

## ОТЗЫВ

официального рецензента на диссертационную работу

**Чинибаевой Толғанай Темірболатқызы на тему «Модели и методы управления данными с гетерогенными структурой (Big Data)», предоставленную на соискание степени доктора философии (PhD) по специальности «6D070400 – Вычислительная техника и программное обеспечения»**

№п/п	Критерии	Соответствие критериям (необходимо отметить один из вариантов ответа)	Обоснование позиции официального рецензента
1.	Тема диссертации (на дату ее утверждения) соответствует направлениям развития науки и/или государственным программам	1.1 Соответствие приоритетным направлениям развития науки или государственным программам:	Диссертация соответствует задачам государственной программы «Цифровой Казахстан». (12 декабря 2017 г. Постановление Правительства Республики Казахстан № 827)
		1) Диссертация выполнена в рамках проекта или целевой программы, финансируемого(ой) из государственного бюджета (указать название и номер проекта или программы) 2) <u>Диссертация выполнена в рамках другой государственной программы (указать название программы)</u> 3) Диссертация соответствует приоритетному направлению развития науки, утвержденному Высшей научно-технической комиссией при Правительстве Республики Казахстан (указать направление)	Диссертация соответствует приоритетному направлению развития науки на 2021-2023 годы: 4. Информационные, телекоммуникационные и космические технологии; от 29.04.2020 протокол заседания Высшей научно-технической комиссией при Правительстве Республики Казахстан
2.	Важность для науки	Работа <u>вносит/не вносит</u> существенный вклад в науку, а ее важность хорошо <u>раскрыта/не раскрыта</u>	Работа вносит существенный вклад в науку, а ее важность хорошо раскрыта
3.	Принцип самостоятельности и	Уровень самостоятельности: 1) Высокий; 2) <u>Средний</u> ; 3) Низкий; 4) Самостоятельности нет	Докторанту принадлежит гипотеза, экспериментальные подтверждения, разработанная методика, формулировка научных положений, разработка системы
4.	Принцип	4.1 Обоснование актуальности диссертации:	Актуальность диссертации

215

	внутреннего единства	1) <u>Обоснована</u> ; 2) Частично обоснована; 3) Не обоснована.	определяется изложением моделей и методов управления большими данными с неоднородной структурой, используемых для мониторинга и анализа информации, описывающей деятельность научных организаций.
		4.2 Содержание диссертации отражает тему диссертации: 1) <u>Отражает</u> ; 2) Частично отражает; 3) Не отражает	Содержание полностью отражает тему диссертации
		4.3. Цель и задачи соответствуют теме диссертации: 1) <u>соответствуют</u> ; 2) частично соответствуют; 3) не соответствуют	Целью исследования является разработка моделей и методов управления научной информацией с гетерогенной структурой.
		4.4 Все разделы и положения диссертации логически взаимосвязаны: 1) <u>полностью взаимосвязаны</u> ; 2) взаимосвязь частичная; 3) взаимосвязь отсутствует	Диссертация состоит из введения, четырех разделов, заключения и списка литературы. Все разделы взаимосвязаны
		4.5 Предложенные автором новые решения (принципы, методы) аргументированы и оценены по сравнению с известными решениями: 1) <u>критический анализ есть</u> ; 2) анализ частичный; 3) анализ представляет собой не собственные мнения, а цитаты других авторов	Разработанные алгоритмы извлечение терминов-пар и построение онтологии в научной области были протестированы. Результаты тестирования представлены в графиках.
5.	Принцип научной новизны	5.1 Научные результаты и положения являются новыми? 1) полностью новые; 2) <u>частично новые (новыми являются 25-75%)</u> ; 3) не новые (новыми являются менее 25%)	Автором разработан новый алгоритм построения онтологии отдельной области научного знания на основе выделения терминов из анонсов научных конференций, а также путем использования информации из поисковых систем в Интернет. Также разработан новый алгоритм выделения терминов-пар слов из коллекций текстов с заданным тематическим делением, эффективность которого в сравнении

			с классическими алгоритмами продемонстрирована на задачах классификации и кластеризации текстов.
		5.2 Выводы диссертации являются новыми? 1) полностью новые; 2) <u>частично новые (новыми являются 25-75%);</u> 3) <u>не новые (новыми являются менее 25%)</u>	Выводы диссертации являются частично новыми для управления гетерогенными научными данными.
		5.3 Технические, технологические, экономические или управленческие решения являются новыми и обоснованными: 1) полностью новые; 2) <u>частично новые (новыми являются 25-75%);</u> 3) <u>не новые (новыми являются менее 25%)</u>	При реализации интеллектуальной системы были использованы библиотеки FreeCite++ для английского языка и модернизированы при работе с русским и казаским языком
6.	Обоснованность основных выводов	Все основные выводы <u>основаны/не основаны</u> на весомых с научной точки зрения доказательствах либо достаточно хорошо обоснованы (для qualitative research и направлений подготовки по искусству и гуманитарным наукам)	Теоретические исследования, проведенные в рамках диссертации, позволили получить результаты, имеющие практическое и научное значение.
7.	Основные положения, выносимые на защиту	Необходимо ответить на следующие вопросы по каждому положению в отдельности: 7.1 Доказано ли положение? 1) доказано; 2) скорее доказано; 3) скорее не доказано; 4) не доказано 7.2 Является ли тривиальным? 1) да; 2) нет 7.3 Является ли новым? 1) да; 2) нет 7.4 Уровень для применения: 1) узкий; 2) средний; 3) широкий	Диссертацияның негізгі қағидаттары: 1) разработаны алгоритмы при работе с большими данными с гетерогенной структурой; 7.1 положение доказано; 7.2 не тривиальный; 7.3 положение является новым; 7.4 уровень применения: широкий; 7.5 доказано в статье «Model, data integration algorithms of information systems based on ontology»; 2) Предоставляется формальное описание запросов на SPARQL; 7.1 положение доказано; 7.2 не тривиальный; 7.3 положение является новым; 7.4 уровень применения: широкий;



		<p>7.5 Доказано ли в статье?</p> <p>1) да; 2) нет</p>	<p>7.5 доказано в статье «Recursive decomposition as a method for integrating heterogeneous data sources»;</p> <p>3) Разработан алгоритм создания онтологии в научной области;</p> <p>7.1 положение доказано;</p> <p>7.2 не тривиальный;</p> <p>7.3 положение является новым;</p> <p>7.4 уровень применения: широкий;</p> <p>7.5 доказано в статье «Algorithm for the construction of an ontology in the field of scientific knowledge»;</p> <p>4) Разработан алгоритм выделения терминов-пар из набора текстов.</p> <p>7.1 положение доказано;</p> <p>7.2 не тривиальный;</p> <p>7.3 положение является новым;</p> <p>7.4 уровень применения: широкий;</p> <p>7.5 доказано в статье «Method of extracting meta description from databases»;</p>
8.	Принцип достоверности Достоверность источников и предоставляемой информации	<p>8.1 Выбор методологии - обоснован или методология достаточно подробно подробно описана</p> <p>1) <u>да</u>; 2) нет</p>	Интеллектуальная системы была разработана на основе онтологий, модель управления информации IDF5 (SAD). В качестве описание была выбран математический аппарат дескриптивная логика первого порядка.
		<p>8.2 Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий:</p> <p>1) <u>да</u>; 2) нет</p>	Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий. Алгоритм извлечения терминов-пар была реализована в

			основном на языке Java. В качестве базы данных была использована OrientDB
		8.3 Теоретические выводы, модели, выявленные взаимосвязи и закономерности доказаны и подтверждены экспериментальным исследованием (для направлений подготовки по педагогическим наукам результаты доказаны на основе педагогического эксперимента): 1) <u>да</u> ; 2) нет	Достоверность теоретических результатов исследования подтверждается научными данными, экспериментальными исследованиями, а также промышленными испытаниями в реальных условиях
		8.4 Важные утверждения <u>подтверждены</u> /частично подтверждены/не подтверждены ссылками на актуальную и достоверную научную литературу	Важные утверждения подтверждены ссылками на актуальную и достоверную научную литературу во всех разделах диссертации
		8.5 Используемые источники литературы достаточны/ <u>не достаточны</u> для литературного обзора	Используемые источники литературы состоит из 77 наименований использованных источников
9	Принцип практической ценности	9.1 Диссертация имеет теоретическое значение: 1) <u>да</u> ; 2) нет	Теоретические и прикладные результаты, полученные в работе, могут быть использованы при проектировании и разработке интеграции систем.
		9.2 Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике: 1) <u>да</u> ; 2) нет	Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике.
		9.3 Предложения для практики являются новыми? 1) полностью новые; 2) частично новые (новыми являются 25-75%); 3) не новые (новыми являются менее 25%)	Результат экспериментального исследования разработанной системы, демонстрирующий повышение качества эффективности за счет семантического

			представления ресурсов создает условия для автоматизации сбора научных публикации
10.	Качество написания и оформления	Качество академического письма: 1) высокое; 2) <u>среднее</u> ; 3) ниже среднего; 4) низкое.	Качество написания и оформления диссертационной работы среднее, структура и правила оформления соблюдены

Диссертационная работа Чинибаевой Толғанай Темірболатқызы по теме «Модели и методы управления данными с гетерогенными структурой (Big Data)» соответствует всем требованиям к диссертационным работам на соискание степени доктора философии (PhD) по специальности 6D070400 – «Вычислительная техника и программное обеспечения», а автор, Чинибаева Толғанай, заслуживает присуждения степени доктора философии (PhD) по специальности 6D070400 – «Вычислительная техника и программное обеспечения».

Имеются следующие замечание по данной работе:

1. При решении поставленных задачи использованы не самые передовые методы NLP. В частности, использование таких методов как контекстные эмбединги, потенциально могло бы дополнительно повысить эффективность предложенных алгоритмов.
2. Предложенная концепция потенциально обладает достаточной степенью универсальностью, однако недостаточно хорошо формализована с точки зрения ее более широкого применения также и других областях, таких, например, как образования и медицина.
3. Не достаточно полно раскрыт принцип выбора использованных методов в обзорной части данной работы.

**Официальный рецензент:**

ассоц. проф., к.т.н.  
заведующий лабораторий «Анализа и моделирования информационных процессов»  
Института вычислительных и информационных технологии КН МОН РК



Мусабаев Р.Р.