

Отзыв научного консультанта

на диссертацию Шигарова Алексея Олеговича «Методы и инструментальные средства автоматизации процессов извлечения данных из таблиц электронных документов неструктурированного формата», представленную на соискание ученой степени доктора технических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Шигаров А. О. получил высшее образование по специальности «Прикладная математика» в Иркутском государственном университете (1999–2004 гг.); с 2001 года по настоящее время работает в ИДСТУ СО РАН в лаборатории комплексных информационных систем; окончил аспирантуру ИДСТУ СО РАН (2004–2006 гг.); в 2010 г. защитил диссертацию на соискание кандидата технических наук по специальности 05.25.05 — «Информационные системы и процессы». В настоящее время он занимает должность ведущего научного сотрудника в ИДСТУ СО РАН.

Шигаров А. О. ведет активную научную деятельность. Всего, начиная с 2005 г., им было опубликовано 177 научных работ. Основным направлением его исследований является автоматизация обработки документной и табличной информации, представленной в неструктурированном и полуструктурированном виде. Именно в рамках данного направления было выполнено его диссертационное исследование.

В диссертации Шигарова А. О. предложены новые методы, модели, языковые и инструментальные средства для повышения эффективности и надежности комплексных процессов извлечения реляционных данных из документных таблиц, представленных в неструктурированном формате. Полученные в рамках его диссертации результаты расширяют текущее состояние исследований по актуальному направлению, а именно «автоматизированному пониманию табличной информации». В том числе впервые предложены теоретические основы и технологические решения автоматизированного понимания табличной информации, допускающие в отличие от имеющихся аналогов произвольность структуры документных таблиц, включая свободную компоновку, структурированность ячеек и иерархические заголовки. Предложенные методы, модели, языковые и инструментальные средства позволяют снизить трудозатраты и сократить сроки разработки программного обеспечения извлечения данных из документных таблиц за счет применения правил анализа и интерпретации таблиц. Программное обеспечение, реализованное на основе результатов, представленных в диссертации, переменялось в ряде научных и промышленных проектов. Получены письменные подтверждения об использовании этих результатов в разработках компаний: «CloudSDS Inc.», «Mosaik Risk Solutions Inc.» и «ООО Адвентум Консалтинг».

Основные результаты диссертационного исследования Шигарова А. О. получены в ходе выполнения проектов по грантам научных фондов, по которым он являлся руководителем, а именно: «Методология и инструментальная платформа разработки систем извлечения данных из произвольных электронных таблиц» (РНФ, 2018–2023 гг., № 18-71-10001); «Методы интеграции неструктурированных табличных данных» (РФФИ, 2015–2016 гг., № 15-37-20042); «Методы извлечения табличной информации из неструктурированных текстовых источников» (Совет по грантам Президента РФ, 2013–2015 гг., № СП-3387.2013.5); «Интеллектуальная система извлечения информации из слабоструктурированных таблиц со сложной компоновкой» (РФФИ, 2012–2013 гг., № 12-07-31051). Отдельные результаты были получены при выполнении базовых проектов программы фундаментальных исследований СО РАН, в том числе: № IV.38.1.2. (2017-2021 гг.), № IV.38.2.3 (2013-2015 гг.), № IV.31.2.4 (2010-2012 гг.), а также крупного научного проекта по приоритетным направлениям научно-технологического развития по гранту Министерства науки и высшего образования РФ (№ 075-15-2024-533, 2024–2026 гг.). Представленные в диссертации результаты,

касающиеся разработки метода и инструментальных средств анализа и интерпретации таблиц на основе исполнения правил, были включены в «Доклад о состоянии фундаментальных наук в РФ и о важнейших научных достижениях российских ученых в 2015 году» как одно из важнейших достижений по разделу «Нанотехнологии и информационные технологии».

Результаты диссертационного исследования Шигарова А. О. в полной мере были представлены на международных, всероссийских и региональных научных мероприятиях, а также опубликованы в 70 работах, включая главы в 2 монографиях, 12 статей в журналах, рекомендованных ВАК для опубликования научных результатов диссертации, из которых четыре издания относятся к первому квартилю по индексу SJR и еще одно — ко второму; всего из них 34 статьи опубликованы в изданиях, проиндексированных в международных базах цитирования: «Web of Science», Scopus и др. В составе коллектива автором получено 6 свидетельств о государственной регистрации программ для ЭВМ.

Тема и основные результаты диссертации соответствуют области исследований паспорта специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей». Достоверность и обоснованность результатов исследований, а также личный вклад автора диссертации в эти результаты не вызывают сомнений.

Считаю, что диссертация Шигарова А. О. является оригинальным и значимым научным исследованием, отвечает всем требованиям ВАК, предъявляемым к диссертационным работам, а ее автор, Шигаров А. О., заслуживает присуждения степени доктора технических наук по специальности 2.3.5 — «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Научный консультант,
директор ИДСТУ СО РАН
академик



И.В. Бычков