

Цифровая библиография: библиотеки в поисках инновационных инструментов библиографической деятельности

По материалам статьи М. Ю. Нецерет «Цифровая библиография в поисках инновационных инструментов библиографической деятельности» (Научные и технические библиотеки. 2021. № 7. С. 33–50).



Цифровизация – глобальный процесс современности, преобразующий социально-политическую, финансово-экономическую и духовно-идеологическую сферы жизнедеятельности социума. Основными факторами, определяющими сегодня развитие общества, являются глобализация информационных потоков и, как следствие, накопление значительного объёма данных в различных областях человеческой деятельности. Предполагается, что в ближайшие 7 лет глобальное информационное пространство вырастет более чем в 5 раз.

Цифровые технологии проникают во все сферы общественной и экономической жизни большинства стран мира. К ним относятся: большие данные (*big data*); искусственный интеллект (*artificial intelligence*); машинное обучение (*machine learning*); технологии беспроводной связи (*wireless connection*); компоненты робототехники и сенсорики; системы распределённого реестра; технологии виртуальной и дополненной реальностей и др. Благодаря элементам цифровизации, *составляющим в комплексе набор методологических подходов и инструментов*, «мощные компьютерные системы могут находить в своих массивах памяти данные, которые изначально могли и не предполагаться, поскольку были не известны их новые взаимосвязи и закономерности».

Перспективы применения цифровых технологий в библиотечной деятельности рассматривались на сессии Всемирного библиотечного и информационного конгресса ИФЛА (24–30 авг. 2019 г., Афины, Греция), на котором было представлено для обсуждения Руководство ИФЛА по запуску проектов цифровой унификации. Оно содержало рекомендации для библиотек по планированию проектов с целью предоставления цифрового доступа к библиотечным фондам и тематическим коллекциям.

В России национальным проектом «Культура», *реализация которого осуществляется в соответствии с Указом Президента Российской Федерации от 7 мая 2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года»*, предусмотрена глобальная цифровизация сферы культуры. Согласно проекту, для формирования информационного пространства знаний необходимо создать безопасную информационную среду; усовершенствовать механизмы обмена знаниями; обеспечить функционирование Национальной электронной библиотеки и иных государственных информационных систем; создать условия для популяризации русской культуры и науки за рубежом; принять меры по эффективному использованию современных информационных платформ для распространения достоверной и качественной информации; обеспечить насыщение рынка доступными и легальными медиапродуктами и сервисами российского производства.

Цифровизация библиографической деятельности библиотек

Количественный рост объёмов библиографической информации обуславливает необходимость перехода к цифровому инструментарию её сбора, анализа, обработки и распространения. Цифровые технологии обладают значительным потенциалом для усовершенствования процессов библиографирования, каталогизации, индексирования, поиска ресурсов, извлечения ключевой информации из текстов и др. Много возможностей для применения в библиографической деятельности предоставляют **системы искусственного интеллекта (СИИ)**, предназначенные для работы с мега объёмами информации.



Иллюстрацией, в частности, может служить опыт Библиотеки Конгресса США, где действует система оптического распознавания и классификации визуального контента – **«Газетный навигатор»** (*Newspaper navigator*). Технология машинного обучения позволила обеспечить быструю обработку и сортировку оцифрованных материалов. Всего за 19 дней «Газетный навигатор» классифицировал более 16 млн страниц. Сегодня с помощью этого инструмента пользователи Библиотеки Конгресса США могут искать карты, комиксы, фотографии, иллюстрации, рекламу и многое др. в обширном архиве оцифрованных газет, датируемых 1900–1963 гг. Интерфейс навигатора позволяет не только осуществлять поиск, но и загружать изображения, просматривать полные выпуски газет, получать библиографическую информацию.

В библиотеках и архивах цифровые технологии используются при создании метаданных в процессе оцифровки фотоизображений, так как они могут быть обучены распознавать местоположение объектов, запечатлённых на фотографии (включая место, где стоял фотограф). Таким образом, эти системы могут быть чрезвычайно полезны для обработки и каталогизации фотоколлекций – благодаря набору метаданных, установленных при помощи искусственного интеллекта, объекты из фотоколлекций легко идентифицируются.

Интеллектуальный анализ текстов

Технологии интеллектуального анализа текстов широко используются в управлении знаниями (*knowledge management*). Они способны извлекать смысл из большого количества научных публикаций в виде концептов, кластеризовать и индексировать тексты в соответствии с деталями содержания, устанавливая связи между источниками.

СИИ не только выполняют автоматический сбор информации из широкого спектра структурированных и неструктурированных источников, но и приводят её к единому формату посредством компиляции. Результатом такой обработки текста становится

«метазнание». При этом СИИ используют комбинацию полнотекстового анализа, вычислительной семантики, нейронных сетей и алгоритмов машинного обучения. СИИ предоставляют только актуальную информацию, почерпнутую из авторитетных источников, позволяя избежать информационной перегрузки, что особенно важно в контексте справочно-библиографического обслуживания.

СИИ, предназначенные для анализа текстов (*Yewnofor Education*), тестируются академическими библиотеками в нескольких престижных учебных заведениях, включая Стэнфордский и Гарвардский университеты, Калифорнийский университет в Беркли, Массачусетский технологический институт. СИИ играют важную роль в формировании критического мышления у начинающих исследователей, так как дают возможность проанализировать «ландшафт» избранной темы.

Современные СИИ используют «концептуальный» поиск, который более эффективен, чем поиск по ключевым словам. Системы искусственного интеллекта, основанные на технологии интеллектуального анализа текста, могут успешно применяться при поиске научной литературы, а также позволяют реализовать идею концептографического обслуживания на практике.

Кроме того, технологии интеллектуального анализа текстов могут использоваться для извлечения описательных метаданных из сетевого контента и формирования стандартных библиографических записей, а также информационных пакетов – комплексных библиографических продуктов, включающих различные виды документов по определённой тематике (библиографический список, аналитическую справку, бэкграундер, факт-лист, фрагменты текстов, иллюстративный материал и пр.).

Библиографические менеджеры

Всё большую популярность в мире приобретают персональные библиографические инструменты – **библиографические менеджеры** (*reference manager*), создаваемые для сбора, хранения, обработки и последующего использования библиографических метаданных (*EndNote, Mendeley, ReadCubePapers, RefWorks, Sciwheel, Zotero, Paperpile* и др.). С помощью автоматического анализа текстов публикаций можно получить такие метаданные, как название, авторы, аннотация, список литературы. Программы управления библиографической информацией позволяют формировать персональную библиографическую БД.

Менеджер состоит из трёх основных частей: базы данных, в которой хранится библиографическая информация; модуля импорта, позволяющего автоматически извлекать библиографические записи из сетевого контента; текстового редактора, предназначенного для формирования списка литературы. Как правило, библиографические менеджеры осуществляют прямой импорт записей, а в ряде случаев – полнотекстовых документов в формате PDF из баз данных и других источников.

Системы управления библиографической информацией могут быть интегрированы с текстовыми процессорами таким образом, что список ссылок создаётся автоматически и добавляется в документ. Ссылки оформляются в выбранном стиле цитирования. В частности, *Mendeley, Zotero* и *Papers* используют более 6 тыс. стилей.

С помощью библиографических менеджеров пользователи формируют личные библиотеки. Встроенные средства просмотра PDF-файлов позволяют им читать тексты, делать заметки, выделять ключевые отрывки и комментировать документы. Библиографические менеджеры могут быть дополнены инструментами онлайн-синхронизации баз данных, совместного доступа к базам, автоматической проверки достоверности ссылок.

Цифровые сервисы избирательного распространения информации



Цифровые технологии существенно изменили процесс **избирательного распространения информации (ИРИ)**. С середины 1990-х гг. за рубежом развиваются онлайн-сервисы, основанные на применении СИИ: *Selectiv edissemination of information*, *Current awareness services*, *Alerting services* и др. Они обеспечивают непрерывность в информировании пользователей о выходе актуальных публикаций, способствуют повышению целенаправленности научной информации, экономят время учёных и специалистов при поиске релевантной литературы.

Функциями ИРИ оснащены сайты библиотек, издательств (*Elsevier*, *Springer*, *Wiley* и др.), компании, формирующие базы научной информации, – *Web of Science*, *Scopus*, *CurrentContents*, *ProQuest* и др. Пользователь самостоятельно регистрируется в выбранной им системе, создаёт индивидуальный тематический профиль, вводит ключевые слова по теме своих научных интересов, выбирает ресурсы, устанавливает частоту и формат оповещений. Пакетные сообщения система формирует автоматически.

«Цифровые помощники» в виртуальном справочно-библиографическом обслуживании

Сегодня во многих библиотеках действуют виртуальные справочные службы для дистанционного **справочно-библиографического обслуживания (СБО)** пользователей, в том числе в режиме реального времени. Цифровые технологии позволяют решать многие практические задачи этого вида обслуживания: автоматическую обработку запроса, представленного на естественном языке; его «перевод» на профессиональный библиографический язык; реализацию диалога СИИ с пользователем в ходе уточнения запроса и формирования ответа системы; библиографический поиск с учётом смыслового содержания многозначных слов, синонимов и тематически связанных слов и др.

Применение **чат-ботов** в онлайн-СБО – пример прикладного использования искусственного интеллекта в библиотеках. «Виртуальный собеседник» может задавать типовые вопросы и отвечать на них, искать информацию по запросу и выполнять простейшие задания. Наибольшее распространение получили самообучающиеся чат-боты, функционирующие на базе алгоритмов *machine learning* и методов понимания естественного языка. Их эффективность обусловлена способностью накапливать знания в процессе взаимодействия с пользователями.

Типовой цикл работы **чат-бота** включает следующие действия: получение запроса от пользователя; анализ запроса и его «понимание»; выполнение действий согласно заранее определённому сценарию – скрипту; генерация ответа на естественном языке и его отправка пользователю; сохранение запроса, контекста и параметров диалога для обработки последующих обращений.

В зарубежных библиотеках используются «цифровые помощники» – *Amazon Alexa*, *Google Assistant*, *Apple Siri*, *Microsoft Cortana* и др. Компании, специализирующиеся на

производстве телекоммуникационного оборудования, помогают библиотекам создавать адаптированные к библиотечно-информационным продуктам и услугам приложения, действующие на платформах *Alexa* или *Google Assistan*.

Для Центральной библиотеки Хельсинки «Ооди» (Oodi) финская технологическая компания *Headai* создала приложение, помогающее ориентироваться в информационном пространстве библиотеки. Виртуальные ассистенты определяют местоположение книг, рекомендуют пользователям наиболее популярные издания, знакомят с новыми поступлениями. Приложение способно работать на финском, шведском и английском языках; не требует входа в систему и не собирает персональные данные.

В библиотеках США цифровые ассистенты оказывают базовые справочные услуги и помогают проводить поиск в каталогах.

Системы интегрированного поиска (discovery-сервисы)



Можно предположить, что внедрение СИИ в библиографическую деятельность позволит в перспективе решить проблему, которая уже давно беспокоит профессиональное библиографическое сообщество. А. В. Соколов в статье «Задачи русского библиографоведения в XXI веке» обозначил эту проблему как противоречие «между грандиозностью библиографических ресурсов и возможностью ориентации в них».

В соответствии с «Основами государственной культурной политики», утверждёнными Указом Президента РФ от 24.12.2014 г. № 808, одной из основополагающих задач в ходе формирования благоприятной для развития личности информационной среды стало образование единого российского электронного пространства знаний, основанного на оцифрованных книжных, архивных и музейных фондах по различным сферам деятельности и отраслям знаний, собранных в Национальную электронную библиотеку и другие национальные электронные ресурсы.

Задача организации единого поискового пространства и создания комфортной поисковой среды в отдельных библиотеках решается при помощи специальных *discovery-сервисов* интегрированного поиска, цель которых сводится к объединению в доступный для эффективного поиска информационный универсум как можно большего количества различных источников информации. Единый поисковый интерфейс обеспечивает одновременный поиск во всех ресурсах библиотеки, включая локальные электронные коллекции и удалённые полнотекстовые БД, к которым библиотека имеет лицензионный доступ.

Системы типа *Web Scale Discovery* являются естественным развитием каталогов нового поколения. Наиболее известны среди них *EBSCO Discovery Service (EBSCO Information Services)*, *WorldCat (OCLC)*, *Summon (Serials Solutions)*. Поисковая система *Primo (ExLibris)* определяет взаимосвязи между информационными ресурсами и таким

образом выявляет релевантные материалы. Система *Primo* (последняя версия – *Primo VE*) особенно эффективна в комплекте с *Primo Central Index*, обеспечивающей доступ к более чем 1 млрд научных материалов, полученных из различных агрегаторов информационных ресурсов и хранилищ открытого доступа.

Вполне вероятно, что распространение *discovery-сервисов* является началом будущей интеграции мировых библиографических ресурсов. Возможно, формирование «всемирного каталога» произойдёт благодаря СИИ, способным осуществлять отбор, анализ и обработку метаданных из сетевого пространства.

Цифровая библиография. Подготовка «цифровых библиографов»



Итак, мы видим, что цифровизация кардинально изменила условия, в которых осуществляется библиографическая деятельность. Исторически и традиционно она была направлена на стабильные объекты – документы, в которых «закреплена информация, предназначенная для преднамеренной и устойчивой фиксации знаний». Библиографические методы создавались с учётом документной реальности, где приоритет имел факт публикации. Цифровая реальность характеризуется разнородностью информационных явлений, сменой коммуникативных форматов, неустойчивостью и изменчивостью объектов библиографического отражения и поиска. Цифровая библиография предполагает взаимодействие преимущественно с содержанием, вне зависимости от формы.

Цифровизация библиографической деятельности выражается в появлении новых методов, предполагающих использование инновационных программных инструментов в качестве альтернативы традиционному библиографическому подходу. Цифровизация требует подготовки не только технических специалистов, но и «цифровых библиографов», призванных вывести библиотеки на новую ступень эволюции, отвечающую современному уровню развития информационных технологий. Как отмечает И. В. Тимошенко, «область применения интеллектуальных экспертных систем в библиотеках чрезвычайно широка. Они могут быть применены практически во всём спектре информационно-справочных услуг библиотеки. Роль библиотечных работников в таких библиотеках кардинально меняется. Их область деятельности – инженерия знаний, содержащихся в библиотечных фондах».

Прогнозируется спрос на библиографов, обладающих знаниями и навыками в области анализа информации и способных раскрыть информационный потенциал, содержащийся в материалах – объектах библиографической деятельности.

Для того чтобы цифровизация библиографических услуг стала реальностью, библиографам необходимо выйти за пределы своего традиционного мышления, отказаться

от тезиса «мы всегда так делали!», осваивать новые инструменты и методы деятельности и активно включиться в соревнование за успех на интенсивно развивающемся рынке услуг поиска и распространения информации.

В конце 2020 г. было подписано соглашение о сотрудничестве между Российской государственной библиотекой и Московским государственным лингвистическим университетом, предметом которого стала «подготовка специалистов, которые будут способны разрабатывать стратегии цифровой трансформации деятельности библиотек, поиска информации в цифровой среде, принимать решения, опираясь на результаты анализа данных компьютерных информационных библиотечных систем».

В рамках сотрудничества была представлена программа «Цифровая библиография». Генеральный директор РГБ В. В. Дуда отметил: «Цифровизация в библиотечной отрасли не сводится к внедрению информационных технологий – это обеспечение возможности получить нужную информацию, значимую, экспертно оценённую или критически осмысленную, максимально соответствующую запросу пользователя... Для того чтобы библиотеки и дальше могли выполнять свои важнейшие функции навигации в мире информации, необходим совершенно новый подход».

Из выше сказанного можно сделать вывод, что применение технологий искусственного интеллекта и машинного обучения означает неизбежное реформирование всей библиотечно-информационной сферы. Прежняя модель организации библиографической деятельности устарела, потому что «ориентирована на идеал унитарного и конвергентного знания». Задачи библиографов сегодняшнего дня – активно осваивать и применять цифровые технологии, учиться управлять информационными мегамаассивами в реальном времени. По мере развития цифровых технологий библиографы должны будут непосредственно участвовать в разработке систем, сервисов, программ и приложений, ориентированных на предоставление библиотечно-информационных продуктов и услуг, – для того, чтобы традиционные библиографические принципы, ценности и этика были изначально заложены и сохранены в инновационных технологиях искусственного интеллекта.

<https://doi.org/10.33186/1027-3689-2021-7-33-50>

Материал подготовлен ведущим библиографом ИБО Ириной Русановой