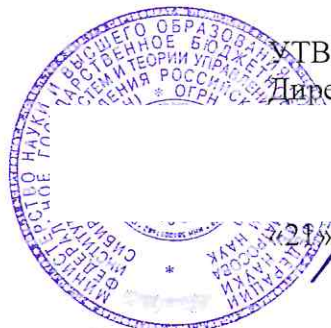


МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное учреждение науки
ИНСТИТУТ ДИНАМИКИ СИСТЕМ И ТЕОРИИ УПРАВЛЕНИЯ
имени В.М. МАТРОСОВА
Сибирского отделения Российской академии наук
(ИДСТУ СО РАН)



УТВЕРЖДАЮ
Директор ИДСТУ СО РАН

И.В. Бычков

«21» мая 2025 г.

ЗАКЛЮЧЕНИЕ

Федерального государственного бюджетного учреждения науки
Института динамики систем и теории управления имени В.М. Матросова
Сибирского отделения Российской академии наук
на диссертацию Шигарова Алексея Олеговича
**«Методы и инструментальные средства автоматизации процессов извлечения данных из
таблиц электронных документов неструктурированного формата»**

Диссертация «Методы и инструментальные средства автоматизации процессов извлечения данных из таблиц электронных документов неструктурированного формата» выполнена в Федеральном государственном бюджетном учреждении науки Институте динамики систем и теории управления имени В.М. Матросова Сибирского отделения Российской академии наук (ИДСТУ СО РАН).

В период подготовки диссертации соискатель **Шигаров Алексей Олегович** работал в ИДСТУ СО РАН в должности ведущего научного сотрудника лаборатории 4.1 Комплексных информационных систем.

В 2004 г. Шигаров А.О. окончил Иркутский государственный университет по специальности «Прикладная математика». В 2010 г. защитил кандидатскую диссертацию по специальности 05.25.05 «Информационные системы и процессы» (технические науки) в диссертационном совете Д 003.046.01 при Институте вычислительных технологий СО РАН, что подтверждается дипломом кандидата наук, регистрационный номер ДКТ № 113218, выданным 11.06.2010 ВАК РФ.

Научный консультант – академик, доктор технических наук Бычков Игорь Вячеславович, директор ИДСТУ СО РАН.

Диссертация Шигарова А.О. посвящена созданию методов и инструментальных средств автоматизации процессов распознавания, анализа и интерпретации документных таблиц, позволяющих упростить разработку прикладного программного обеспечения извлечения данных из документных таблиц с машиночитаемым текстовым содержанием, представленных в неструктурированном виде, за счет поддержки произвольности табличной структуры.

Актуальность темы диссертации. Документные таблицы представляют собой способ хранения реляционных данных в неструктурированных электронных документах, таких как изображения, веб-страницы и электронные книги. Эти таблицы являются важным источником информации для систем поддержки принятия решений, анализа данных и построения баз знаний. Для того чтобы табличная информация могла быть проиндексирована, запрошена и применена в перечисленных приложениях, прежде всего она должна быть приведена к структурированному

представлению (базам данных или графам знаний). В научной литературе проблематику извлечения реляционных данных из таких неструктурированных источников принято называть «автоматическим пониманием таблиц» (далее АПТ).

Сложность АПТ обусловлена разнообразием форматов таблиц и отсутствием формальных моделей для их интерпретации. Являясь частью неструктурированного источника, документная таблица обычно не сопровождается формальной моделью, позволяющей интерпретировать представленную там информацию в соответствии со смыслом, заложенным в нее автором. Обычно неизвестно местоположение документной таблицы внутри источника, структура ячеек, логический порядок чтения данных и пр. В общем случае процесс АПТ включает следующие этапы: распознавание — обнаружение местоположения таблицы и ее ячеек в источнике; анализ — восстановление логического порядка чтения данных; интерпретация — восстановление соответствующего ей набора записей в канонической форме, приведение его к некоторой совокупности отношений в терминах реляционной модели данных и сопоставление их с внешними словарями и схемами. При текущем уровне развития технологий искусственного интеллекта данные процессы в общем случае не могут выполняться без участия человека. Поэтому автоматизация этих процессов нацелена на сокращение операций, производимых человеком.

Современные обзоры тематической литературы показывают, что все известные решения автоматизации извлечения данных из документных таблиц являются частными. Их общим ограничением является отсутствие поддержки произвольности структуры документных таблиц. Известные методы полагаются на типичную компоновку, атомарность ячеек и плоскую структуру заголовков, игнорируя большое количество случаев, когда эти предположения не выполняются. Таким образом, автоматизация процессов извлечения данных из документных таблиц произвольной структуры является актуальной научной проблемой. В частности, ее решение имеет важное хозяйственное значение для массовой обработки таких источников табличной информации, как неотредактируемые печатно-ориентированные документы (далее НПОД), представленные в форматах языков описания страниц (далее PDL): PDF, PostScript и др., а также рабочие книги (далее РК), представленные в форматах табличных процессоров: Excel, Sheets и др.

Научную новизну и значимость составляют следующие основные результаты исследования, выносимые на защиту:

1. Усовершенствована структура совокупности задач АПТ, в которой согласована терминология, сложившаяся в родственных направлениях исследований: компьютерном зрении, управлении данными, информационном поиске и «Семантической паутине». По сравнению с имеющимися формулировками АПТ, она является более релевантной за счет согласованной терминологии и многоуровневой декомпозиции.
2. Предложен метод автоматизации распознавания таблиц НПОД, т. е. конвертирования их в редактируемый формат РК. Впервые данный процесс базируется на использовании PDL-специфичной информации НПОД. Показано, как ее можно задействовать для анализа компоновки страниц, обнаружения и сегментации таблиц НПОД.
3. Создана модель таблицы, обеспечивающая представление табличной информации в процессах АПТ. В отличие от известных аналогов, она не ограничивает структуру таблицы предопределенными типами компоновки, атомарностью ячеек и плоскими заголовками.
4. Предложен метод автоматизации анализа и интерпретации таблиц редактируемого формата РК. Впервые данный процесс реализуется посредством пользовательских правил. Обеспечена поддержка произвольности структуры таблицы: свободной компоновки, структурированности ячеек и иерархичности заголовков.
5. Создан проблемно-ориентированный язык правил анализа и интерпретации таблиц. В отличие от формальных языков общего назначения, он позволяет сфокусироваться исключительно на реализации логики соответствующих этапов АПТ.
6. Разработан комплекс инструментальных средств, реализующий предлагаемые теоретические основы. По сравнению с другими доступными инструментами АПТ он обеспечивает

конвертирование таблиц НПОД в редактируемый формат РК с возможностью пользовательской коррекции результатов и извлечение наборов записей в канонической форме из полученных таблиц РК с помощью правил, предоставляемых пользователями.

Практическая значимость результатов проведенного исследования состоит в том, что в совокупности они составляют теоретические и технологические основы для упрощения разработки прикладного программного обеспечения извлечения данных из документных таблиц неструктурированного формата (НПОД/РК) за счет поддержки произвольности табличной структуры. Разработанные инструментальные средства нашли применение в пяти научных и трех промышленных проектах при решении прикладных задач анализа технических, финансовых и научных документов, интеграции данных государственной статистики и медиапланирования, конструирования онтологии предметной области и кросс-контекстного обмена бизнес-документами.

Основные результаты диссертационного исследования были получены при выполнении следующих научных проектов:

Базовые проекты Программы фундаментальных исследований СО РАН: № IV.38.1.2. «Методы и технологии облачной сервис-ориентированной цифровой платформы сбора, хранения и обработки больших объемов разноформатных междисциплинарных данных и знаний, основанные на применении искусственного интеллекта, модельно-управляемого подхода и машинного обучения» (2020-2021 гг.); № IV.38.1.2 «Методы и технологии создания распределенной сервисно-ориентированной среды сбора, хранения, обработки больших объемов разноформатных междисциплинарных научных данных и знаний, основанные на конструктивных средствах спецификации, порождающем программировании и интеллектуализации» (2017-2019 гг.); № IV.38.2.3 «Новые методы, технологии и сервисы обработки пространственных и тематических данных, основанные на декларативных спецификациях и знаниях» (2013-2015 гг.); № IV.31.2.4 «Методы и технологии разработки программного обеспечения для анализа, обработки и хранения разноформатных междисциплинарных данных и знаний, основанные на применении декларативных спецификаций форматов представления информации и программных систем» (2010-2012 гг.).

Проекты, поддержанные грантами научных фондов: «Методология и инструментальная платформа разработки систем извлечения данных из произвольных электронных таблиц» (РНФ № 18-71-10001, 2018–2023 гг.); «Методы интеграции неструктурированных табличных данных» (РФФИ № 15-37-20042, 2015–2016 гг.); «Методы извлечения табличной информации из неструктурированных текстовых источников» (Совет по грантам Президента РФ № СП-3387.2013.5, 2013–2015 гг.); «Интеллектуальная система извлечения информации из слабоструктурированных таблиц со сложной компоновкой» (РФФИ № 12-07-31051, 2012–2013 гг.); «Моделирование информационной инфраструктуры цифровой экономики Иркутской области» (РФФИ, грант № 17-47-380007, 2017–2019 гг.); «Технология разработки информационно-аналитических систем с функциями пространственного анализа для поддержки управления социально-экономическим развитием территорий» (РФФИ, грант № 16-57-44034, 2016–2017 гг.), а также грантом Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды», № 075-15-2024-533, 2024–2025 гг.).

Реализованное программное обеспечение зарегистрировано в Роспатенте; всего автором в составе коллектива получено 6 свидетельств о регистрации программ для электронных вычислительных машин (ЭВМ). В диссертации представлены письменные подтверждения об использовании разработанного программного обеспечения в промышленной сфере.

Достоверность и обоснованность полученных в диссертации результатов подтверждается корректным использованием методов исследования; работоспособностью разработанного программного обеспечения; представленными экспериментальными данными, качественным и количественным сравнением с имеющимися аналогами, а также программной

реализацией; применением предложенных методов в практических задачах. Разработанные инструментальные средства и материалы для воспроизведения проведенных экспериментов опубликованы в открытом доступе под свободными лицензиями.

Полнота изложения результатов диссертации в печатных работах, опубликованных соискателем, подтверждается следующим перечнем основных работ:

В журналах, рекомендованных ВАК:

1. Шигаров А.О. Извлечение реляционных данных из произвольных таблиц электронных документов редактируемых форматов на основе пользовательских правил // Вычислительные технологии. 2025. Т. 30, № 3.
2. Шигаров А.О. Распознавание таблиц неаннотированных PDF-документов на основе использования PDF-специфичных свойств // Вычислительные технологии. 2024. Т. 29, № 6. № 6(29). С. 125–146.
3. Shigarov A. Table understanding: problem overview // WIREs Data Mining and Knowledge Discovery. 2023. 13(1), e1482. (Q1 SJR)
4. Шигаров А.О., Парамонов В.В. Сегментация текста неразмеченных PDF-документов // Вычислительные технологии. 2022. Т. 27, № 5. С. 69–78.
5. Yurin A., Dorodnykh N., Shigarov A. Semi-automated formalization and representation of the engineering knowledge extracted from spreadsheet data // IEEE Access. 2021. 9, 157468–157481. (Q1 SJR)
6. Shigarov A., Khristyuk V., Mikhailov A. TabbyXL: Software platform for rule-based spreadsheet data extraction and transformation // SoftwareX. 2019. 10, 100270. (Q2 SJR)
7. Shigarov A., Mikhailov A. Rule-based spreadsheet data transformation from arbitrary to relational tables // Information Systems. 2017. 71, 123–136. (Q1 SJR)
8. Shigarov A. Table understanding using a rule engine // Expert Systems with Applications. 2015. 42(2), 929–937. (Q1 SJR)
9. Шигаров А.О., Бычков И.В., Парамонов В.В., Белых П.В. Анализ и интерпретация произвольных таблиц на основе исполнения CRL-правил // Вычислительные технологии. 2015. Т. 20, № 6. С. 87–112.
10. Шигаров А.О. Восстановление логической структуры таблиц из неструктурированных текстов на основе логического вывода // Вычислительные технологии. 2014. Т. 19, № 1. С. 87–99.
11. Шигаров А.О., Бычков И.В., Ружников Г.М. [и др.] Система трансформации таблиц // Информационные технологии и вычислительные системы. 2013. № 3. С. 15–26.
12. Shigarov A., Fedorov R. Simple algorithm page layout analysis // Pattern Recognition and Image Analysis. 2011. 21(2), 324–327.

В прочих рецензируемых журналах:

13. Шигаров А.О., Бычков И.В., Ружников Г.М. [и др.] Проект интеллектуальной системы извлечения табличной информации из неструктурированных текстов // Вестн. Бурят. гос. ун-та. Математика, информатика. 2013. № 9. С. 110–118.
14. Ветров А.А., Фереферов Е.С., Хмельнов А.Е., Шигаров А.О. Формирование хранилища данных для систем MDAtr // Вестн. Бурят. гос. ун-та. Математика, информатика. 2012. № 9. С. 22–28.
15. Шигаров А.О., Федоров Р.К. Алгоритм обнаружения пустого места на странице документа // Современные технологии. Системный анализ. Моделирование. 2011. № 3(31). С. 42–46.
16. Shigarov A., Bychkov I., Ruzhnikov G., Khmel'nov A. A method of table detection in metafiles // Pattern Recognition and Image Analysis. 2009. 19(4), 693–697.
17. Бычков И.В., Ружников Г.М., Хмельнов А.Е., Шигаров А.О. Эвристический метод обнаружения таблиц в разноформатных документах // Вычислительные технологии. 2009. Т. 14, № 2. С. 58–73.
18. Шигаров А.О. Технология извлечения табличной информации из электронных документов разных форматов // Современные технологии. Системный анализ. Моделирование. 2009. № 3(23). С. 97–102.

19. Хмельнов А.Е., Шигаров А.О. Метод извлечения таблиц из неформатированного текста // Вычислительные технологии. 2008. Т. 13, спец. выпуск 1. С. 93–101.
 20. Бычков И.В., Ружников Г.М., Хмельнов А.Е., Шигаров А.О. Метод обнаружения таблиц в метафайлах // Современные технологии. Системный анализ. Моделирование. 2008. Спецвыпуск. С. 47–51.
- В прочих изданиях, входящих в базы цитирования Web of Science и Scopus:*
21. Kostyleva O., Paramonov V., Shigarov A., Vetrova V. Towards comparison of table type taxonomies // Proc. 45th Int. ICT and Electronics Conv. 2022. P. 1461–1465.
 22. Shigarov A., Dorodnykh N., Yurin A. et al. From web-tables to a knowledge graph: prospects of an end-to-end solution // Proc. 4th Scientific-Practical W. on Information Technologies: Algorithms, Models, Systems. 2021. CEUR WS 2984. P. 23–33.
 23. Shigarov A., Dorodnykh N., Mikhailov A. et al. Table extraction, analysis, and interpretation: the current state of the TabbyDOC project // Proc. 4th Scientific-Practical W. on Information Technologies: Algorithms, Models, Systems. 2021. Irkutsk, Russia. CEUR WS 2984. P. 11–22.
 24. Paramonov V., Shigarov A., Vetrova V. Rule driven spreadsheet data extraction from statistical tables: case study // Proc. 27th Int. Conf. on Inf. and Softw. Technol. 2021. CCIS 1486. P. 84–95.
 25. Dorodnykh N., Shigarov A., Yurin A. Using the semantic annotation of web table data for knowledge base construction // Proc. 4th Int. Conf. on Artificial Intelligence and Cloud Computing. 2021. P. 122–129.
 26. Dorodnykh N., Yurin A., Shigarov A., Turdakov D. Ontology engineering at the assertion level based on semantic annotation of tabular data // Proc. 2021 Ivannikov Memorial Workshop. 2021. P. 28–34.
 27. Cherepanov I., Mikhailov A., Shigarov A., Paramonov V. On automated workflow for fine-tuning deep neural network models for table detection in document images // Proc. 43rd Int. ICT and Electronics Conv. 2020. P. 1130–1133.
 28. Mikhailov A., Shigarov A., Rozhkov E., Cherepanov I. On graph-based verification for PDF table detection // Proc. 2020 Ivannikov ISPRAS Open Conf. 2020. P. 91–95.
 29. Paramonov V., Shigarov A., Vetrova V. Table header correction algorithm based on heuristics for improving spreadsheet data extraction // Proc. 26th Int. Conf. on Inf. and Softw. Technol. 2020. CCIS 1283. P. 147–158.
 30. Dorodnykh N., Yurin A., Shigarov A. Conceptual model engineering for industrial safety inspection based on spreadsheet data analysis // Proc. 6th Int. Conf. on Model. and Develop. of Intelli. Sys. 2020. CCIS 1126. P. 51–65.
 31. Paramonov V., Shigarov A., Vetrova V., Mikhailov A. Heuristic algorithm for recovering a physical structure of spreadsheet header // Proc. 40th Int. Conf. on Inf. Sys. Architect. and Technol. 2019. AISC 1050. P. 140–149.
 32. Shigarov A., Khristyuk V., Mikhailov A., Paramonov V. TabbyXL: rule-based spreadsheet data extraction and transformation // Proc. 25th Int. Conf. on Inf. and Softw. Technol. 2019. CCIS 1078. P. 59–75.
 33. Shigarov A., Cherepanov I., Cherkashin E. et al. Towards end-to-end transformation of arbitrary tables from untagged portable documents {(PDF)} to linked data // Proc. 2nd Scientific-practical W. on Information Technologies: Algorithms, Models, Systems. 2019. Irkutsk, Russia. CEUR WS 2463. P. 1–12.
 34. Shigarov A., Khristyuk V., Mikhailov A., Paramonov V. Software development for rule-based spreadsheet data extraction and transformation // Proc. 42nd Int. ICT and Electronics Conv. 2019. P. 1132–1137.
 35. Yang S., Wei R., Shigarov A. Semantic interoperability for electronic business through a novel cross-context semantic document exchange approach // Proc. ACM Symp. on Document Engineering. 2018. V. 28. P. 1–10.
 36. Shigarov A., Altaev A., Mikhailov A. et al. TabbyPDF: Web-based system for PDF table extraction // Proc. 24th Int. Conf. on Inf. and Softw. Technol. 2018. CCIS 920. P. 257–269.

37. Shigarov A., Khristyuk V., Paramonov V. et al. Toward framework for development of spreadsheet data extraction systems // Proc. 1st Scientific-Practical W. on Information Technologies: Algorithms, Models, Systems. 2018. Irkutsk, Russia. CEUR WS 2221. P. 90–96.
38. Shigarov A., Mikhailov A., Altaev A. Configurable table structure recognition in untagged PDF documents // Proc. 2016 ACM Symp. on Document Engineering. 2016. P. 119–122.
39. Shigarov A., Paramonov V., Belykh P., Bondarev A. Rule-based canonicalization of arbitrary tables in spreadsheets // Proc. 22nd Int. Conf. on Inf. and Softw. Technol. 2016. CCIS 639, P. 78–91.
40. Shigarov A., Paramonov V. CRL: a rule language for analysis and interpretation of arbitrary tables // Selected Papers XVII Int. Conf. on Data Analytics and Management in Data Intensive Domains. 2015. CEUR-WS 1536. P. 22–29.
41. Shigarov A., Bychkov I. From unstructured to structured tabular data using a rule engine // Proc. 5th Int. W. on Comp. Sci. and Eng. 2015. P. 85–91.
42. Shigarov A. Rule-based table analysis and interpretation // Proc. 21st Int. Conf. on Inf. and Softw. Technol. 2015. CCIS 538. P. 175–186.

Главы в монографиях

43. Инфраструктура информационных ресурсов и технологии создания информационно-аналитических систем территориального управления / И.В. Бычков [и др.]; под ред. И.В. Быčkova; Ин-т динамики систем и теории управления имени В.М. Матросова СО РАН. – Новосибирск: Изд-во СО РАН, 2016. – 242 с.
44. Интеграция информационно-аналитических ресурсов и обработка пространственных данных в задачах управления территориальным развитием / И.В. Бычков [и др.]; под ред. И.В. Быčkova; Ин-т динамики систем и теории управления имени СО РАН. – Новосибирск: Изд-во СО РАН, 2012. – 369 с.

Личный вклад автора диссертации. Все выносимые на защиту положения получены соискателем лично. Из совместных работ, в том числе [4–7, 9, 11, 12], в диссертацию включены только те результаты, которые принадлежат непосредственно автору, включая постановку задач, разработку предлагаемых методов, моделей, языковых и инструментальных средств для АПТ, а также планирование и проведение экспериментов. В неделимом соавторстве с А. А. Алтаевым, А. И. Бондаревым, А. А. Михайловым, В. В. Парамоновым, Е. В. Рожковым, В. В. Христюком и И. А. Черепановым выполнена программная реализация и оценка производительности предлагаемых методов. Совместно с А. А. Ветровым, Н. О. Дородных, В. В. Парамоновым и А. Ю. Юриным реализованы примеры их практического применения, с И. В. Бычковым, Г. М. Ружниковым и А. Е. Хмельновым определены направления и методы исследования.

Апробация научных работ соискателя подтверждена представлением результатов диссертационной работы на следующих международных и всероссийских научных мероприятиях, в том числе: «Информационные и математические технологии в науке и управлении» (Иркутск, 2007–2009, 2013 гг.); «Pattern recognition and image analysis» (Нижний Новгород 2008 г.; Санкт-Петербург 2010 г.); «Matematičke i informacione tehnologije» (Белград, Сербия, 2010 г.); «Математическое моделирование и вычислительно-информационные технологии в междисциплинарных научных исследованиях» (Иркутск, 2011, 2013–14 гг.); «Математическое моделирование и информационные технологии» (Иркутск, 2012, 2017 гг.); «Digitization of cultural and scientific heritage» (Белград, Сербия, 2014 г.); «Распределенные информационно-вычислительные ресурсы. Наука – цифровой экономике» (Новосибирск, 2017 г.); «Data analytics and management in data intensive domains» (Дубна, 2014 г.; Обнинск, 2015 г.; Ершово, 2016 г.); «Information and software technologies» (Каунас-Вильнюс, 2015, 2016, 2018–2021 гг.); «ACM Document engineering» (Вена, Австрия, 2016 г.; Галифакс, Канада, 2018 г.); «Information systems architecture and technology» (Вроцлав, Польша, 2019); «Information and communication technology, electronics and microelectronics» (Опатия, Хорватия, 2019, 2020, 2022 гг.); «Information, computation, and control systems for distributed environments» (Иркутск, 2023 г.); «Марчуковские научные чтения» (Новосибирск, 2024 г.). Они также обсуждались на ряде совещаний и семинаров в ИДСТУ СО РАН (Иркутск), ОНИТ СО РАН (Новосибирск, 2016 г.), Институте национального

развития Монголии (Улан-Батор, Монголия, 2012 г.), Харбинском политехническом университете (Харбин, Китай, 2014 г.) и др.

Соответствие паспорту специальности. Основные результаты диссертационной работы Шигарова А.О. направлены на решение задач повышения эффективности автоматизации процессов извлечения данных из таблиц электронных документов неструктурированного формата, что соответствует специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей». Разработанные в рамках диссертационной работы модели, методы и методики могут использоваться при проектировании программных систем извлечения данных из документных таблиц (см. защищаемые положения 2–5). Созданные инструментальные средства обеспечивают разработку прикладного ПО извлечения данных из таблиц РК на основе проблемно-ориентированного языка правил (см. защищаемые положения 5 и 6). Представленные модели, методы и методики позволяют организовать взаимодействие программ распознавания, анализа и интерпретации документных таблиц (см. защищаемые положения 1–6).

Основные результаты диссертации соответствуют следующим направлениям исследований паспорта специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»:

1. «Модели, методы и алгоритмы проектирования, анализа, трансформации, верификации и тестирования программ и программных систем» — защищаемые положения 2–5.
2. «Языки программирования и системы программирования, семантика программ» — защищаемые положения 5 и 6.
3. «Модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем» — защищаемые положения 1–6.

Диссертационное исследование Шигарова А.О. является завершенной, самостоятельной научно-исследовательской работой, подводящей итог многолетним исследованиям соискателя. Полученные в диссертации результаты можно квалифицировать как совокупность новых научно обоснованных технологических решений, внедрение которых вносит значительный вклад в развитие страны посредством повышения эффективности и надежности процессов обработки данных и знаний в вычислительных машинах, комплексах и компьютерных сетях при распознавании, анализе и интерпретации таблиц, представленных в электронных документах неструктурированного формата.

Диссертация соответствует требованиям ВАК РФ, предъявляемым к докторским диссертациям, не содержит заимствованного материала без ссылок на автора и (или) источник заимствования.

Диссертационная работа Шигарова А.О. «Методы и инструментальные средства автоматизации процессов извлечения данных из таблиц электронных документов неструктурированного формата» рекомендуется к защите на соискание ученой степени доктора технических наук по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Обсуждение диссертации проведено на заседании Объединенного семинара ИДСТУ СО РАН (протокол № 2 от 1 июля 2024 г.).

Заключение принято на заседании Ученого совета ИДСТУ СО РАН. На заседании присутствовало 17 чел. Результаты голосования: «за» – 17 чел., «против» – нет, «воздержались» – нет (протокол № 4 от «21» мая 2025 г.).

Председатель Ученого совета,
директор ИДСТУ СО РАН,
академик



И.В. Бычков