательского проекта от постановки цели и получения ответов на вопросы, но и для развития навыков интерпретации результатов анализа.

Релевантные проекты, реализованные в области веб-истории, для последующего рассмотрения в ходе практических занятий могут быть опубли-

ликованы в специализированных сборниках научных статей<sup>250</sup> и периодических изданиях<sup>251</sup>, материалах конференций и публикациях профессиональных сообществ историков, использующих информационные технологии в исследованиях<sup>252</sup>, веб-архивистов и исследователей Всемирной паутины<sup>253</sup>, членов объединения исследователей Интернета<sup>254</sup> и т. д.

Доступность публикаций – необходимое условие отбора кейсов для анализа в студенческой аудитории. В этом плане особую значимость приобретают принципы открытой науки при реализации исследовательских проектов, что постепенно становится научным трендом. При формировании списка проектов для рассмотрения в качестве примеров приоритеты должны распределяться в пользу тех публикаций, которые доступны для ознакомления и содержат более подробное описание необходимых этапов проведения исследования, методологию, используемый инструментарий и т.д.

Для анализа исследовательских кейсов эффективными являются методы групповой работы. Преподаватель может предложить список отдельных проектов для предварительного изучения в малых студенческих группах по общей для всех схеме:

- анализ цели и задач исследования;
- определение источниковой базы и способов ее формирования;
- выделение примененных методов, средств и инструментов для поиска, сбора, организации, обработки и визуализации данных;
- обсуждение результатов и их интерпретация исследователями.

По возможности могут быть проанализированы наборы данных и их модели, методы формирования, особенности обработки и визуализации, а также способы интерпретации результатов. Обсуждение кейсов включает

---

<sup>250</sup> The Web as History. Using Web Archives to Understand the Past and the Present. Eds. Niels Brügger, Ralph Schroeder. London, UCL Press, 2017. <https://doi.org/10.14324/111.9781911307563>; The SAGE Handbook of Web History. Brügger N., Milligan I. (eds.) First edition. SAGE Publishing. 2018. DOI: 10.4135/9781526470546

<sup>251</sup> Internet Histories. Digital Technology, Culture and Society. URL: <https://www.tandfonline.com/toc/rint20/current> (дата обращения: 01.05.2022); Журнал «Историческая информатика». URL: <https://e-notabene.ru/istinf/> (дата обращения: 01.05.2022); Journal of Digital History. URL: <https://journalofdigitalhistory.org/> (дата обращения: 01.05.2022).

<sup>252</sup> Российская ассоциация «История и компьютер». URL: <http://aik-sng.ru/> (дата обращения: 01.05.2022).

<sup>253</sup> Research Infrastructure for the Study of Archived Web Materials (RESAW). URL: <https://resaw.eu/> (дата обращения: 01.05.2022); International Internet Preservation Consortium. URL: <https://netpreserve.org/> (дата обращения: 01.05.2022).

<sup>254</sup> Клуб любителей интернета и общества. URL: <http://clubforinternet.net/> (дата обращения: 01.05.2022).

оценку предвзятостей, связанных с выборкой и использованными наборами данных. Студенты учатся применять полученные в ходе лекционных занятий знания по критике цифровых источников и преодолению выявленных ограничений.

После ознакомления с проектами и проведения анализа группы представляют изученные проекты по предложенному плану, демонстрируя особенности, описанные в нем.

Критика цифровых источников и источниковедческий анализ являются важными этапами исследования, и их необходимо прорабатывать в любом проекте. Выявление информационного потенциала веб-архивов для изучения тех или иных вопросов истории и оценка полноты сохраненных коллекций также лежат в основе проведения качественного исследования и должны быть рассмотрены в учебном курсе. Изучение теории сквозь призму реализованных проектов – важный этап подготовки студентов к проведению исследования на следующих этапах освоения дисциплины.

Выделенные в данном учебном пособии направления веб-истории и рассмотренные исследовательские проекты (см. Гл. 4) способны стать примерами, на которые можно ориентироваться при осуществлении самостоятельного научного изыскания, и которые способны трансформировать любопытство студентов в плоскость постановки новых вопросов и формулировки исследовательских задач.

На следующем этапе обучения студентам может быть предложено придумать и спланировать «воображаемое» исследование (Imaginary Research). Такой прием используется в образовании и подразумевает практику творческого мышления и письма как своеобразный художественный акт, создание рассказа о потенциальном исследовании и возможных его результатах. Такие истории не всегда относятся к реальности, основаны на существующих материалах или полученных аргументах. Воображаемое исследование может быть посвящено гипотетическому диалогу, беседам по теоретическим вопросам и практическим решениям некоторых задач, которые можно отнести к «воображаемым» дебатам или спорам, практикой, пришедшей из древних культур<sup>255</sup>. В античной традиции оратор мог спорить с воображаемым оппонентом по тем или иным вопросам, изобретая аргументы для обеих сторон диалога. По сравнению с этой древней

---

<sup>255</sup> Светенко Т. В. Путеводитель по дебатам: учеб. пособие для педагогов и учащихся. М.: Бонфи, 2001. 135 с.

практикой воображаемое исследование можно рассматривать как художественный жанр письма или подготовленное устное выступление.

Практика воображаемого диалога используется сегодня в образовании как своего рода эксперимент в разных отраслях знания. И идея о воображаемом исследовании стала центральной для блога, ориентированного на студентов, под названием *Journal of Imaginary Research*<sup>256</sup> («Журнал воображаемых исследований»). На этом ресурсе публикуются аннотации исследований, созданные студентами из различных направлений науки, и он предназначен для раскрытия идей возможных (воображаемых) исследовательских проектов. Эти идеи, так или иначе, связаны с наукой (в широком смысле), но изложены свободно, без ограничений на владение автором определенными методами исследования, на техническую, технологическую или даже физическую возможности реализации этих методов, даже и абсурдно звучащий предмет исследования может находиться в центре внимания автора. Однако эти образцы текстов предназначены для того, чтобы сформулировать смелые предположения, вывести мышление за рамки стандартного, обойти ограничения и мотивировать исследователей творчески рассуждать.

Подходы к воображаемому исследованию рекомендуются к применению и в образовательном процессе в рамках курса по веб-истории. Можно предложить студентам практические занятия с планированием такого воображаемого исследовательского проекта, который не будет реализован. В качестве заданий рекомендуется предложить общий план масштабного исследования с постановкой цели и задач исследования, выбором источников и их источниковедческой критикой, моделированием данных для их организации и анализа. Студентам можно предложить масштабировать исследование, планируя его амбициознее, не обращая внимания на требуемые навыки для его реализации, в том числе с потенциальным использованием больших объемов данных, технологий искусственного интеллекта и пр. «Воображаемое» содержание исследования не предполагает необходимости выполнять реализованный план и осуществлять предложенное на практике. Однако польза от такого планирования существенная, поскольку проект относится к рассмотрению конкретной части действительности, реальным проблемам, стоящим перед исследова-

---

<sup>256</sup> Journal of Imaginary Research. URL: <https://journalofimaginaryresearch.home.blog> (дата обращения: 01.05.2022).

телем, формулировке реальных целей и задач и подразумевает применение определенных методов обработки данных и т.д. В этом случае воображаемое исследование выступает как способ мыслить нестандартно и генерировать новые идеи.

Вместе с тем умение формулировать исследовательский вопрос, понимание значимости изучаемого предмета являются ключевыми для историка. Подробное рассмотрение реализованных проектов в области веб-истории и планирование воображаемого исследования необходимы для формирования четкого представления у студентов об этапах проведения исследования и подготовки их к реализации индивидуального исследовательского проекта на практике.

Заключительным этапом образовательного курса является проведение студентом самостоятельного исследования в области веб-истории. В этой части курса важно закрепить изученную теорию и практические навыки работы с веб-архивами. Цель исследования в рамках данного проектно-ориентированного курса должна быть достаточно значимой и в то же время выполнимой в краткие сроки освоения дисциплины. При консультациях с преподавателем студент ставит исследовательский вопрос, который может быть решен в рамках учебного проекта, подбирает методы достижения поставленных целей и задач и проходит через все этапы его реализации.

В рамках проектной работы должен быть реализован гибкий подход к выбору методологии для использования в зависимости от уровня подготовки студентов и степени владения ими информационными технологиями. Так, например, методы моделирования, качественного анализа, подходы ближнего чтения и изучения индивидуальных записей веб-архивов могут быть введены в курс в качестве базовых, предназначенных для всех студентов, а другие методы, как например, контент-анализ, сетевой анализ и др. могут быть рекомендованы студентам, обладающим продвинутыми навыками использования цифровых инструментов, поскольку предполагают более сложные манипуляции, связанные с организацией источниковой базы, сбором данных, их обработкой и визуализацией.

В качестве итогового контроля в рамках курса студентам рекомендуется представлять результаты проведенного исследования как в устной форме с демонстрацией презентации, так и в письменной, посредством написания блога для веб-сайта курса.

Блог для данного курса имеет существенное педагогическое значение. При проектировании данного курса к его педагогическому дизайну была применена цифровая таксономия Блума<sup>257</sup> (Bloom's Digital Taxonomy), предназначенная для объяснения того, как использовать цифровые технологии значимым образом, чтобы улучшить учебный опыт студентов и получить более весомые результаты<sup>258</sup>. Структура таксономии показывает порядок развития навыков мышления и включает следующие шаги, представленные от базового до самого продвинутого уровня: запоминание, понимание, применение, анализ, оценка, создание. Указанные шаги внедрены в структуру данного курса по веб-истории и логичным образом формируют его педагогический каркас. Итоговое задание при освоении курса, связанное с реализацией проекта и созданием сообщения (поста) в блоге, относится к высшему уровню таксономии и развивает навыки мышления более высокого порядка.

Блог на сайте курса имеет и существенное практическое значение. Дело в том, что научные публикации практически не содержат примеров изучения веб-архивов «простыми» методами, понятными и адаптированными для учебного процесса, что вполне естественно и объясняется необходимостью для научного издания глубины проработки темы. Поэтому тематический блог с опубликованными студенческими постами, содержащими результаты исследований в научно-популярной форме, способен решить эту задачу и стать своеобразной библиотекой студенческих проектов по веб-истории.

Сайт Webhistoria.ru<sup>259</sup>, созданный в поддержку курса, открыт для использования и содержит разнообразные материалы, в том числе, раздел для публикации студенческих исследовательских проектов. Такой подход позволяет развить не только профессиональные компетенции специалиста в области истории, но и медийные.

---

<sup>257</sup> Churches A. (2008) Bloom's Digital Taxonomy Guide. URL: [https://www.celt.iastate.edu/wp-content/uploads/2020/07/Churches\\_2008\\_DigitalBloomsTaxonomyGuide.pdf](https://www.celt.iastate.edu/wp-content/uploads/2020/07/Churches_2008_DigitalBloomsTaxonomyGuide.pdf) (дата обращения: 01.05.2022).

<sup>258</sup> Глотова М. Ю., Самохвалова Е. А. Цифровая таксономия Блума и модель цифровой трансформации образования в учебном процессе вуза // Информатика и образование. 2019. № 6. С. 42–48. <https://doi.org/10.32517/0234-0453-2019-34-6-42-48> (дата обращения: 01.05.2022).

<sup>259</sup> WebHistoria. Сайт курса «Веб-история социума и его институтов». URL: <http://webhistoria.ru/> (дата обращения: 01.05.2022).

## ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА И ИСТОЧНИКИ

1. Глотова М. Ю., Самохвалова Е. А. Цифровая таксономия Блума и модель цифровой трансформации образования в учебном процессе вуза // Информатика и образование. 2019. № 6. С. 42–48. URL: <https://doi.org/10.32517/0234-0453-2019-34-6-42-48> (дата обращения: 01.05.2022).
2. Журнал «Историческая информатика». URL: <https://e-notabene.ru/istinf/> (дата обращения: 01.05.2022).
3. Клуб любителей интернета и общества. URL: <http://clubforinternet.net/> (дата обращения: 01.05.2022).
4. Российская ассоциация «История и компьютер». URL: <http://aik-sng.ru/> (дата обращения: 01.05.2022).
5. Светенко Т. В. Путеводитель по дебатам: учеб. пособие для педагогов и учащихся. М.: Бонфи, 2001. 135 с.
6. Churches A. (2008) Bloom's Digital Taxonomy Guide. URL: [https://www.celt.iastate.edu/wp-content/uploads/2020/07/Churches\\_2008\\_DigitalBloomsTaxonomyGuide.pdf](https://www.celt.iastate.edu/wp-content/uploads/2020/07/Churches_2008_DigitalBloomsTaxonomyGuide.pdf) (дата обращения: 01.05.2022).
7. International Internet Preservation Consortium. URL: <https://netpreserve.org/> (дата обращения: 01.05.2022).
8. Internet Histories. Digital Technology, Culture and Society. URL: <https://www.tandfonline.com/toc/rint20/current> (дата обращения: 01.05.2022).
9. Journal of Digital History. URL: <https://journalofdigitalhistory.org/> (дата обращения: 01.05.2022).
10. Journal of Imaginary Research. URL: <https://journalofimaginary-research.home.blog> (дата обращения: 01.05.2022).
11. Research Infrastructure for the Study of Archived Web Materials (RESAW). URL: <https://resaw.eu/> (дата обращения: 01.05.2022).
12. The SAGE Handbook of Web History. Brügger, N., Milligan I. (eds.) First edition. SAGE Publishing. 2018. DOI: 10.4135/9781526470546
13. The Web as History. Using Web Archives to Understand the Past and the Present. Eds. Niels Brügger, Ralph Schroeder. London, UCL Press, 2017. <https://doi.org/10.14324/111.9781911307563/>
14. WebHistoria. Сайт курса «Веб-история социума и его институтов». URL: <http://webhistoria.ru/> (дата обращения: 01.05.2022).



## ПРИЛОЖЕНИЯ

### Приложение 1

#### Глоссарий

**Веб-архивирование** – это процесс сбора фрагментов Всемирной паутины (WWW), сохранения коллекций в архивном формате, а также обеспечения доступа к ним для их использования.

**Веб-архив** – особая цифровая платформа, предназначенная для хранения архивированных веб-ресурсов и обеспечения доступа к ним.

**Доменное сканирование** – способ сохранения веба, принадлежащего к определенному домену.

**Интернет-архив (Internet Archive)** – это международный проект, одна из самых ранних инициатив по архивированию веба, миссия, которой заключается в том, чтобы доступ к разным видам цифрового контента – веб-сайтам, оцифрованным и электронным книгам, рукописям и другие текстовым источникам, аудио- и видеозаписям, изображениям и программному обеспечению различного назначения, включая игры.

**Киберсообщество** – это объединение людей в виртуальной среде, чаще всего по интересам или роду деятельности, которые поддерживают общение посредством сайтов, мессенджеров, форумов или иных средств цифровой коммуникации.

**Краулер** (от англ. crawl, «ползать») – это поисковый робот, особая программа, написанная для сбора и сохранения веб-страниц и их содержимого для последующего архивирования. Краулер следует по гиперссылкам, переходя с одной страницы на другую, сохраняя взаимосвязи между страницами, чтобы в дальнейшем можно было воспроизвести структуру и наполнение сайта в том виде, в котором было отсканировано на определенный момент времени.

**Логирование интервенций** – создание логов, записей о корректировках в специальном журнале событий (подобные регистры ведутся на веб-ресурсах для отслеживания событий – авторизаций, изменений контента и др.).

**Снапшот** – это копия веб-страницы, в совокупности ее содержимого с сохраненными связями, с которой можно осуществлять действия – переходить по ссылкам, производить действия с контентом (текстом, изображениями и др.).

**Скриншот** – это фрагмент экрана компьютера, сохраненный в формате изображения.

## Приложение 2

### Веб-архивы, репозитории и сервисы с открытым доступом к веб-архивным материалам

1. Archive-It. URL: <https://archive-it.org> (дата обращения: 01.05.2022).
2. Hrvatski Arhiv Weba. URL: <https://haw.nsk.hr> (дата обращения: 01.05.2022).
3. Internet Archive. URL: <https://archive.org/> (дата обращения: 01.05.2022).
4. New Zealand Web Archive. URL: <https://natlib.govt.nz/collections/az/new-zealand-web-archive> (дата обращения: 01.05.2022).
5. The Latin American Web Archiving Project (LAWAP). URL: <http://lanic.utexas.edu/project/archives> (дата обращения: 01.05.2022).
6. The UK Web Archive (UKWA). URL: <https://www.webarchive.org.uk/> (дата обращения: 01.05.2022).
7. TROVE. National Library of Australia. URL: <https://webarchive.nla.gov.au> (дата обращения: 01.05.2022).
8. Wayback Machine. Internet Archive. URL: <https://archive.org/web/> (дата обращения: 01.05.2022).
9. Web Archive Singapore. URL: <https://eresources.nlb.gov.sg/webarchives/landing-page> (дата обращения: 01.05.2022).
10. Национальный цифровой архив России. URL: <https://ruarxive.org> (дата обращения: 01.05.2022).
11. Список всех аккаунтов в Twitter, подвергаемых регулярной архивации в Национальном цифровом архиве России. URL: <https://hubofdata.ru/dataset/govtwitters-archives/resource/312580af-725a-4feb-9665-663e14b573ea> (дата обращения: 01.05.2022).
12. Список инициативных проектов по веб-архивированию. Wikipedia. URL: [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives) (дата обращения: 01.05.2022).

## Приложение 3

### Специализированные сервисы и приложения

1. Archive.today. URL: <http://archive.today> (дата обращения: 01.05.2022).
2. BA-WA. bibalex.org – Web Archiving. GitHub. URL: <https://github.com/arcalex> (дата обращения: 01.05.2022).
3. Conifer. Collect and revisit web pages. URL: <https://conifer.rhizome.org> (дата обращения: 01.05.2022).
4. GLAM Workbench. URLs: <https://glam-workbench.github.io/>; <https://glam-workbench.net/> (дата обращения: 01.05.2022)
5. Memento Tracer. Memento. <http://tracer.mementoweb.org/> (дата обращения: 01.05.2022).
6. Mink – Integrate Live & Archived Web +Memento. URL: <https://chrome.google.com/webstore/detail/mink-integrate-live-archi/jemoalkmipibchioofomhkgimhofbbem> (дата обращения: 01.05.2022).
7. Sketch Engine. URL: <https://www.sketchengine.eu/guide/create-a-corpus-from-the-web/> (дата обращения: 01.05.2022).
8. Time Travel. Memento. URL: <https://timetravel.mementoweb.org/> (дата обращения: 01.05.2022).
9. WARCcreate. URL: <https://warcreate.com/> (дата обращения: 01.05.2022).
10. Wayback Machine. Internet Archive. URL: <https://archive.org/web/> (дата обращения: 01.05.2022).
11. Web Archiving Integration Layer (WAIL). URL: <http://machawk1.github.io/wail/> (дата обращения: 01.05.2022).
12. Webrecorder Player. URL: <https://github.com/webrecorder/webrecorder-player> (дата обращения: 01.05.2022).
13. Webrecorder.io. URL: <https://webrecorder.net/> (дата обращения: 01.05.2022).

*Учебное издание*

**Поврозник Надежда Георгиевна**

## **Веб-история социума и его институтов**

*Учебное пособие*

*Редактор Л. Л. Савенкова*

*Корректор Л. Л. Соболева*

*Компьютерная вёрстка: Н. Г. Поврозник*

---

Подписано в печать 13.05.2022. Формат 60×84/16.

Усл. печ. л. 9,07. Тираж 300 экз. Заказ 105

---

Издательский центр  
Пермского государственного  
национального исследовательского университета.  
614990 г. Пермь, ул. Букирева, 15

Типография ПГНИУ.  
614990 г. Пермь, ул. Букирева, 15