

АВТОМАТИЗАЦИЯ ПРОЦЕССОВ ОБРАБОТКИ ДАННЫХ ИНФОРМАЦИОННЫХ КАРТ СОХРАННОСТИ РЕДКИХ И ЦЕННЫХ ИЗДАНИЙ

Коробковский В.А.¹

Научный руководитель – канд. техн. наук, доцент Горлушкина Н.Н.¹

¹Университет ИТМО

vkorobkovskiy@gmail.com

Введение

В условиях цифровизации библиотечных фондов Библиотеки Российской академии наук возрастает важность контроля за состоянием редких и ценных изданий, даже несмотря на то, что доступ к ним ограничен. Для фиксации результатов обследования экземпляров используются информационные карты сохранности, содержащие табличные формы и систему отметок. В настоящее время обработка таких карт и перенос данных в АБИС ИРБИС 64+ [1] выполняются вручную, что требует значительного времени и повышает вероятность ошибок. На основании этого задача автоматизации извлечения и интерпретации данных из карт сохранности на основе методов анализа изображений и компьютерного зрения является актуальной. Целью работы является исследование подходов к решению данной задачи и разработка алгоритма автоматизированной обработке информационных карт сохранности для последующей загрузки полученных результатов в библиотечную систему.

Основная часть

В рамках данной работы в качестве объектов исследования выступают информационные карты сохранности редких и ценных изданий, который используются в Библиотеке Российской академии наук. Они представляют собой двусторонние бланки формата А4, на которых расположены табличные формы, каждая из которых отвечает за соответствующий блок, отражающий состояние экземпляра. Подход к их обработке развивает ранее полученные мною результаты по автоматизации обработки библиографических карточек и переносу данных в библиотечные информационные системы [2, 3] Так как эти данные собираются ещё с 1953 года, то за такой длительный промежуток времени карты претерпели множество изменений, касающиеся формы и структуры, которые необходимо учитывать. В работе был проведён анализ структуры карт и вариативности их оформления, а также рассмотрены два подхода к решению данной задачи: нейросетевой и алгоритмический.

Нейросетевой подход на основе детекции чекбоксов с использованием архитектуры модели YOLO доказал возможность практического применения, однако он требует большего количества размеченных данных и значительных трудозатрат для сбора обучающей выборки, особенно для оборотной стороны карточки.

В качестве основного был выбран подход с использованием классических методов компьютерного зрения [4, 5]. Был реализован алгоритм, который позволяет выравнивать изображение, выделять нужные блоки на обеих страницах, будь то отдельные таблицы или одна общая, сегментировать строки и ячейки, проводить проверку количества горизонтальных и вертикальных линий на каждом этапе сегментации и определять наличие различных отметок (галочек, крестиков, плюсики и так далее) в нужных ячейках.

Дополнительно была сформирована схема соответствия ячеек кодам полей, что позволило реализовать это в программном коде и автоматически формировать структурированный текстовый файл для последующей загрузки в библиотечную

систему.

Выводы

Результатом работы стали исследование, которое позволило выбрать наиболее оптимальные методы для решения задачи, и реализованный алгоритм, который позволяет автоматизировать процесс переноса информации с бумажного носителя в информационную систему. Полученные результаты подтверждают работоспособность подхода и оставляют возможность для улучшения, например, расширения на остальные блоки информационных карт сохранности и обработки наиболее старых типов карт, структура которых серьезно отличается от более современных.

Литература

1. Система ИРБИС64+ [Электронный ресурс]. — URL: https://elnit.org/index.php?option=com_content&view=article&id=255:irbis64&catid=18:kharakteristiki-produktov (дата обращения: 04.02.2025).
2. Коробковский В.А. (науч. рук. Горлушкина Н.Н.) Разработка программы для автоматизации процесса обработки информации библиографических карточек // Сборник тезисов докладов конгресса молодых ученых. Электронное издание. – СПб: Университет ИТМО, [2024]. [Электронный ресурс]. – Режим доступа: <https://kmu.itmo.ru/digests/article/12882>.
3. Коробковский В.А., Горлушкина Н.Н., Белинская М.А. Разработка алгоритма автоматизации ретроконверсии для создания электронного каталога. *Научные и технические библиотеки*. 2025;(2).
4. Shlok Nandurbarkar, Archana Chaudhari, Rahesha Mulla et al. SPARTAN – Automated Table Detection and Extraction from Documents using Advanced OpenCV Heuristics and OCR Techniques, 17 July 2025, PREPRINT (Version 1). <https://doi.org/10.21203/rs.3.rs-6644838/v1>.
5. Sreekiran A.R. Checkbox/Table cell detection using OpenCV-Python, 22 November 2020. [Электронный ресурс]. – Режим доступа: <https://towardsdatascience.com/checkbox-table-cell-detection-using-opencv-python-332c57d25171/>.

Коробковский В.А. (автор)  _____

Горлушкина Н.Н. (научный руководитель)  _____