

ЧИТАЛЬНЫЙ ЗАЛ • READING ROOM



Статья • Article

Р. Р. Туманян

Всеармянская цифровая библиотека – современная платформа для научных исследований

Аннотация. Цель статьи – представить Всеармянскую цифровую библиотеку (ВЦБ) – уникальный проект, реализованный совместными усилиями Познанского суперкомпьютерного центра, Института проблем информатики и автоматизации Национальной академии наук Республики Армении и Фундаментальной научной библиотеки Национальной Академии наук Республики Армении (ФНБ НАН РА) с 2018 г. Оцифровка культурного наследия стала неотъемлемым фактором современного информационного общества. Цифровая библиотека сохраняет историческое наследие, делает его доступным для широкой аудитории, в ВЦБ включены статьи академических журналов различных научных сфер. В ФНБ НАН РА статьям научных журналов присваиваются УДК и DOI (Digital Object Identifier), которые важны для повышения видимости научных публикаций. В статье рассказывается также о работе ФНБ НАН по внедрению технологии OCR (оптическое распознавание символов): преобразование текста из отсканированных изображений в редактируемый текст делает научную литературу более доступной для исследователей, так как позволяет использовать функции поиска и копирования текста.

Ключевые слова: Всеармянская цифровая библиотека, Фундаментальная научная библиотека Национальной Академии наук Республики Армении, библиографический поиск, культурное наследие, OCR (оптическое распознавание символов)

Для цитирования: Туманян Р. Р. Всеармянская цифровая библиотека – современная платформа для научных исследований // Книга. Чтение. Медиасреда. 2024. Т. 2, № 3. С. 220–226. <https://doi.org/10.20913/BRM-2-3-8>.

Статья поступила в редакцию 05.05.2024

Получена после доработки 04.11.2024

Принята для публикации 14.11.2024



Туманян Рузанна Рубеновна

Фундаментальная научная библиотека Национальной Академии наук
Республики Армении,

пр. Маршала Баграмяна, 24/6, Ереван, 0019, Армения,

заведующая информационно-библиографическим отделом

ORCID: 0000-0002-3349-6677

e-mail: rtumanyan32@gmail.com

© Р. Р. Туманян, 2024

R. R. Tumanyan

Pan-Armenian Digital Library as a Platform for Scientific Research

Tumanyan Ruzanna R.

Fundamental Scientific Library of the National Academy of Sciences of the Republic of Armenia,
Marshal Baghramyan Ave., 24/6, Yerevan, 0019, Armenia,
Head of Information and Bibliographic Department
ORCID: 0000-0002-3349-6677
e-mail: rtumanyan32@gmail.com

Abstract. The main purpose of the article is to present the Pan-Armenian Digital Library. The project was implemented by the Poznan Supercomputer Centre, the Institute of Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia, the Fundamental Scientific Library. Digitalization of cultural heritage became an integral part of the modern information society. The platform preserves historical heritage and makes it accessible to a wide audience. The Digital library includes articles from academic journals in various scientific fields. The Fundamental Scientific Library of the National Academy of Sciences of the Republic of Armenia assigns UDC (Universal Decimal Classification) and DOI (Digital Object Identifier) to scientific articles. Notably, the library is doing important work to introduce OCR technology into scientific journals. Converting text from scanned images into editable text makes scientific literature more accessible to researchers by providing search and copy functionality.

Keywords: Pan-Armenian Digital Library, Fundamental Scientific Library, Bibliographic search, Cultural Heritage, OCR (Optical Character Recognition)

Citation: Tumanyan R. R. Pan-Armenian Digital Library as a Platform for Scientific Research // Book. Reading. Media. 2024. Vol. 2, No. 3. P. 220–226. <https://doi.org/10.20913/BRM-2-3-8>.

Received 05.05.2024

Revised 04.11.2024

Accepted 14.11.2024

Введение

В конце XX и начале XXI в. развитие информационных технологий заставило библиотечное сообщество внести ключевые изменения в библиотечную работу, пересмотреть ключевые функции библиотек. Информационные технологии сыграли в этом значительную роль, открыв более широкие возможности для библиотек, которые стали создавать цифровые архивы, онлайн-каталоги для обеспечения доступа к информации в электронном формате (Арефьев и др., 2006, с. 311–317; Ципинова, 2021, с. 25). Цифровые библиотеки стали важным средством хранения знаний и управления ими, обладая такими преимуществами, как доступность информации и возможность сохранения уникальных и редких документов, их защиты от физического разрушения и утраты.

В Армении сформированы гибридные – бумажные и цифровые – библиотеки, которые предлагают всевозможные варианты поиска и чтения материалов (Չարգարյան, 2002, էջ 168; Vidal-Gorene, 2023, p. 21–25). Наряду с традиционными библиографическими списками формируются электронные базы данных (БД), которые своим содержанием и гибкой системой поиска открывают библиотекарям и читателям новые возможности. Такие репозитории

представляют собой уникальные библиографические списки с цифровыми версиями, широкими возможностями и различными способами доставки. При создании подобных БД необходимо учитывать их цель, содержание, программное обеспечение, технологическую насыщенность, соблюдение авторских прав (Գրիգորյան, 2022, էջ 98–107).

Фундаментальная научная библиотека Национальной Академии наук Республики Армении (ФНБ НАН РА) является крупнейшим научно-информационным учреждением по объему фондов, возможностям внедрения и использования информационных технологий, наличию профессиональных кадров. Библиотека оказывает услуги населению в соответствии с международными стандартами. ФНБ НАН РА стала одним из ведущих разработчиков цифровых версий научных изданий и различных модулей электронной библиотеки, что дает возможность на современном уровне предоставлять общественности богатые разнообразные коллекции научной литературы и арменоведение (Zargaryan, Astsatryan and Matela, 2020, p. 68–79). ФНБ НАН РА сегодня предлагает исследователям

- библиографический поиск материала;
- документы для чтения в электронной и бумажной версиях;

- предоставление письменных рекомендаций;
- электронное оформление заказа;
- присвоение научным статьям DOI и УДК;
- размещение научных статей в интернете и доступ к ним.

Цель статьи – описать структуру, содержание и возможности Всеармянской цифровой библиотеки (ВЦБ), координируемой ФНБ НАН РА. Основной проблемой для ВЦБ стало программное обеспечение, поскольку в Армении используется ряд программ, которые предназначены скорее для создания электронной картотеки или архивирования цифровых материалов. Между тем программное обеспечение, лежащее в основе ВЦБ, предназначено для создания цифровой библиотеки. Представление содержания цифровой библиотеки также является важным вопросом, поскольку ни одна БД сегодня не может собрать воедино все виды рукописных и печатных материалов.

Всеармянская цифровая библиотека

Всеармянская цифровая библиотека организована при активной поддержке Познанского суперкомпьютерного центра. Техническое обслуживание ВЦБ осуществляет Институт проблем информатики и автоматизации НАН РА, а за содержание отвечает ФНБ НАН РА (рис. 1).

ФНБ НАН РА – самое крупное научно-информационное учреждение в республике, обладающее значительными объемами научных коллекций. Благодаря полученным возможностям, ВЦБ сегодня – лучшая цифровая платформа для развития науки и образования в Армении.

Первоначально в ФНБ НАН РА работали с программой ePrint, что позволило накопить ценный опыт в управлении цифровыми ресурсами и организации цифровых коллекций, научиться эффективно структурировать и каталогизировать цифровые документы; сотрудники приобрели навыки в продвижении научных публикаций;

работа с интерфейсом ePrint помогла лучше понять потребности пользователей и дала возможность улучшить сервисы и услуги библиотеки. Однако ePrint имела и свои недостатки, в частности, не интегрировалась с другими системами или БД, что затрудняло поиск информации из разных источников. Пользователи вынуждены были искать информацию по каждому журналу отдельно, что увеличивало время поиска нужных материалов. Эти плюсы и минусы дали ФНБ НАН РА возможность понять, как оптимизировать предоставление услуг, и рассмотреть альтернативные решения для управления цифровыми ресурсами.

Всеармянская цифровая библиотека, состоящая на сегодняшний день из 260 тыс. цифровых объектов, отражает огромный объем работы, проведенной командой ФНБ НАН РА по созданию и пополнению коллекций. В отличие от других цифровых библиотек Армении, работающих только с печатными материалами, ВЦБ дает возможность собрать все рукописные и печатные документы.

В академических и научных учреждениях Армении используют системы DSpace, но важно отметить, что существует также Dlibra, и эти две разные системы для управления цифровыми библиотеками отличаются по функциональности. DSpace предназначена для создания институциональных репозиторий, которые позволяют университетам и исследовательским учреждениям управлять научными публикациями, диссертациями, исследовательскими данными и другими академическими материалами и хранить их, тогда как Dlibra – это система цифровой библиотеки с открытым исходным кодом, позволяющая быстро находить нужные документы по метаданным, ключевым словам и содержанию

Основные функции Dlibra:

- индексация и поиск, позволяющие осуществлять быстрый поиск по содержанию, включая текстовые документы и библиографические записи;

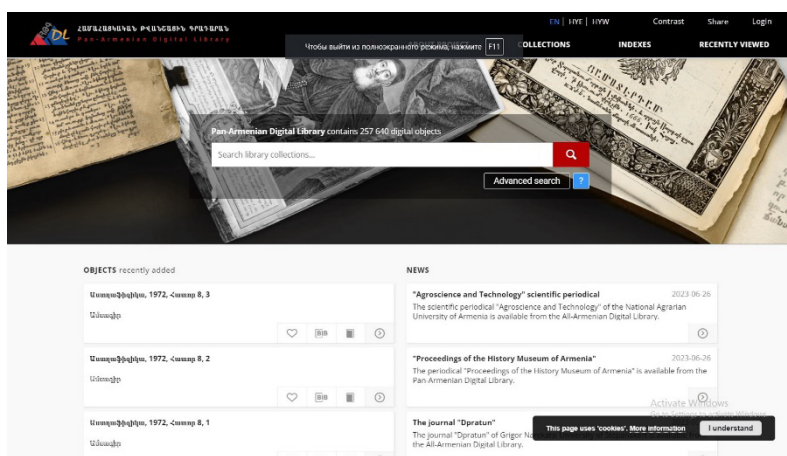


Рис. 1. Главная страница Всеармянской цифровой библиотеки <https://arar.sci.am/dlibra>

Fig. 1. The Main Page of the Pan-Armenian Digital library <https://arar.sci.am/dlibra>

- поддержка разных форматов файлов, что делает ее универсальной для работы с различными типами контента;
- эффективное управление метаданными, что упрощает каталогизацию и организацию материалов;
- интеграция с внешними поисковыми системами, что позволяет повысить доступность контента.

На основе Dlibra работает ВЦБ, коллекции которой представляют собой библиографический поиск с открытым доступом научных публикаций, книг, журналов, религиозных, учебных, архивных материалов, которые становятся доступными для пользователей со всего мира. В обеспечении доступности и удобства ее использования она дает выбор языка в интерфейсе ВЦБ (английский, восточноармянский, западноармянский), позволяющем привлечь международную аудиторию. ВЦБ имеет возможность выполнять расширенный поиск по полям метаданных в своей системе, которая является важным инструментом для пользователей, ищущих конкретные публикации или информацию (рис. 2). Этот функционал предоставляет дополнительные возможности точного и быстрого поиска, делая его более эффективным и удобным. Например, читатель может искать публикации по автору, году выхода, тому, номеру выпуска, названию статьи, журнала и другим характеристикам.

В коллекциях ВЦБ содержатся оцифрованные копии армянских книг, вышедших в свет начиная с 1512 г. (рис. 3): редкие издания и современные публикации, отражающие новейшие армянские исследования и литературу. Значимую часть библиотеки занимают копии статей, опубликованных в армянских научных журналах, сборниках научных

трудов НАН РА, а также представляющие огромную ценность для исследователей документы конференций и библиография академиков Армении.

Помимо изданий НАН РА, в ВЦБ включены печатные документы и издания различных научных, образовательных, культурных структур Армении, Арцаха¹ и диаспоры. Представлены сборники периодических и газетных изданий 1794–1920 гг. из венской монастырской библиотеки ордена мхитаристов (рис. 4).

Издания того времени представляют собой ценный источник информации для всех интересующихся армянской наукой, культурой и историей, в том числе и для армянской диаспоры, играющей огромную роль в сохранении культурного наследия армянского народа. В отличие от традиционных бумажных изданий, которые могут быть утрачены или повреждены со временем, цифровые копии обеспечивают долгосрочное хранение и доступность материалов для будущих поколений.

Не так давно библиотека пополнилась новой коллекцией – «Цифровая репатриация», которая представляет собой библиографические описания публикаций, отсутствующих в библиотеках Армении и переданных в дар ВЦБ со всего мира. Коллекция включает книги, периодические издания, рукописи, карты, фотографии и другие публикации, которые имеют историческое, культурное или научное значение для армянской общины (рис. 5).

Фундаментальная научная библиотека НАН РА также оказывает услуги находящимся за пределами страны исследователям, которые обращаются с письменным запросом о предоставлении книги в онлайн-формате, помогает пользователям в поиске, составлении точных и эффективных запросов.

Рис. 2. Поля расширенного поиска <https://arar.sci.am/dlibra/advsearch>

Fig. 2. Advanced Search Option <https://arar.sci.am/dlibra/advsearch>

¹ Арцах, также Малый Сюник, Глубинная Армения, Орхистена – историческая область на территории современного Азербайджана.

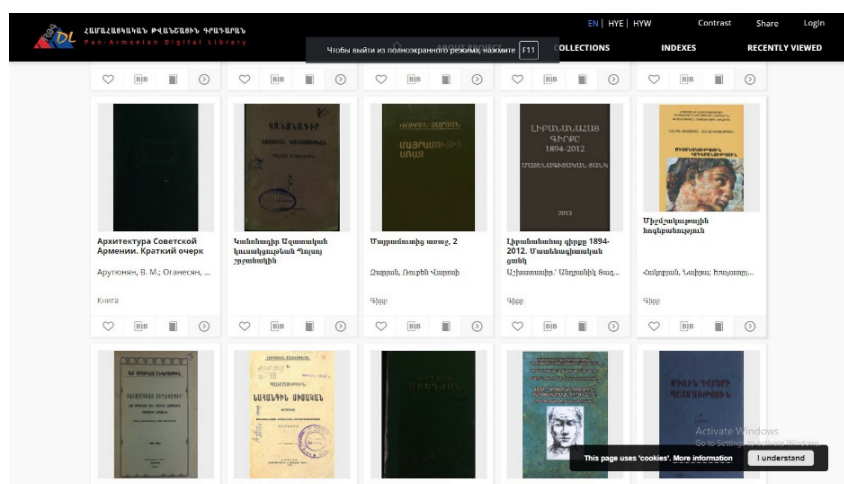


Рис. 3. Коллекция армянских книг
Fig. 3. Collection of Armenian Books

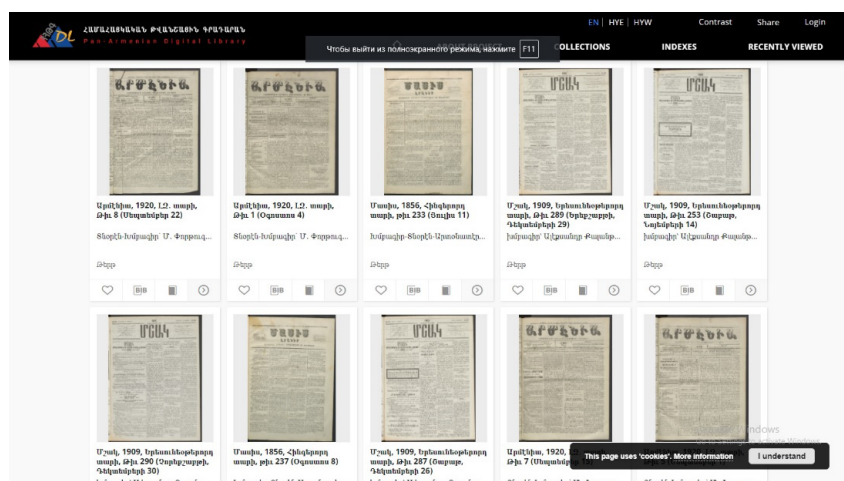


Рис. 4. Коллекция газетных изданий библиотеки ордена мхитаристов
Fig. 4. Collection of Newspapers

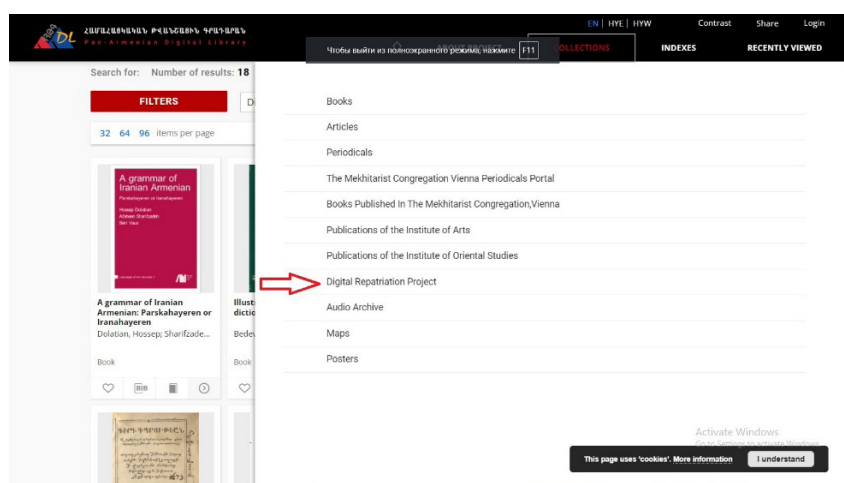


Рис. 5. Коллекции цифровой библиотеки
Fig. 5. Digital Library Collections

Важную работу ФНБ НАН РА осуществляет по внедрению технологии OCR (оптическое распознавание символов). Преобразование отсканированных изображений в редактируемый текст делает научную литературу более доступной для исследователей, так как при этом используется поиск по ключевым словам и копирование текста. Распознанный текст из научных журналов может быть легко интегрирован с системами искусственного интеллекта для выполнения различных задач.

В дальнейшем внимание ФНБ НАН РА будет уделено созданным компанией «Калфа» технологиям для оцифровки и автоматизации процессов, связанных с документами. «Калфа» разработала искусственный интеллект для арабского, армянского, грузинского, сирийского и других языков. Эта технология основана на адаптированных для работы с древними материалами методах сканирования и обработки изображений.

Конвертация оцифрованного документа в текст с использованием технологий оптического распознавания символов (англ. optical character recognition – OCR) для армянского языка представляет собой сложную задачу. Армянский алфавит имеет уникальные графемы и свои особенности, что затрудняет разработку эффективных алгоритмов распознавания по сравнению с латинскими и кириллическими системами. Важно продолжать работу над улучшением технологий OCR для армянского языка, чтобы обеспечить высокую точность распознавания. Компании «Калфа» удалось достичь результата в 99 % случаев распознавания символов армянского языка – и это значительный успех (рис. 6).

В ФНБ НАН РА статьям научных журналов присваивается определитель УДК для обозначения тематической принадлежности статьи и DOI (Digital Object Identifier). Это является важным шагом для улучшения качества и усиления видимости научных

публикаций, обеспечивает стабильность ссылок и устойчивость к изменениям в публикации, в том числе при индексировании публикаций в ведущих научных БД, таких как Scopus и Web of Science.

Заключение

Фундаментальная научная библиотека НАН РА сегодня является платформой, на которой внедрены различные инструменты и разработаны системы, позволяющие быстро и легко доставить необходимый материал исследователю. Усилиями сотрудников ФНБ НАН РА создана ВЦБ, которая играет огромную роль в современном информационном обществе, предоставляя доступ к обширным коллекциям научных материалов. Этот репозиторий доступен в онлайн-формате и постоянно пополняется, поэтому количество онлайн-читателей, пользующихся ВЦБ, растет с каждым днем. Так, за три месяца (август – октябрь) в ФНБ НАН РА обратилось 66 613 читателей, что позволяет сделать следующие выводы: оцифрованные документы облегчают доступ к важным исследованиям и публикациям, переход на цифровые форматы позволяет ученым получать доступ к материалам из любой точки, что особенно актуально в условиях глобальной науки и исследований. Эти достижения демонстрируют, что цифровая библиотека может играть важную роль в поддержке науки и образования, а также в сохранении культурного наследия.

Автор прочитал и одобрил окончательный вариант рукописи.

Конфликт интересов

Автор заявляет об отсутствии конфликтов интересов, требующих раскрытия в этой статье.

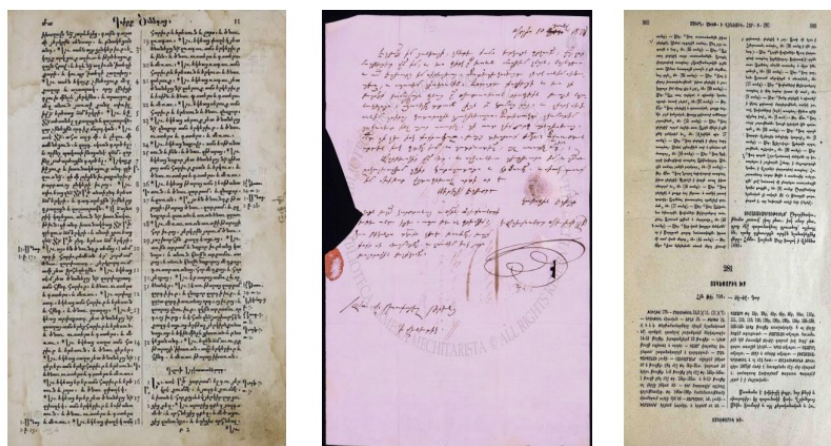


Рис. 6. Конвертированный с помощью OCR и оцифрованный документ
Fig. 6. Converting a Digitized Document (OCR)

Список источников / References

Գրիգորյան Մ. Էլեկտրոնային շտեմարանները՝ գրադարանների զարգացման հարթակ // Մրցունակ կրթության և գիտության հիմնախնդիրները հարափոփոխ աշխարհում : միջազգային գիտաժողովի նյութերի ժողովածու. Երևան : Հյուսիսային համալսարան, 2022. 336 էջ. [Grigoryan M (2022) Electronic databases as a platform for library development. Issues of competitive education and science in the ever-changing world: proc. of the intern. sci. conf. Yerevan: Northern Univ. (In Arm.)].

Զարգարյան Տ. Թվային գրադարաններ. Երևան, 2002. 168 էջ. [Zargaryan T (2002) Digital libraries. Yerevan. (In Arm.)].

Арефьев М. Ю., Дмитрученко П. Г., Коняшкина Е. С., Трегубов А. В. Использование современных информационных технологий в библиотечном деле (на примере библиотеки ИУБиП) // Образовательные технологии и общество : электрон. журн. 2006. Т. 9, № 2. С. 311–317 [Arefiev MYu, Dmitruchenko PG, Konyashkina ES and Tregubov AB (2006) The use of modern information

technologies in librarianship (using the example of the library of the Institute of Management, Business and Law). *Obrazovatel'nye tekhnologii i obshchestvo: e-journal* 9 (2): 311–317. (In Russ.)). URL: <https://sciup.org/14062088> (дата обращения = accessed 30.09.2024).

Ципинова Н. М. Онлайн-сервис современных библиотек : метод. пособие. Майкоп : Нац. б-ка Респ. Адыгея, 2021. 25 с. [Tsipinova NM (2021) Online service of modern libraries: methodological manual. Maykop: Nats. b-ka Resp. Adygeya. (in Russ.)].

Vidal-Gorene C (2023) OCR/HTR technologies and Armenian heritage preservation. *Bulletin of Armenian libraries*, spec. iss.: 61–65. DOI: <https://doi.org/10.52027/18294685-cvo2023.sp>.

Zargaryan T, Astsatryan H and Matela M (2020) Armenian Research & Academic Repository in action: towards challenges of the 21st century. *Katchar: collection of sci. art.* Yerevan, no. 2, pp. 67–79.