

REVIEW COMMENTS

of dissertation work by Chinibayeva T.T. "Models and Methods of Management of Data with Heterogeneous Structures (Big Data)", submitted for the degree of Doctor of Philosophy (PhD) in the specialty 6D070400 - Computer Science and Software Engineering.

This thesis is well organized and studied. Summarized as the followings.

The relevance of the research topic is determined by the presentation of models and methods for managing big data with a heterogeneous structure, used to monitor and analyze information describing the activities of scientific organizations.

The aim of the research is to develop software for searching, organizing, storing, annotating and analyzing information describing the publications of scientists in this field of science, using mathematical models, algorithms and document corpus.

The object of the research is heterogeneous scientific data.

The subject of the research is models and methods of managing data with a heterogeneous structure in order to ensure the semantic compatibility of documents.

Research methods. The tasks posed in the course of the research were solved by methods of analysis of natural language texts, classification and software engineering. The results were presented by the apparatus of mathematical statistics and mathematical logic.

The scientific novelty of the dissertation research is presented in the form of an intellectual system, where the author's developments are applied, namely, algorithms for constructing an ontology of a separate area of scientific knowledge and the extraction of terms-pairs of words from a collection of texts with a given thematic division, as well as a formal description of requests to the system using ontologies and SPARQL language.

Theoretical and practical significance of the work: scientific novelty and practical significance of the research are high. The research results are used to combine heterogeneous data and use them for further processing.

Chinibayeva T.T. was studied and discussed with me at scientific seminars organized by the Department of Computer Engineering and Information Security of the International University of Information Technologies, scientific seminars organized by Gachon University (South Korea, Seoul)

Her thesis is regarding heterogeneous science data. So, I strongly recommend to review the science research data principle when you organize using mathematical models. See FAIR principle at <https://www.go-fair.org/fair->

research data in EU. Also, I recommend to add real data examples and simulation results.

IT is very important how to organize data (hetero- and/or homogeneous data) in AI, because we found algorithms are very dependent by data which is organized. So, data modeling techniques are mandatory work for digital transformation era. For the reason, her thesis is very meaningful work.

Thank you so much and your great work.

Hope to successfully receive a PH.D degree.

Sincerely yours,

2021. June 28

Young Im Cho



A handwritten signature in black ink, appearing to read 'Young Im Cho' in a cursive style.

Professor at Dept. of Computer Engineering,
Gachon University,
Seoul, KOREA

ОТЗЫВ

на диссертационную работу Чинибаевой Т.Т. «Модели и методы управления данными с гетерогенными структурами (большие данные)», представленная на соискание ученой степени доктора философии (PhD) по специальности 6D070400 – Вычислительная техника и программное обеспечение.

данная работа достаточно хорошо организована и изучена. В кратце данную работу можно характеризовать следующим образом.

Актуальность темы исследования определяется представлением моделей и методов управления большими данными с гетерогенной структурой, используемых для мониторинга и анализа информации, описывающей деятельность научных организаций.

Целью исследования является разработка программного обеспечения для поиска, систематизации, хранения, аннотирования и анализа информации, описывающей публикации ученых в этой области науки, с использованием математических моделей, алгоритмов и формата документов.

Объект исследования - гетерогенные научные данные.

Предмет исследования - модели и методы управления данными с гетерогенной структурой с целью обеспечения семантической совместимости документов.

Методы исследования. Поставленные в ходе исследования задачи решались методами анализа текстов на естественном языке, классификации и программной инженерии. Результаты были представлены аппаратом математической статистики и математической логики.

Научная новизна диссертационного исследования представлена в виде интеллектуальной системы, в которой применены разработки автора, а именно алгоритмы построения онтологии отдельной области научного знания и извлечения терминов-пар слов из набор текстов с заданным тематическим разделением, а также формальное описание запросов к системе с использованием онтологий и языка SPARQL.

Теоретическая и практическая значимость работы: научная новизна и практическая значимость исследования высоки. Результаты исследования используются для объединения гетерогенных данных и использования их для дальнейшей обработки.

Чинибаева Т.Т. изучался и обсуждался со мной на научных семинарах, организованных кафедрой компьютерной инженерии и информационной безопасности Международного университета информационных технологий, научных семинарах, организованных Университетом Гачон (Южная Корея, Сеул)

Ее диссертация посвящена гетерогенным научным данным. Поэтому я настоятельно рекомендую пересмотреть принцип данных научных исследований при организации с использованием математических моделей. См. Принцип FAIR на <https://www.go-fair.org/fair-principles/>. Это очень известный принцип данных научных исследований в ЕС. Также рекомендую добавить примеры реальных данных и результаты моделирования.

Очень важно, как организовать данные (гетеро- и/или однородные данные) в ИИ, потому что мы обнаружили, что алгоритмы очень зависят от данных, которые организованы. Итак, в эпоху цифровой трансформации методы моделирования данных являются обязательной работой.

Большое вам спасибо и за вашу отличную работу.

Надеюсь на успешное получение степени доктора философии.

Искренне Ваш,
2021. 28 июня

Йонг Им Чо
Профессор кафедры компьютерной инженерии,
Университет Гачон,
Сеул, КОРЕЯ

Город Алматы, Республика Казахстан.
Двадцать шестое августа две тысячи двадцать первого года.

Перевод текста документа с английского языка на русский язык выполнен дипломированным переводчиком Миргиязовой Санам Анваровной, 01.10.1989 года рождения, ИИН 891001400246, в городе Алматы, Республика Казахстан.

Подпись: Миргиязова Санам Анваровна

Я, **Оршабекова Гаухар Булегеновна**, нотариус г. Алматы, действующий на основании государственной лицензии № 14013776 от 18 сентября 2014 года, выданной Комитетом регистрационной службы и оказания правовой помощи Министерства юстиции Республики Казахстан, свидетельствую подлинность подписи, сделанной переводчиком Миргиязовой Санам Анваровной. Личность подписавшего документ установлена, дееспособность и полномочия его проверены.



Зарегистрировано в реестре за № 2246

Взыскано: 1546 тенге

Нотариус Оршабекова



ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҚҰҚЫҚТЫҚ ҚҰРМАҚ КӨРСЕТУ КОМИТЕТІ БЕРГЕН
Пронумеровано и прошнуровано
№ 31746
Оршабекова