

ОТЗЫВ

официального оппонента на диссертационную работу **Бурлаевой Екатерины Игоревны** «Совершенствование методов системного анализа в задачах управления и систематизации специализированной информации», представленную на соискание учёной степени кандидата технических наук по специальности 05.13.01 – «Системный анализ, управление и обработка информации (по отраслям) (технические науки)».

1. Актуальность темы диссертации.

Согласно всеобщему признанию, информатизация всех сторон деятельности человека является ведущей парадигмой развития общества в XXI веке. По оценкам экспертов, более 70% накопленной и используемой информации представлено в несистематизированной текстовой форме, вследствие чего получение требуемых сведений по конкретной тематике весьма затруднено.

В этой связи задача развития методов и систем классификации, структуризации и управления специализированной текстовой информацией является, безусловно, актуальной, имеющей важное значение практически во всех отраслях.

2. Обоснованность научных положений, выводов и рекомендаций, сформулированных в диссертации.

Обоснованность положений, выводов и рекомендаций обеспечивается полнотой анализа теоретических и практических исследований отечественных и зарубежных авторов, корректным применением фундаментальных положений при разработке методов и осуществлении теоретических и лабораторных исследований, положительной оценкой результатов работы на научных конференциях и семинарах, а также выполненных публикаций.

Достоверность положений подтверждается достаточным объемом экспериментальных исследований в лабораторных условиях и удовлетворительной степенью соответствия результатов теоретических и лабораторных данных.

3. Анализ содержания диссертации.

В **первой главе** выполнен анализ разработанных к настоящему времени методов и средств управления информацией, систематизации документов, оформляемых в виде текстовых массивов.

ОБЛАСТНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Вх. № 16/121.
« 24 » 10. 20 19 г.

Отсутствие возможности вовремя и быстро получить необходимую информацию по нужной теме делает бесполезной большую часть накопленных знаний, вследствие чего появляется необходимость в управлении, анализе и систематизации текстовой информации, выполняемой автоматическими системами обработки информации.

Для создания эффективных систем управления и систематизации информацией требуется применение методов классификации на основе машинного обучения и методов, основанных на знаниях.

В результате выполненного анализа сформулированы цель и задачи исследования.

В качестве замечания можно отметить, что в первой главе встречаются отдельные недостаточно строгие формулировки, например, на стр. 32 «...при решении реальных задач все перечисленные требования могут оказаться неточными, а два последних редко используются...». Если для решения задач выдвигаются определённые требования, то они не могут быть неточными.

Вторая глава посвящена теоретическим исследованиям способов автоматического управления текстовыми документами и обоснованию выбора базовых методов для построения модернизированной автоматической компьютерной системы управления и систематизации специализированной текстовой информацией.

В результате исследований проведён выбор методов для построения модели модернизированной системы классификации и управления текстовой информацией. Для дальнейших исследований проведён анализ и выбран один из наиболее эффективных методов систематизации и управления информацией, использование принимается в качестве стандарта для сравнения качества работы с другими модернизированными моделями, обрабатывающими специализированную текстовую информацию.

Следует сказать, что анализируемые в данной главе методы представляются не всегда с исчерпывающей полнотой. В частности, при описании векторного представления документов приведен пример на рис. 2.3 (с. 49), но комментарий не даётся. То же можно сказать о представлении алгоритма k -ближайших соседей (рис. 2.4, с. 53) и др.

Кроме того, некоторые характеристики алгоритмов не сопровождаются числовыми показателями (например, вычислительная сложность алгоритма, с. 54, табл. 2.1).

При оценке качества классификации (п. 2.5, с. 66 – 68) рассматривается ряд аналитических зависимостей, однако пример их применения не приводится.

В **третьей главе** выполнено построение модернизированной модели управления и систематизации информации на базе предложенного оптимального способа анализа знаний, представляемого в виде стандарта IDEF. Формализация данных единым универсальным системным способом представления знаний позволяет создать соответствующие алгоритмы и подобрать инструментальные средства для обработки знаний с помощью единого формального аппарата.

Разработана обобщенная архитектура систематизации информации на основе предложенной модели, которая показывает общее представление ее построения и позволяет перейти к практической реализации ее прототипа и исследованию практической эффективности предложенных решений.

Следует отметить, что приведенная архитектура описана теоретически, вопрос практической реализации не рассмотрен.

В **четвертой главе** представлено практическое исследование эффективности разработанных средств, при использовании тематического анализа коллекции документов. Осуществлено практическое сравнение методов машинного обучения при использовании различных средств классификации и предложены эффективные модели модернизированных композиций, объединяющих в себе несколько методик, для получения наилучших показателей качества систематизации информации.

Для улучшения характеристик обобщенной модели, осуществляющей систематизацию информации, предложены и обоснованы модели композиций методов систематизации.

Экспериментальное исследование предложенной общей модели автоматической систематизации и управления информацией, основанной на выполненном объединении достоинств синтагматических и парадигматических подходов, и всех предложенных модернизаций показало увеличение полноты и точности работы в среднем на 32,5% и 31,5% соответственно.

Экспериментальное исследование предложенной общей модернизированной модели управления и систематизации текстовых документов на основе всех разработанных средств показало повышение скорости и качества обработки текстовой информации в среднем более чем в 4 раза по сравнению с ручным способом (так, на обработку одного текстового документа вручную в среднем затрачивается 10 - 30 минут, в то время как автоматическая обработка на основе разработанных средств – в среднем затрачивает 3 - 7 минут).

Следует отметить, что в данной главе автором приведены результаты экспериментальных исследований с указанием конкретных числовых данных, однако недостаточно чётко показан процесс получения этих данных.

4. Научная новизна работы.

Новизна разработанных в диссертации научных положений, выводов и рекомендаций состоит в следующем.

1. Впервые предложена усовершенствованная общая модель автоматической систематизации и управления информацией, основанная на объединении достоинств синтагматических и парадигматических подходов. Использование новой модели позволяет повысить полноту и точность работы модели в среднем на 32,5% и 31,5% соответственно.

2. Получила дальнейшее развитие и модернизирована модель классификации текстовой информации на основе внесения изменений в

структуру алгоритма её построения, что позволяет повысить полноту и точность работы модели еще на 5,5% и 8,5 % соответственно.

3. Экспериментально обоснована модель вычислительной композиции распределения веса слов в текстовом документе. При разном распределении веса термина, повышение качества работы предложенной общей модели систематизации и управления информацией варьируется в пределах 10% в зависимости от используемой композиции.

Практическая ценность результатов работы состоит в обосновании решения задачи повышения эффективности предложенной общей модели систематизации и управления текстовой информацией за счет внесения изменений в структуру алгоритма её построения и вычислительных композиций распределения веса слов в документах, правил отбора неинформативных признаков и способов взвешивания термов, что обеспечивает дополнительное повышение качества распределения информации на 27,5%.

5. Ценность для науки и практики полученных автором результатов.

Научное значение работы состоит в том, что предложенная комплексная методика построения модели автоматической классификации и статистического анализа является совершенствованием существующих подходов к обработке информации и в дальнейшем может быть расширена и дополнена функциями автоматического и автоматизированного тематического анализа потоков текстовой информации для расширения количества тематик, по которым распределяются текстовые документы, а также повышением качества модели автоматической обработки информации.

Практическое значение результатов работы состоит в том, что предложенная практическая реализация усовершенствованной модели систематизации и управления информацией позволяет формировать текстовые базы данных классифицированной информации в автоматическом режиме, а на основании результатов классификации имеется возможность формировать аналитические задачи и статистические базы данных по результатам обработки текстов, автоматизировать работу специалистов-аналитиков, осуществляющих тематический анализ текстовой информации, и ведение аналитических задач в различных предметных областях, что может послужить функциональным дополнением и развитием информационных систем различных организаций.

6. Степень полноты опубликования полученных результатов.

Основные научные результаты диссертации представлены в 9 научных публикациях, 5 из них в рецензируемых научных изданиях, в том числе 2 – в рецензируемых научных журналах и изданиях, в которых должны быть опубликованы основные научные результаты диссертаций на соискание

ученой степени кандидата и доктора наук в Российской Федерации, и 3 - в специализированных научных изданиях, рекомендованных ВАК ДНР, 4 – по материалам научных конференций.

7. Реализация результатов диссертации в промышленности и предложения по их дальнейшему использованию.

Практическое значение полученных результатов подтверждается:

– внедрением в практику организации информационных массивов и баз данных с целью совершенствования компьютерной технологии прогноза в отделе сдвижения земной поверхности и охраны подрабатываемых объектов (СЗПО) Республиканского академического научно–исследовательского и проектно–конструкторского института горной геологии, геомеханики, геофизики и маркшейдерского дела (РАНИМИ) (справка о внедрении № 01/140 от 20.05.2019 г.);

– внедрением в учебный процесс ГОУВПО «ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» (справка № 52.1–05/19 от 14.05.2019 об использовании в учебном процессе при чтении лекций и проведении практических занятий на кафедрах «Искусственный интеллект и системный анализ» и «Прикладная математика» по дисциплинам: «Организация баз данных и знаний», «Информационные системы и технологии», «Стандартизация и сертификация в сфере информационных технологий», «Распределённые информационно–аналитические системы», «Корпоративные информационные системы»).

На основании результатов классификации в дальнейшем имеется возможность формировать аналитические задачи и статистические базы данных по результатам обработки текстов, автоматизировать работу специалистов–аналитиков, осуществляющих тематический анализ текстовой информации и ведение аналитических задач в различных предметных областях, что может послужить функциональным дополнением и развитием информационных систем различных организаций.

8. Соответствие содержания диссертации специальности 05.13.01 – «Системный анализ, управление и обработка информации (по отраслям), (технические науки)».

Содержание рецензируемой диссертационной работы, выдвинутые научные положения, полученные выводы и рекомендации дают основание сделать заключение о том, что диссертация Бурлаевой Е.И., представленная к защите на соискание ученой степени кандидата технических наук, соответствует паспорту специальности 05.13.01 – «Системный анализ, управление и обработка информации (по отраслям), (технические науки)».

9. Соответствие автореферата содержанию диссертации.

В автореферате в достаточном объеме изложены основные результаты исследований, приведенных в диссертации. В нем представлены научные положения, новизна и практическая значимость, основные выводы по 4 главам диссертации, заключение, реализация результатов работы в промышленности и предложения по их дальнейшему использованию.

10. Структура, стиль и язык диссертации.

Структурное построение диссертации соответствует цели и задачам исследований. Стиль изложения содержания исследований и подача материала вполне логичны, последовательны и связаны единой идеей. Язык диссертации достаточно ясен и доступен для восприятия.

11. Замечания по диссертации.

11.1. В первой главе встречаются отдельные недостаточно строгие формулировки, например, на стр. 32 «...при решении реальных задач все перечисленные требования могут оказаться неточными, а два последних редко используются...». Если для решения задач выдвигаются определённые требования, то они не могут быть неточными.

11.2. Анализируемые во второй главе методы представляются не всегда с исчерпывающей полнотой. В частности, при описании векторного представления документов приведен пример на рис. 2.3 (с. 49), но комментарий к рисунку не даётся. То же можно сказать о представлении алгоритма k -ближайших соседей (рис. 2.4, с. 53) и др.

11.3. В главе 2 некоторые характеристики алгоритмов не сопровождаются числовыми показателями (например, вычислительная сложность алгоритма, с. 54, табл. 2.1).

11.4. Во второй главе при оценке качества классификации (п. 2.5, с. 66 – 68) рассматривается ряд аналитических зависимостей, однако пример их применения не приводится.

11.5. Приведенная в третьей главе архитектура описана теоретически без конкретной привязки к возможной элементной базе.

11.6. В четвертой главе автором приведены результаты экспериментальных исследований с указанием конкретных числовых данных, однако недостаточно чётко показан процесс получения этих данных.

Заключение.

Отмеченные в отзыве недостатки не исключают общей положительной оценки диссертации. Научные положения, выводы и рекомендации, сформулированные в диссертации, несмотря на недочеты, вполне приемлемы и не вызывают принципиальных возражений.

В целом диссертационная работа Бурлаевой Е.И. «Совершенствование методов системного анализа в задачах управления и систематизации специализированной информации» является законченной научно-исследовательской работой, посвященной разработке актуальной тематики, имеет научную новизну, практическое значение и достаточную реализацию в отрасли, выполнена на достаточно высоком научном уровне, соответствует паспорту специальности 05.13.01– «Системный анализ, управление и обработка информации (по отраслям), (технические науки)» и отвечает требованиям, предъявляемым ВАК к кандидатским диссертациям.

За получение новых научно обоснованных результатов в области классификации, систематизации и управления специализированной текстовой информацией, заключающихся в установлении эффективности комбинированных методов обработки информации, что позволило дать новое решение актуальной научной задачи повышения эффективности управления и систематизации специализированной текстовой информации больших объемов с учетом особенностей русскоязычных текстовых документов, реализованных с помощью модернизации моделей, методик, и алгоритмов системного анализа, искусственного интеллекта и компьютерной лингвистики, автор представленной диссертации Бурлаева Екатерина Игоревна заслуживает присуждения ей ученой степени кандидата технических наук по специальности 05.13.01– «Системный анализ, управление и обработка информации (по отраслям) (технические науки)».

Официальный оппонент, доцент кафедры «Высшая математика и информатика» ГОУ ВПО «Донбасская национальная академия строительства и архитектуры»,

кандидат экономических наук по специальности

08.03.02 – «Экономико-математическое моделирование»,

доцент по специальности 05.13.01– «Системный анализ,

управление и обработка информации»

 С.В. Сторожев

Адрес: 286123, ДНР, г. Макеевка, ул. Державина, 2,

тел.: (062) 343 70 33

эл. почта: mailbox@donnasa.org

Я, Сторожев Сергей Валерьевич, даю согласие на автоматизированную обработку моих персональных данных.

Кандидат экономических наук, доцент

 С.В. Сторожев

Подпись доцента кафедры

«Высшая математика и информатика» Сторожева С.В. подтверждаю

  Иванов А.А.

