

УДК 004.9:930.25+930.22

«Цифровые технологии в архивоведении и документоведении»

**“Digital technologies
in Archival science and
documentation”**

22 сентября 2023 года в г. Казань в рамках Международного форума KAZAN DIGITAL WEEK – 2023 состоялась секция «Цифровые технологии в архивоведении и документоведении».

Целью проведения секции являлось содействие в обмене научно-технической информацией в области архивного дела, консолидация научно-практических активов, способных эффективно решить задачи импортозамещения и импортоопережения, профессиональная ориентация с целью трудоустройства в архивную отрасль.

В ходе работы секции был рассмотрен ряд вопросов, отражающих актуальные проблемы и перспективные направления развития архивной отрасли: практики использования в архивной отрасли «сквозных» цифровых технологий, таких как системы распределенного реестра в части управления интеллектуальными правами; искусственно-интеллекта в части распознавания рукописных текстов метрических книг, в том числе на арабской графике, что необходимо для масштабных историко-демографических работ, а также в части аннотирования и анализа тематических комплексов аудиовизуальных документов.

В мероприятии приняло участие более 200 человек. Среди них: ученые-исследователи, эксперты в сфере трансформации права, интеллектуальной собственности, креативных индустрий и цифровой экономики, сотрудники архивных и научных учреждений.

Спикерский состав был представлен ведущими учеными и практиками, работающими в области управления интеллектуальными правами из Москвы и Санкт-Петербурга.

В рамках Международного форума KAZAN DIGITAL WEEK было подписано соглашение о сотрудничестве по созданию, развитию и внедрению цифровых решений для формирования единого информационного пространства по аудиовизуальным

архивным материалам между ГБУ «Государственный архив Республики Татарстан» и Киноархивом Нэт-Фильм*.

Из резолюции по итогам работы секции «Цифровые технологии в архивоведении и документоведении» Международного форума KAZAN DIGITAL WEEK – 2023

22 сентября 2023 г.
г. Казань

[...] Значительная часть выступлений была посвящена изучению складывающихся мировых практик по управлению интеллектуальными правами с использованием технологий блокчейн.

В Российской Федерации инфраструктура управления правами и объектами интеллектуальной собственности на основе технологии распределенных реестров, была создана по инициативе крупнейших игроков рынка, а к 2021 году в соответствии с указом Президента Российской Федерации приобрела общественно-государственный статус.

В координационный совет ООГО РЦИС вошли главы федеральных министерств цифрового развития, связи и массовых коммуникаций, культуры, высшего образования и науки.

С докладами на данную тему выступили, кроме специалистов из архивной сферы, эксперты в области цифровой трансформации права, интеллектуальной собственности, креативных индустрий и цифровой экономики. Принимая участие в таких мероприятиях, эксперты транслируют правообладателям информацию о том, как развивается институт интеллектуальной собственности и какие тренды преобладают. Также для евразийского пространства нельзя забывать о научно-познавательной составляющей мероприятия, так как именно на Казанском Форуме говорят о будущем и о том, что и как полезно делать в настоящем.

В своих выступлениях эксперты затронули следующие вопросы:

- интеграция каталогов обязательного экземпляра аудиовизуальной продукции и кино-, фото-, фонодокументов Архивного фонда с информационно-телекоммуникационной инфраструктурой управления правами на результаты творческой деятельности и один из первых в России практический опыт по интеграции каталогов ГБУ «Государственный архив Республики Татарстан» и Киноархива Нэт-Фильм; возможность создания безопасных цифровых каналов, позволяющих с разрешения правообладателя подключение к банку эталонных копий произведений, хранящихся в государственном депозитарии (в том числе в государственных архивах) для различных сервисов, выявляющих и предотвращающих нарушения исключительных прав;

- упрощение доступа к историко-культурному наследию (архивам, музеям фондам, государственным депозитариям аудиовизуальной, книжной продукции и т.д.);

- снижение транзакционных издержек при заключении лицензионных договоров и о возможности заключения таких договоров в цифровом виде;

- перспективы развития законодательства в сфере интеллектуальной собственности, в том числе в целях доступа к историко-культурному наследию с учетом законных интересов авторов и иных правообладателей, включая неизвестных.

* Фоторепортаж работы секции «Цифровые технологии в архивоведении и документоведении» см. на третьей стороне обложки журнала.

Заслушав и обсудив доклады и выступления по управлению интеллектуальными правами с использованием технологий блокчейн, участники конференции пришли к выводу о необходимости:

- положительно оценить опыт сотрудничества ГБУ «Государственный архив Республики Татарстан» и Киноархива Нэт-Фильм в части развития отечественного программного обеспечения для создания публичного доступа к каталогам аудиовизуальных произведений и интеграции каталогов с общественно-государственной блокчейн-инфраструктурой – сетью РЦИС.РФ;

- продолжения работы, направленной на выработку научно обоснованных предложений по развитию законодательства в сфере интеллектуальной собственности, с учетом передового мирового опыта и российского опыта;

- построения стандартов и алгоритмов взаимодействия участников медиа и интернет рынка с целью снижения количества ссылок на пиратский контент и издержек, связанных с правомерным использованием;

- обеспечения принятия эффективных мер, направленных на пресечение недобросовестных действий правообладателей, с целью защиты прав национальных участников рынка, в том числе потребителей, с соблюдением соответствующего мировым стандартам уровня правовой охраны в сфере интеллектуальной собственности;

- формирования рабочей группы для проработки организационного плана сотрудничества в развитии объединенного информационного пространства по архивным аудиовизуальным материалам – централизованного сервиса доступа к архивным кино-, фото-, фонодокументам.

Для обеспечения эффективного поиска в соответствии с порядком, установленным постановлением Правительства Российской Федерации от 29.09.2021 г. №1644 «Об общероссийской общественно-государственной организации «Российский центр оборота прав на результаты творческой деятельности» имеется необходимость стандартизировать требования к описаниям кино-, фото-, фонодокументам и обеспечить автоматизированную актуализацию информации, опубликованной на централизованном сервисе.

В части обеспечения соблюдения авторских, смежных и иных интеллектуальных прав реализовать интеграцию с сетью РЦИС.РФ, а именно:

- обеспечивать регистрацию предоставления лицензии (содержания сделки) в реестре РЦИС.РФ;

- предоставлять информацию для обществ по коллективному управлению правами, необходимую для корректного расчета отчислений авторам и иным правообладателям в соответствии с их компетенцией;

- обеспечить возможность полностью безбумажного оформления требуемых лицензий, с прозрачными условиями ценообразования, согласованными с правообладателями.

[...] Значительная часть выступлений на секции была посвящена вопросам использования искусственного интеллекта, которые были бы важны для исторических исследований. Это в первую очередь распознавание речи и компьютерная идентификация символов арабской графики.

Подавляющая часть исторического наследия, касающегося Республики Татарстан, представлена в арабографических рукописях на арабском, татарском и фарси. Значительная их часть – это метрические книги, основа генеалогических исследований, в связи с чем особую значимость приобретает задача машинного распознавания символов арабской графики, так как распознавание и перевод метрических записей в ручном режиме, требует большого количества времени и является

затратным. Только в Государственном архиве Республики Татарстан количество листов, подлежащих распознаванию и переводу, превышает 7 млн. 514 тыс.

Особенности символов арабской графики и, в частности, трудности ввода в ЭВМ, неприемлемость существующих алгоритмов распознавания символов (так как не учитывают особенностей и специфики их написания) делают проблемы распознавания и идентификации их особенно важными.

Критический обзор выступлений, посвященных распознаванию и идентификации, показал, что распознавание письменных знаков, в особенности арабских представляет собой сложный процесс переработки информации, задача распознавания рукописных знаков окончательно еще не решена, так как существует много как теоретических, так и практических трудностей, связанных с огромным многообразием возможных написаний отдельных рукописных знаков, и поэтому общего решения для распознавания обычного почерка не существует. Для решения же различных задач распознавания необходимо в каждом случае построить свое, соответствующее задаче, описание объектов.

Немаловажным фактором для надежного распознавания рукописных текстов является необходимость в предварительном восстановлении угасающего текста документа. Архивные документы, которые хранились десятки и сотни лет без должных условий, постепенно утрачивают свою читаемость: от воздействия воды и сырости, появления грибка, разрушения чернил и проч. Тогда буквы и изображения на бумажном носителе могут стать плохо различимыми. Однако, некоторая информация может остаться в невидимой для человеческого глаза части оптического спектра.

Использование гиперспектральных сканеров, правильно подобранные параметры сканирования по скорости и по скорости формирования изображения позволяют сформировать так называемый гиперкуб (множество спектральных) каналов. Компьютерная программа выявляет ключевые и наиболее значимые спектральные каналы, которые фиксируют плесень, чернила или просто бумагу. Таким образом, на определенной длине волн может проявится угасающий текст. Новая методика позволит значительно повысить информативность угасающих текстов исторических документов.

По данной тематике на сессии выступили, кроме специалистов из архивной сферы, эксперты в области технологий применения ИИ, отраслевые эксперты и представители науки. В своих выступлениях эксперты затронули следующие вопросы:

- усовершенствование технологии распознавания рукописных текстов и транскрибирование исторических документов;
- проблема автоматизированного распознавания рукописного текста, связанная с недостаточной точностью исходного набора данных;
- подходы обучения нейронной сети с применением различных алгоритмов;
- особенности распознавания и интеллектуальной обработки информации, содержательно подобных друг другу, но не обладающих подобной структурой;
- проблема доверия к информации, которую генерирует ИИ, методическое обеспечение создания и развития прикладных видов ИИ.