

УДК 004.932.75'1

РАЗРАБОТКА СЕРВИСА РАСПОЗНАВАНИЯ СТРУКТУРИРОВАННЫХ ДАННЫХ ТИТУЛЬНЫХ ЛИСТОВ ВСЕРОССИЙСКИХ ПРОВЕРОЧНЫХ РАБОТ С ПРИМЕНЕНИЕМ НЕЙРОСЕТЕВЫХ МЕТОДОВ

Смирнов Д.А. (ИТМО) Сагайдак А.А. (ИТМО)

Научный руководитель - кандидат технических наук, доцент Федоров Д.А. (ИТМО)

Введение. В работе рассматривается развитие ранее разработанного сервиса автоматизированного распознавания структурированных данных титульных листов всеоссийских проверочных работ (ВПР). Сервис на первом этапе апробации в школах обеспечивал извлечение печатных, рукописных и табличных данных с приемлемой точностью, однако в процессе эксплуатации был выявлен ряд недостатков, которые влияли на устойчивость распознавания и гибкость использования сервиса. На текущем этапе производится совершенствование системы, устранение выявленных недостатков и повышение точности и надежности распознавания [1].

Основная часть. Исходный сервис реализовывал последовательность этапов обработки, включающих предобработку изображения, выделение областей документа, распознавание печатного текста и рукописных цифр, извлечение данных из таблицы. После апробации выяснилось, что при сохранении общей работоспособности качество распознавания снижается на некачественных сканах и при нестандартном рукописном вводе, а ошибки на ранних этапах приводят к порче итогового результата [2]. Для устранения выявленных недостатков сначала была выполнена переработка архитектуры: выделены независимые модули и внедрен централизованный механизм конфигурирования параметров распознавания. Это значительно упростило дальнейший процесс модификации сервиса и позволило настраивать ключевые пороги и правила обработки без изменения программного кода. Повышение надежности было достигнуто за счет логической поствалидации результатов, а также дополнительных проверок на основе использования оценок уверенности исходной модели так, чтобы результаты с низкой степенью уверенности распознавания отмечались для возможной отправки на дополнительную проверку. Для этого был реализован механизм передачи спорного изображения на вход мультимодальной языковой модели с использованием соответствующего промта, ограничивающего доменную область распознавания [3]. Такой подход позволил повысить точность распознавания в нестандартных случаях, однако увеличил вычислительные затраты и потребовал дополнительных вычислительных ресурсов, поэтому был установлен опциональным при развертывании сервиса. Таким образом, данный подход может быть использован как дополнительная мера для повышения общей точности работы сервиса [4]. Также для повышения устойчивости выделения и обработки структурных элементов документа при изменениях оформления представляется целесообразным применение современных методов анализа макета документа, основанных на моделях детекции и структурной разметки [5].

Выводы. В ходе работы устранены выявленные при первичной апробации недостатки исходного сервиса. Архитектурная переработка и внедрение конфигурирования повысили гибкость и удобство сопровождения. Использование оценок уверенности и выборочное применение мультимодальной языковой модели для распознавания отдельных цифр повысили точность в сложных случаях, однако требуют дополнительных вычислительных ресурсов. Намечены направления дальнейшего развития, такие как усиление автоматической валидации, повышение устойчивости к изменениям форматов бланков и оптимизация вычислительных затрат.

Список использованных источников:

1. Nishant S., Alexandre M., Malcolm G., Adrian L. A Survey of Deep Learning Approaches for OCR and Document Understanding // arXiv. 2021.
2. Guo C., Pleiss G., Sun Y., Weinberger K.Q. On Calibration of Modern Neural Networks // ICML. 2017.
3. Qwen Team. Qwen2.5-VL: A Series of Multimodal Large Language Models // arXiv. 2024.
4. Tang Z. et al. Unifying Vision, Text, and Layout for Universal Document Processing // CVPR. 2023.
5. Zhao Z., Kang H., Wang B., He C. DocLayout-YOLO: Enhancing Document Layout Analysis // ICLR. 2024.