

ОТЗЫВ

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

Представленная работа посвящена важной проблеме извлечения неструктурированных данных как носителей разнородной и важной информации, заключенной во множестве неструктурированных электронных документов типа всевозможных разноформатных таблиц, множества различных вебстраниц, рабочих книг, растровых изображений и много другого.

Одним из наиболее востребованных видов такой информации являются всевозможные документные таблицы, которые скрывают реляционные данные в произвольной неструктурированной форме

Учитывая гигантский объем информации, заключенной в неструктурированных данных, научная проблема разработки методов автоматизации процессов извлечения их из документных таблиц произвольной структуры по определенным целевым критериям является актуальной и весьма востребованной. Основа ее состоит в получении структурированных данных в виде сущностей и связей между ними из неструктурированных источников.

Отсутствие на сегодняшний день какого-либо общего решения, которое позволило бы автоматически извлекать и структурировать данные из таких источников, обуславливает актуальным исследование и разработки, направленные на то, чтобы сделать этот процесс более эффективным и надежным за счет увеличения автоматизации и сокращения ручной работы.

Диссертационная работа Шигарова А.О. посвящена решению этой задачи в части создания моделей, методов, языковых и инструментальных средств для разработки прикладного программного обеспечения извлечения реляционных данных из документных таблиц, представленных в неструктурированном виде, за счет поддержки произвольности табличной структуры. Разработанные средства в совокупности служат для создания технологии первичной подготовки табличных данных и их дальнейшего применения в аналитике, конструировании графов знаний, информационном обмене и других приложениях.

Научная новизна диссертационного исследования заключается в исследовании и развитии комплексного подхода к решению проблемы автоматизации процессов преобразования разнородных табличных данных, представленных изначально в неструктурированном виде (т.е. без формальной схемы), к структурированному виду (т.е. с формальной схемой), включающего в себя новые модели, методы, языковые и инструментальные средства. Созданные средства позволяют повысить эффективность и надежность разрабатываемых на их основе прикладных программ извлечения данных из документных таблиц как наиболее распространенных форматов языков описания страниц и рабочих книг.

Задачи исследования по главам сформулированы достаточно точно и полно. Методологические подходы, примененные автором при выполнении исследования,

адекватны цели и задачам исследования, характеризуются современным уровнем разработок. Полученные результаты и сформулированные выводы полностью соответствуют поставленным задачам.

Обоснованность и достоверность научных положений и выводов подтверждается использованием автором теоретически обоснованных современных методов исследования, результатами экспериментов по решению научных и прикладных задач. Представленные в диссертации количественные оценки результатов экспериментальных исследований находятся в ряду мировых по данной тематике.

Результаты исследований, представленные в автореферате, соответствуют специальности 2.3.5. – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей», опубликованы автором в высокорейтинговых журнальных статьях из баз Web of Science и Scopus, представлялись на ряде ведущих международных и всероссийских научных конференций.

По результатам диссертационного исследования зарегистрировано 6 программ для ЭВМ в Роспатенте.

Замечание. В связи с рассмотрением в главе 2 метода автоматизации распознавания таблиц, возникает вопрос о возможности использования метода нейросетевого распознавания, как адекватного метода для рассматриваемой задачи. Это обусловлено и большим объемом восстанавливаемых сущностей и их семантическим разнообразием и, наконец, существующими возможностями использования иерархических нейросетей.

В целом, судя по автореферату, диссертация Шигарова А.О. представляет собой самостоятельную научно-квалификационную работу, которая характеризуется логически выдержанной последовательностью выполнения всех этапов исследования, их внутренним единством и завершенностью, а также существенной теоретической и практической значимостью полученных результатов. Представленную диссертацию можно рассматривать как значимый научный вклад в решение проблемы повышения эффективности и надежности автоматизированных процессов извлечения и структурирования данных, изначально представленных в неструктурированном виде в документных таблицах произвольной формы и содержания. В ней предложены новые научно-обоснованные теоретические положения и технологические решения, которые в совокупности можно считать важным научным достижением в области развития математического и программного обеспечения для автоматизации обработки табличных данных.

Диссертация соответствует критериям, установленным пп. 9–14 «Положения о присуждении ученых степеней» для ученой степени доктора наук, а ее автор, Шигаров Алексей Олегович, заслуживает присуждения ученой степени доктора технических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

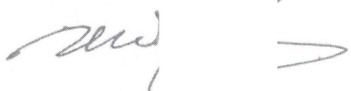
Отзыв представил:

Хайретдинов Марат Саматович

доктор технических наук, главный научный сотрудник Лаборатории геофизической информатики Федерального государственного бюджетного учреждения науки Института вычислительной математики и математической геофизики Сибирского отделения Российской академии наук (ИВМиМГ СО РАН), профессор Новосибирского государственного технического университета.

Даю свое согласие на включение своих персональных данных в документы, связанные с работой диссертационного совета, их дальнейшую обработку и передачу в соответствии с требованиями Минобрнауки России.

Подпись Хайретдинова М.С. заверяю
Заведующая отделом кадров
Трофимкина Елена Юрьевна



М. С. Хайретдинов

10.12. 2025 г.

Контактная информация:

ИВМиМГ СО РАН

Адрес: 630090, Россия, Новосибирск, Проспект академика Лаврентьева, 6.

Телефон: +7 (383) 330-87-43, эл. почта: marat@org.sccc.ru

