

ОТЗЫВ

на диссертационную работу **Әйтiм Әйгерiм Қайратқызы** «Модели и методы автоматической обработки неструктурированной информации», представленную на соискание ученой степени доктора философии (PhD) по ОП: 8D06101 – «Интеллектуальные системы»

Представленная диссертация **Әйтiм Ә.Қ.** является полноценным, самостоятельным и актуальным научным исследованием, выполненным на высоком академическом уровне. Работа посвящена разработке моделей и методов автоматической обработки неструктурированной текстовой информации на казахском языке, что представляет собой крайне важную и своевременную задачу в области компьютерной лингвистики и интеллектуальных систем.

Современная мировая практика в области искусственного интеллекта демонстрирует устойчивую тенденцию к расширению языкового охвата цифровых систем. Однако для многих агглютинативных и слабообеспеченных ресурсов языков, таких как казахский, по-прежнему отсутствуют качественные, открытые и полнофункциональные инструменты для автоматической обработки текстов. Диссертационная работа **Әйтiм Ә.Қ.** отвечает на этот вызов, предлагая комплексное решение, охватывающее все основные уровни лингвистического анализа: от морфологии до семантики.

В рамках диссертации разработан полный стек интеллектуальных модулей: морфологический анализ и генерация, частеречная разметка (POS), распознавание именованных сущностей (NER), семантическая ролевая разметка (SRL), разрешение кореференции. Особое внимание уделено адаптации современных нейросетевых архитектур, таких как KazBERT, BiLSTM+CRF и трансформерные модели, к структуре казахского языка. Реализован авторский морфологический генератор с учетом корневой структуры, суффиксации, сингармонизма и морфонологических правил.

Следует особо отметить оригинальную разработку – библиотеку QNLP, представляющую собой модульную, расширяемую и открытую систему для комплексной NLP-обработки казахских текстов. Данная библиотека уже применяется в ряде образовательных и исследовательских проектов, а также демонстрирует потенциал для промышленного применения в задачах интеллектуального анализа текстов на казахском языке.

Диссертация отличается высокой научной и методической проработанностью. Автор самостоятельно выполнила все этапы исследования – от сбора и препроцессинга текстового корпуса до обучения, валидации и сравнительного анализа моделей. Все построенные модули

протестированы на специально размеченном корпусе, а результаты оценены с использованием современных метрик (F1, Ассигасу и др.). В работе соблюдены требования научной добросовестности, а ключевые результаты представлены в публикациях в международных научных журналах и на конференциях.

Как зарубежный научный консультант, я могу с уверенностью отметить высокий уровень самостоятельности, академической зрелости и исследовательского мышления соискателя. Эйтім Ә.Қ. продемонстрировала глубокое понимание как лингвистических, так и вычислительных аспектов исследуемой проблемы. Её способность интегрировать знания в области прикладной лингвистики и искусственного интеллекта на практике соответствует лучшим международным стандартам подготовки PhD-исследователей.

Учитывая вышесказанное, считаю, что докторская диссертация **«Модели и методы автоматической обработки неструктурированной информации»** является законченной самостоятельной научно-исследовательской работой, отвечающей всем требованиям предъявляемым к диссертационным работам по данной специальности, а ее автор **Эйтім Әйгерім Қайратқызы** заслуживает присуждения ученой степени доктора философии (PhD) по ОП: 8D06101 – «Интеллектуальные системы»

Зарубежный консультант,
д.т.н, профессор,
факультет электротехники,
инженерные и компьютерные науки,
Люблинский технологический университет,
Польша



В.Вуйцик

POLITECHNIKA LUBELSKA
Katedra Elektroniki i Technik Informacyjnych
ul. Nadbystrzycka 38A, 20-618 Lublin
tel. 81 538 43 09, fax 81 538 43 12

REVIEW

on the PhD thesis by Aitim Aigerim Kairatkyzy entitled "Models and methods for the automatic processing of unstructured information", submitted for the degree of Doctor of Philosophy (PhD) in the EP: 8D06101 – Clever Systems

The presented thesis by Aitim A.K. is a full-fledged, independent and relevant scientific research performed at a high academic level. The work is devoted to the development of models and methods for automatic processing of unstructured textual information in the Kazakh language, which is an extremely important and timely task in the field of computational linguistics and intelligent systems.

Modern world practice in the field of artificial intelligence demonstrates a steady trend towards expanding the linguistic coverage of digital systems. However, many agglutinative and low-resource languages, such as Kazakh, still lack high-quality, open and fully functional tools for automatic text processing. Author's dissertation responds to this challenge by offering a comprehensive solution covering all major levels of linguistic analysis, from morphology to semantics.

Within the framework of the dissertation, a full stack of intelligent modules was developed: morphological analysis and generation, partial markup (POS), named entity recognition (NER), semantic role labeling (SRL), coreference resolution. Special attention is paid to the adaptation of modern neural network architectures such as KazBERT, BiLSTM+CRF and transformer models to the structure of the Kazakh language. The author's morphological generator is implemented taking into account the root structure, suffixation, synharmonism and morphological rules.

Special mention should be made of the original development – the QNLP library, which is a modular, extensible and open system for complex NLP processing of Kazakh texts. This library is already used in a number of educational and research projects, and also demonstrates the potential for industrial applications in the tasks of intellectual analysis of texts in the Kazakh language.

The dissertation is characterized by high scientific and methodological elaboration. The author independently completed all the stages of the research – from the collection and preprocessing of the text corpus to training, validation and comparative analysis of models. All built modules have been tested on a specially marked case, and the results have been evaluated using modern metrics (F1, Accuracy). The work meets the requirements of scientific integrity, and key results are presented in publications in international scientific journals and conferences.

As a foreign consultant, I can confidently note the high level of independence, academic maturity and research thinking of the applicant. Aitim

A.K. demonstrated a deep understanding of both the linguistic and computational aspects of the problem under study. Her ability to integrate knowledge in the field of applied linguistics and artificial intelligence in practice meets the best international standards for the training of PhD researchers.

Considering the above, I believe that the doctoral thesis "**Models and methods for the automatic processing of unstructured information**" is a completed independent research work that meets all the requirements of the dissertations in this specialty, and its author is **Aitim Aigerim Kairatkyzy** deserves to be awarded the degree of Doctor of Philosophy (PhD) in the EP: 8D06101 – "Clever Systems".

Foreign consultant,
Doctor of Technical Sciences, Professor,
Faculty of Electrical Engineering and Computer Science
Lublin University of Technology,
Poland



W. Wojcik

POLITECHNIKA LUBELSKA
Katedra Elektroniki i Technik Informacyjnych
ul. Nadbystrzycka 39A, 20-618 Lublin
tel. 81 538 43 09, fax 81 538 43 12