

**ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.1.120.01,
созданного на базе
Федерального государственного бюджетного учреждения науки
Институт системного программирования им. В.П. Иванникова
Российской академии наук
Министерства науки и высшего образования РФ
по диссертации на соискание ученой степени кандидата наук**

аттестационное дело № _____

решение диссертационного совета от 23 апреля 2026 года № 2026/07

О присуждении Беляевой Оксане Владимировне, гражданке РФ, ученой степени кандидата технических наук.

Диссертация «Автоматическое восстановление структуры текстовых документов» по специальности 2.3.5 – «математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» принята к защите 18 февраля 2026 года (протокол № 2026/04) диссертационным советом 24.1.120.01, созданным на базе Федерального государственного бюджетного учреждения науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ведомственная принадлежность: Министерство науки и высшего образования РФ; адрес: 109004, г. Москва, ул. А. Солженицына, дом 25), создан Приказом Минобрнауки России о советах по защите докторских и кандидатских диссертаций от 2 ноября 2012 г. № 714/нк.

Соискатель Беляева Оксана Владимировна, 26.08.1994 года рождения.

В 2018 году соискатель окончила Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», специальность 09.04.04 «Программная инженерия». В 2024 году соискатель окончила аспирантуру Федерального государственного

бюджетного учреждения науки Институт системного программирования им. В.П.Иванникова Российской академии наук.

Работает научным сотрудником в Федеральном государственном бюджетном учреждении науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ведомственная принадлежность: Министерство науки и высшего образования РФ).

Диссертация выполнена в отделе “Информационных систем” Федерального государственного бюджетного учреждения науки Институт системного программирования им. В.П. Иванникова Российской академии наук (ведомственная принадлежность: Министерство науки и высшего образования РФ).

Научный руководитель – кандидат физико-математических наук Турдаков Денис Юрьевич, Федеральное государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук, Центр доверенного искусственного интеллекта ИСП РАН, заведующий центра.

Официальные оппоненты:

1. Котельников Евгений Вячеславович, доктор технических наук, доцент, профессор Школы вычислительных социальных наук автономной некоммерческой образовательной организации высшего образования «Европейский университет в Санкт-Петербурге»;
2. Шигаров Алексей Олегович, кандидат технических наук, Федеральное государственное бюджетное учреждение науки «Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук», ведущий научный сотрудник лаборатории комплексных информационных систем.

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО», Санкт-Петербург, в своем положительном отзыве, подписанном Бухановским Александром

Валерьевичем, доктором технических наук, научным руководителем исследовательского центра в сфере искусственного интеллекта “Сильный искусственный интеллект в промышленности”, указала, что указала, что диссертационная работа содержит новые научные результаты, имеющие существенное значение для науки и практики.

Соискатель имеет 11 опубликованных работ, в том числе по теме диссертации опубликовано 10 работ, из них в рецензируемых научных изданиях опубликовано 8 работ.

Научные работы соискателя направлены на исследования в области интеллектуальной обработки электронных документов и создания их инструментов автоматической обработки. Исследования включают применения методов машинного обучения в решении задач извлечения содержимого и восстановления структуры документов. Три научные работы соискателя индексируются в Scopus и Web of Science, 5 работ опубликованы в журналах, рекомендованных ВАК. Все опубликованные работы выполнены соискателем или при его непосредственном участии.

Наиболее значимые работы:

1. Golodkov A.O. Real Application of CNN Interpretation Methods: Document Image Classification Model Errors' Detection and Validation / Golodkov A.O., Belyaeva O.V., Perminov A.I. // Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). — 2023. — Vol 35. — P. 7-18. — (BAK).
2. Bogatenkova A. O. A.I. Logical structure extraction from scanned documents / Bogatenkova A. O. Kozlov I. S., Belyaeva O. V., Perminov // Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). — 2020. — Vol 32. — P. 175-188. — (BAK).
3. Belyaeva O.V. Synthetic data usage for document segmentation models fine-tuning / Belyaeva O.V., Perminov A.I., Kozlov I.S. // Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). — 2020. — Vol 32. — P. 189-202. — (BAK).
4. Perminov A.I. Loss functions for train document image segmentation models / Perminov A.I., Turdakov D.Yu., Belyaeva O.V. // Programming and Computer Software. – 2023. — Vol 49. — P. 574-589. — (BAK, WoS).

5. A. O. Bogatenkova. Generation of Images with Handwritten Text in Russian / A. O. Bogatenkova, O. V. Belyaeva, A. I. Perminov // Programming and Computer Software. — 2024. — Vol 50. — P. 483-492. — (ВАК, Scopus).

Выбор официальных оппонентов и ведущей организации обосновывается их компетентностью и достижениями в данной отрасли науки, наличием публикаций в сфере исследований, соответствующей теме диссертации, и способностью определить научную и практическую ценность диссертации.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

- разработана новая научная идея автоматической обработки электронных текстовых документов, позволяющая расширить границы применимости восстановления структуры документов разных типов и форматов;
- предложены оригинальные суждения по тематике автоматической обработки документов;
- доказана перспективность использования новых идей как в науке, путем разработки новых методов обработки, так и в практике, за счет применимости разработанного расширяемого программного комплекса в разных сферах деятельности человека;
- введены новые термины и предложены изменения трактовки старых понятий.

Теоретическая значимость исследования обоснована тем, что:

- применительно к проблематике диссертации результативно и корректно использован комплекс существующих базовых методов исследования, в т.ч. численных методов, математического аппарата и экспериментальных методик;
- изложены идеи и основные этапы автоматической обработки текстовых документов, рассмотрены различные условия обработки и факторы, влияющие на данный процесс;
- раскрыты существующие проявления теории: выявлены новые проблемы автоматической обработки PDF документов;
- изучены причины возникновения некорректности PDF документов;

- проведена модернизация существующих алгоритмов, обеспечивающих получения новых и лучших результатов в области по теме диссертации.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

- разработаны и внедрены результаты диссертационного исследования в ряд коммерческих и некоммерческих организаций в виде разработанного программного комплекса, содержащего результаты исследования;
- определены пределы и перспективы практического использования полученных результатов;
- создана расширяемая система автоматической обработки текстовых документов разных форматов и типов;
- представлены предложения по расширению применимости разработанных решений на разные предметные области.

Оценка достоверности результатов исследования выявила:

- показана воспроизводимость результатов исследования для разных предметных областей документов;
- теория построена на известных, проверяемых данных, фактах и согласуется с опубликованными экспериментальными данными диссертации или других работ, связанных с темой диссертации;
- идея базируется на анализе практических задач и реальных данных, встречающихся на практике;
- установлено качественное и количественное совпадение авторских результатов с результатами, представленными в независимых источниках по тематике диссертации;
- использованы современные методики сбора, генерации и обработки исходных данных, показаны представительные выборочные совокупности с обоснованием выбора объектов.

Личный вклад соискателя состоит в непосредственном участии на всех этапах процесса создания диссертационной работы, а также в участии в

получении исходных данных и апробации результатов диссертационного исследования. Лично автором и при его непосредственном участии проводилось построение экспериментов и разработка на их основе экспериментального программного комплекса, обработка данных и интерпретация экспериментальных данных. Автором лично проводилась подготовка всех публикаций по теме диссертации.

В ходе защиты диссертации были высказаны следующие критические замечания:

1) Для классификации строк в методе автоматического восстановления иерархической структуры выбрана библиотека XGBoost, при этом отсутствует сравнение с другими моделями машинного обучения;

2) Метод определения корректности текстового слоя в PDF документах обучался и тестировался на русских и английских текстах, при этом в тексте не раскрыта его применимость и расширяемость для работы с документами на других языках:

3) Предлагаемый метод извлечения содержимого PDF-документа сначала оценивает «корректность» всего документа в целом, а затем принимает решение, каким способом обрабатывать все страницы документа кроме первой: (1) через растеризацию PDF (с OCR) или (2) напрямую через инструкции отрисовки PDF (без OCR). Однако, в общем случае, внутри PDF-документа (более того внутри одной страницы) может быть представлено разнородное содержимое: текст, вставленный в виде растра, и текст, выводимый инструкциями отрисовки (со встроенными CMAP и стандартными кодировками). Вероятно, что в некоторых случаях предпочтительнее было бы оценивать «корректность» и соответственно выбирать способ обработки каждой страницы в отдельности, а не всего документа в целом. В работе следовало бы рассмотреть ограничения предлагаемого метода в явной форме, в особенности те случаи, которые могут повлиять на эффективность и надежность обработки документов.

Соискатель Беляева Оксана Владимировна согласилась с замечаниями, ответила на задаваемые ей в ходе заседания вопросы.

На заседании 23 апреля 2026 года диссертационный совет принял решение за разработку новых научно обоснованных технических решений и разработок, имеющих значение для развития отрасли знаний автоматической обработки текстовых документов, присудить Беляевой Оксане Владимировне ученую степень кандидата технических наук.

При проведении тайного голосования диссертационный совет в количестве 12 человек, из них 7 докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из 15 человек, входящих в состав совета, проголосовали: за – 12, против – 0.

Заместитель председателя диссертационного совета,
доктор физико-математических наук

Петренко А. К.

И.о. ученого секретаря диссертационного совета,
доктор технических наук

Евтушенко Н. В.

23 апреля 2026 года