



## PhD thesis review report

Review report for the PhD thesis entitled “Open Vocabulary Model for Kazakh Language using Deep Neural Networks”, submitted by SULTANOVA NAZERKE ZHOLDYBAYEVNA, from Suleyman Demirel University, Kazakhstan.

The thesis proposes a language generation model for Kazakh language. The document starts with a review of language models, namely the statistical approach and the approach based on artificial neural networks. From there it develops to describe experiments with character-based model, based on a long short-term memory network. From there, the document proceeds to describe different text analysis methods, including part of speech analysis, word stemming, text classification and sentiment analysis. In the last chapter, a model is described, using a deep neural network with an attention mechanism, for Kazakh language.

In general the document reads well. The topics are well introduced, the methods are sufficiently described, the results are coherent and there is sufficient depth in the analysis. A number of different techniques are used, from text analysis and text generation, and they were chosen adequately for the purpose.

The latest version I have received needs some proof-reading. The main results and contributions of the work could be discussed in more detail, comparing to the state of the art, as well as highlighting relevant author's publications.

As a general comment, the work is specially difficult and relevant in the context where the tools for the text analysis and generation in Kazakh language are still in the initial stages of development — even though they are quite mature for other languages, application in Kazakh as a generative language is not straightforward, as highlighted in the document. In the context, my opinion is that the work has a merit, considering the novelty, the diversity of experiments and the utility as precursor of future work in Kazakh natural language processing.

*Coimbra, Portugal*

A handwritten signature in blue ink, appearing to read 'M. Mendes', written over a horizontal line.

Mateus Mendes

*(Adjunct Professor at the Polytechnic Institute of Coimbra,  
Researcher at the Institute of Systems and Robotics of the University of Coimbra)*

## Отчет о рецензии на докторскую диссертацию

Обзорный отчет по кандидатской диссертации на тему "Модель открытого словаря для казахского языка с использованием глубоких нейронных сетей", представленной СУЛТАНОВОЙ НАЗЕРКЕ ЖОЛДЫБАЕВНОЙ из Университета имени Сулеймана Демиреля, Казахстан.

В диссертации предлагается модель языковой генерации для казахского языка. Документ начинается с обзора языковых моделей, а именно статистического подхода и подхода, основанного на искусственных нейронных сетях. Оттуда он развивается для описания экспериментов с символьной моделью, основанной на сети долговременной кратковременной памяти. Далее в документе описываются различные методы анализа текста, включая анализ части речи, выделение слов, классификацию текста и анализ настроений. В последней главе описана модель, использующая глубокие нейронные сети с механизмом внимания, для казахского языка.

В целом документ читается хорошо. Темы хорошо представлены, методы достаточно описаны, результаты последовательны и анализ достаточно глубок. Используется ряд различных методов, начиная с анализа текста и генерации текста, и они были выбраны адекватно для этой цели.

Последняя полученная версия нуждается в корректуре. Основные результаты и вклад в работу можно было бы обсудить более подробно, сравнив их с современным уровнем, а также выделив соответствующие публикации автора.

В качестве общего замечания следует отметить, что работа особенно сложна и актуальна в контексте, когда инструменты для анализа и генерации текста на казахском языке все еще находятся на начальных стадиях разработки - даже несмотря на то, что они достаточно развиты для других языков, применение казахского языка в качестве продуктивного языка не является простым, как подчеркивается в документе. В этом контексте я считаю, что работа имеет свои достоинства, учитывая новизну, разнообразие экспериментов и полезность в качестве предшественника будущей работы по обработке казахского естественного языка.

Коимбра, Португалия

/подписано/  
Матеус Мендес

(Адъюнкт-Профессор Политехнического института Коимбры, научный сотрудник Института Систем и Робототехники Университета Коимбры)

Город Алматы Республика Казахстан. Шестнадцатое ноября две тысячи двадцать первого года.

Перевод документа с **английского языка на русский язык** выполнен агентством переводов ИП «Enjoy Service», в лице представителя **Биржанова Алмаса Болатулы**, ИИН 920916350701, в соответствии Сертификата «Переводческое дело» №003 от 04.12.2014 г.

Подпись: Биржанов Алмас Болатулы

Город Алматы Республика Казахстан.

Шестнадцатое ноября две тысячи двадцать первого года.

Я, Болкебаев Мырзагалий Курманбаевич, нотариус города Алматы, государственная лицензия №0001124 от 07 марта 2008 года, выданной Министерством юстиции Республики Казахстан, **свидетельствую подлинность подписи** сделанной переводчиком **Биржановым Алмасом Болатулы**, с **английского языка на русский язык**.

Личность переводчика установлена, дееспособность и полномочия проверены.



Зарегистрировано в реестре за №  
Взыскано: согласно ст.617 НК РК и 30-1 Закона  
Республика Казахстан «О нотариате».

Нотариус:

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
 АРДАҚТЫҚ ҚАҒАЗАТ  
 АКАДЕМИЯСЫ



ПРОШИТО И ПРОЧУМЕНО  
 ТИЛІГЕН ЖӘНЕ НОМІРЛЕНГЕН  
 ЛИСТОВ

317px

