

АРХИВЫ В ИНФОРМАЦИОННУЮ ЭПОХУ И ВРЕМЯ DATA SCIENCE

Доклад посвящен краткому обзору истории отечественных государственных программ внедрения информационно-коммуникационных технологий и эволюции их основного содержания, а также современному уровню применения технологий и методов Data Science (DS) (машинное обучение, искусственный интеллект (Artificial Intelligence, AI), компьютерное зрение, нейросети и т.п.) в архивной практике и исторических исследованиях, выполненных на основе комплексов архивных документов. Автор представляет деятельность Международного совета архивов и зарубежных сетевых архивных сообществ, специализирующихся на изучении возможностей использования искусственного интеллекта в решении практических задач архивного дела (отбор документов на хранение, экспертиза ценности, составления научно-справочного аппарата, цифровая реставрация, организация доступа к электронным документам и т.п.); характеризует основные итоги выполненных за рубежом научных проектов, а также обозначает перспективные направления исследований.

Ключевые слова: автоматизация, информатизация, цифровая трансформация, основные направления архивной деятельности, Data Science, архивные документы, наборы данных

Эпоха глобального информационного общества, наступление которой было зафиксировано в Окинавской хартии 2000 г., за первые без малого четверть века прошла несколько стадий, затронувших в той или иной форме все сферы человеческой деятельности, в том числе и архивы.

Переход к электронному документообороту, декларированный в 2002 г. в Программе «Электронная Россия» [Федеральная целевая программа..., 2002], в 2010 г. сменился отраслевыми программами информатизации, разработка которых предусматривалась Государственной программой «Информационное общество» [Государственная программа..., 2010], в которой под «информатизацией» подразумевались создание и развитие информационно-коммуникационной инфраструктуры различных отраслей, перевод (адаптация) основных видов деятельности органов управления, предприятий и организаций в цифровой вид, оцифровка активов и оказание услуг населению в цифровом формате. Нельзя сказать, что эти задачи для архивной сферы не ставились и ранее (Программа информатизации

архивного дела России (1997–2000) [Юмашева, 2016, с. 101–162]), однако их реализация в 2010-е гг. приобрела особую государственную важность, которая выразилась в том числе в создании Программы информатизации Федерального архивного агентства и подведомственных ему учреждений (2011–2020 гг.), утвержденной в 2011 г. и ставшей модельной программой информатизации архивной сферы на всех уровнях: от федерального до муниципального [Программа информатизации..., 2011].

Следующий этап внедрения ИКТ в жизнь российского общества происходит в рамках реализации Программы «Цифровая экономика» и созданных в ее развитие ведомственных программ цифровой трансформации [Положение..., 2022], в которых под цифровой трансформацией понимается «совокупность действий, осуществляемых государственным органом, направленных на изменение (трансформацию) государственного управления и деятельности государственного органа по предоставлению им государственных услуг и исполнению государственных функций за счет использования данных в электронном виде и внедрения информационных технологий в свою деятельность».

Очевидно, что решение подобной задачи требует глубинных изменений, затрагивающих все процессы, процедуры и операции в деятельности федеральных и региональных органов государственного управления и подчиненных им организаций. Эти изменения должны быть основаны на активном применении информационно-коммуникационных технологий и методов Data Science, переводе рутинных операций не просто в автоматизированный режим, а их реализацию с использованием алгоритмов искусственного интеллекта, создания наборов данных, позволяющих обеспечивать выполнение запросов и оказания услуг в максимально короткие сроки и с максимально возможным качеством [Прохоров, Коник, 2019].

Так, в Ведомственной программе цифровой трансформации Федерального архивного агентства на 2022–2024 гг. перечисленным технологиям и методам, а также их применению в архивной практике уделяется повышенное внимание [Об утверждении..., 2022].

Вместе с тем необходимо отметить, что в вопросах цифровой трансформации, создания наборов данных, открытых данных и применения методов DS российская архивная сфера в определенном смысле отстает от мировых тенденций.

К примеру, еще в 2019 г. секретарь Международного совета архивов (ICA) А. Селеш заявила, что изучение возможностей применения DS и AI в архивах будет развиваться, как никакое другое [Seles, 2019]. В результате была создана профильная группа ICA – HAIRA,

а также соответствующий информационный ресурс, представляющий итоги исследований (URL: <https://haira.clir.org/artificial-intelligence-and-archives/>).

Большую активность в данном вопросе проявляют национальные и профильные общественные объединения архивистов и специалисты крупнейших архивов мира. В Великобритании функционирует сеть AURA (AURA Network, URL: <https://www.aura-network.net/>), в США – портал AEOLIAN (Artificial Intelligence for Cultural Organizations, URL: <https://www.aeolian-network.net/>), которые принимают активное участие и координируют проекты, посвященные изучению возможностей применения методов DS и алгоритмов AI в архивном деле. На сайтах упомянутых организаций публикуются отчеты о реализованных проектах, а на вебинарах обсуждаются результаты деятельности профессиональных сетевых сообществ по применению AI в архивах.

Кроме того, следует отметить, что с середины 2010-х гг. количество публикаций в профессиональных архивных изданиях по проблематике применения DS и AI исчисляется тысячами. Практически все исследования могут быть сгруппированы в четыре блока [Colavizza et al., 2022]:

- вопросы теории архивного дела в контексте DS, в том числе теоретическое осмысление перспектив профессии архивиста, ее модификации в связи с активным проникновением информационных технологий в архивную практику и возникновением междисциплинарного направления Computational Archival Science (URL: <https://ai-collaboratory.net/cas/>);

- практические результаты автоматизации процессов комплектования и учета архивной документации на основе методов DS (онтологии, базы знаний, наборы данных, машинное обучение, семантические сети, НЛП, методы формирования (в т. ч. обогащения) и состава метаданных и т. п.);

- организация доступа к архивной документации (выявление документов с конфиденциальной информацией, редактирование конфиденциальной информации с использованием инструментов цифровой криминалистики, методы совершенствования поиска («капсульные» тематические модели, «дистанционное чтение»); повышение доверия к оцифрованному/цифровому контенту, «датафикация, интеграция, интероперабельность и доступность для анализа»). При этом во многих исследованиях применение AI для поиска информации в архивных каталогах оценивается весьма неоднозначно. Специалисты отмечают «предвзятость» результатов применения поисковых механизмов, основанных на использовании AI и предлагают механизмы преодоления этой проблемы;

– новые формы цифровых архивов. Растущие объемы оцифрованных и born-digital архивных документов требуют переосмысления и автоматизации организации хранения и организации доступа к архивным документам. Среди методов, которые могут быть использованы для решения данных проблем, рассматриваются распределенное хранение/блокчейн, внедрение хеш-кодов и т. п.

Резюмируя анализ публикации по проблемам применения методов DS в архивах, можно перечислить те направления, использование новых методов в которых оценивается специалистами, как перспективное:

- масштабируемость бизнес-процессов, реализуемых в архивных службах;
- изучение возможностей применения DS для комплектования и научно-технического обеспечения информационных ресурсов, социальных сетей, электронной почты и т. п., а также организация и функционирование аудиовизуальных репозиторий, блокчейн и т. п.;
- цифровая консервация, реставрация и сохранение;
- ИТ-инфраструктура для архивных исследований, создания и хранения коллекций;
- краудсорсинг и архивы;
- анализ текстов документов для архивной обработки, определения возможности доступа к документам, идентификация личных данных, оценки, регистрации, описания документов;
- применение аналитических методов к архивным материалам;
- большие данные и архивная теория и практика;
- большие данные и построение памяти/идентичности; архивные корпуса больших данных и справочные коллекции; связанные данные и архивы; подготовка данных для исследований из архивных материалов; юридические и этические вопросы.

Этим перечнем возможные области и задачи применения DS и AI в архивной сфере не исчерпываются. Однако необходимо подчеркнуть сдержанную позицию руководителей направления цифровой трансформации ведущих архивов мира в вопросах применения AI и их ориентацию на реализацию «малых проектов», призванных апробировать подходы DS для решения узких задач практики архивного дела.

В связи с этим представляется, что изучение возможностей применения методов DS в архивной практике является одним из ключевых направлений в развитии современной архивной науки. Не последнюю роль в этом вопросе играет междисциплинарный подход и активизация сотрудничества архивистов и специалистов-истори-

ков, информатиков, математиков, представляющих научно-исследовательские центры и ВУЗовскую науку.

Государственная программа «Информационное общество (2011–2020 годы)» : одобр. распоряж. Правительства РФ от 20.10.2010 № 1815-р // Министерство цифрового развития, связи и массовых коммуникаций РФ. URL: <https://digital.gov.ru/ru/activity/programs/1/> (дата обращения: 01.12.2022).

Об утверждении ведомственной программы цифровой трансформации Федерального архивного агентства на 2022–2024 годы : приказ Федеральн. архивн. агентства от 6.05.2022 № 57 // Федеральн. архивн. агентство : офиц. сайт. URL: https://archives.gov.ru/documents/prik57_2022.shtml (дата обращения: 01.12.2022).

Положение о ведомственных программах цифровой трансформации (в ред. постановления Правительства РФ от 17.02.2022 № 195) : утв. постановлением Правительства РФ от 10.10.2020 № 1646 // Правительство России. URL: <http://government.ru/docs/all/130305/> (дата обращения: 01.12.2022).

Программа информатизации Федерального архивного агентства и подведомственных ему учреждений на 2011–2020 гг. : утв. приказом Федеральн. архивн. агентства от 2.12.2011 № 104 // Федеральн. архивн. агентство : офиц. сайт. URL: <https://archives.gov.ru/programs/informatization.shtml> (дата обращения: 01.12.2022).

Прохоров А., Коник Л. Цифровая трансформация. Анализ, тренды, мировой опыт // М.: ООО «КомНьюс Групп», 2019.

Федеральная целевая программа «Электронная Россия (2002–2010 годы)» : утв. постановлением Правительства Российской Федерации от 28.01.2002 № 65. // Министерство цифрового развития, связи и массовых коммуникаций РФ. URL: <https://digital.gov.ru/ru/activity/programs/6/> (дата обращения: 01.12.2022).

Юмашева Ю. Ю. Информатизация архивного дела в Российской Федерации (1991–2015). Научные исследования в области применения информационных технологий : моногр. // М. ; Берлин: Директ-Медиа, 2016.

Seles A., Secretary General of International Council on Archives. A Brave New World: Artificial Intelligence and Archives // ICA. 2019. URL: <https://www.ica.org/en/anthea-seles-presentation-on-artificial-intelligence-and-archives> (дата обращения: 01.12.2022).

Colavizza G., Blanke T., Jeurgens C., Noordegraaf J. Archives and AI: An Overview of Current Debates and Future Perspectives. // J. Comput. Cult. Herit. 15., 2022. № 1. Article 4 (February 2022). URL: <https://doi.org/10.1145/3479010> (дата обращения: 01.12.2022).