

ОТЗЫВ

на автореферат диссертационной работы
Смирнова Сергея Владимировича
«Технология и система автоматической корректировки результатов
при распознавании архивных документов»,
представленной на соискание ученой степени кандидата технических наук по
специальности 05.13.11 — Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей.

Как следует из автореферата, диссертационная работа Смирнова Сергея Владимировича направлена на повышение эффективности оптического распознавания оцифрованных архивных документов, путем корректировки ошибок и оптимизации всего процесса обработки с помощью разработанных технологии и системы. Актуальность работы обусловлена интенсивным развитием в наши дни проектов по переводу бумажных фондов архивов, музеев, библиотек, в результате чего формируются большие базы электронных образов, требующие последующего распознавания и обеспечения возможностей эффективного поиска по ним. В описании первой главы отмечается, что поскольку существующие OCR системы нацелены, как правило, на обработку современных качественных изображений, то они допускают большой процент ошибок при распознавании архивных документов, изобилующих узкоспециализированной терминологией и обладающих более низким исходным качеством.

Наиболее существенным результатом диссертационного исследования с точки зрения вклада в науку является разработанный метод автоматической корректировки результатов распознавания на основе рейтинго-ранговых распределений, генерируемых путем статистического анализа всего корпуса распознанных материалов. Среди прочих подходов разработанный метод выделяется своим процессом автоматического формирования тезаурусов, словарей, хэш-таблиц и других структур, используемых для обнаружения ошибок, поиска кандидатов на их замену, ранжирования и выбора наилучших корректировок.

Все теоретические результаты диссертационной работы нашли свою реализацию в разработанной системе распознавания архивных документов с автоматической корректировкой ошибок, которая была испытана на практике в составе государственной информационной системы «Государственные архивы Санкт-Петербурга» при обработке фондов центральных государственных архивов Санкт-Петербурга.

Приведенная в четвертой главе оценка распознавания корпуса архивных документов размером до 800 тысяч образов наглядно демонстрирует практическую значимость, обоснованность и достоверность результатов диссертационной работы.

По тексту автореферата можно сделать следующие замечания.

1. В описании третьей главы говорится о критериях оценки качества распознавания, используемых экспертом для настройки профилей распознавания и корректировки, однако, не приводится ни их описания, ни их перечня, что не позволяет до конца разобраться в разработанном функционале.
2. В автореферате на рисунке 8 «Сравнение оценок полноты и точности до и после корректировки» отсутствуют пояснения обозначений на горизонтальных осях столбиковых диаграмм. Лишь их сопоставление с рисунком 7 позволяет понять, что они определяют группы изображений, соответствующих определенному диапазону точности.

Отмеченные недостатки имеют частный характер и не снижают научной и практической значимости проведенных исследований. Автореферат написан ясным и четким языком.

В целом диссертационная работа, насколько можно судить по автореферату, имеет важное научное и практическое значение, соответствует требованиям «Положения о порядке присуждения ученых степеней», предъявляемым ВАК РФ к кандидатским диссертациям по специальности 05.13.11, а ее автор Смирнов Сергей Владимирович заслуживает присвоение искомой ученой степени кандидата технических наук.

Заведующий кафедрой
теории вероятностей и анализа данных
Федерального государственного бюджетного
образовательного учреждения высшего
профессионального образования
«Петрозаводский государственный университет»
доктор технических наук, профессор

« 28 » апреля 2015 г.

Рогов Александр Александрович



Подпись руки проф. Рогов А.А.
ДОСТОВЕРЯЮ.
секретарь ученого совета Бурмалов А.И.
2015 г.