



**PhD thesis review report**

Coimbra, Portugal, 18 September 2023

Review report for the PhD thesis, by Bekarystankyzy Akbayan, from Satbayev University, Almaty, Kazakhstan, for the PhD in Management information systems

The thesis analyses, discusses and proposes models automatic speech recognition, in Kazakh and other agglutinative languages. The document reviews the different language models, different models for automatic speech recognition, as well as the problems and challenges associated with agglutinative languages and lack of resources for training large machine learning models.

The state of the art, presented in Chapter 1, is fair enough, both in length and depth, giving a sufficiently clear picture of the latest developments in the field, along with their limitations and advantages.

The theoretical foundations, presented in Chapter 2, are also clear and comprehensive, both in terms of topics covered and depth of the analysis. Older and newer models are described, including the working principles and sometimes even important implementation details are hinted.

Regarding the experimental work, described in Chapter 3, it was very extensive and approaches the topic from a wide variety of angles. The PhD candidate not only searched adequate datasets but also collected additional data. The data were also pre-processed and transformed as necessary to fit the desired goals and suit the machine learning models being used. The choice of models was adequate for the goals of the project. The models were also hacked and adapted in a way that was adequate to the purpose. The PhD candidate also showed adequate mastery of the research methodology, correctly following the steps of the scientific method and data analysis projects, while still re-planning when necessary.

The results obtained are in line with what could be expected for the project and in many aspects the students' proposals achieved results above the state of the art, therefore adding valid contributions to the scientific community.

In general the document reads well and is well organised. 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 were described. Those techniques were chosen wisely and adequately used for the purpose.

As a second general comment, the work is specially difficult and relevant in a context where the tools for text analysis and classification in Kazakh language are still in the initial stages of development – even though they are more mature for other languages, application in Kazakh as an agglutinative low-resource language is not straightforward, as highlighted in the document. In this context, my opinion is that the work has merit, considering the novelty, the quality and diversity of experiments and the utility as precursor of future work in Kazakh natural language processing. Therefore, I recommend the candidate Bekarystankyzy Akbayan for a PhD degree in the field of 8D06103 – Management information systems.

Mateus Mendes  
(Adjunct Professor, Polytechnic University of Coimbra)

Assinado por: **MATEUS DANIEL ALMEIDA MENDES**  
Num. de Identificação: 11238412  
Data: 2023.09.20 10:58:03 +0100



Коимбрский политехнический университет

Отзыв на докторскую диссертацию

Коимбра, Португалия, 18 сентября 2023 года

Отзыв на докторскую диссертацию Бекарыстанкызы Акбаян из Университета имени Сатпаева, Алматы, Казахстан, на соискание ученой степени доктора философии в области Информационных систем управления.

В диссертации анализируются, обсуждаются и предлагаются модели автоматического распознавания речи на казахском и других агглютинативных языках. В документе рассматриваются различные языковые модели, различные модели для автоматического распознавания речи, а также проблемы и трудности, связанные с агглютинативными языками и недостатком ресурсов для обучения больших моделей машинного обучения.

Обзор состояния дел, представленный в Главе 1, достаточно полно, как по объему, так и по глубине, дает достаточно четкое представление о последних достижениях в данной области, а также об их ограничениях и преимуществах.

Теоретические основы, представленные в Главе 2, также отличаются ясностью и полнотой, как по охваченным темам, так и по глубине анализа. Описаны старые и новые модели, включая принципы работы, а иногда даже дается указание на важные детали реализации.

Что касается экспериментальной работы, описанной в Главе 3, то она получилась очень объемной и подходит к теме с самых разных сторон. Кандидат на степень доктора философии не только искала подходящие наборы данных, но и собирала дополнительные данные. Данные также подвергались предварительной обработке и преобразованию, что было необходимо для достижения поставленных целей и соответствия используемым моделям машинного обучения. Выбор моделей был адекватен целям проекта. Модели также были разобраны и адаптированы в соответствии с поставленной целью. Кандидат на степень доктора философии также продемонстрировала адекватное владение методологией исследования, правильно выполняя этапы научного метода и проекты анализа данных, при этом при необходимости перепланируя их.

Полученные результаты соответствуют тому, что можно было ожидать от проекта, и во многих аспектах предложения студента позволили достичь результатов, превышающих уровень техники, что вносит весомый вклад в научное сообщество.

В целом документ хорошо читается и хорошо организован. Темы хорошо представлены, методы описаны достаточно полно, результаты последовательны и достаточно глубоки в анализе. Описано несколько различных методик. Эти методики были выбраны грамотно и адекватно использованы для поставленной цели.

В качестве второго замечания общего характера следует отметить, что работа представляет особую сложность и актуальность в условиях, когда инструменты анализа и классификации текстов на казахском языке находятся на начальной стадии развития, хотя для других языков они более совершенны, применение их на казахском языке как агглютинативном малоресурсном языке не является простым, что подчеркивается в документе. В связи с этим я считаю, что работа заслуживает внимания, учитывая новизну, качество и разнообразие экспериментов, а также полезность в качестве предвестника будущих работ по обработке казахского естественного языка. Поэтому я рекомендую кандидата Бекарыстанкызы Акбаян на соискание ученой степени доктора философии по специальности 8D06103-Информационные системы управления.

Матеус Мендес  
(адъюнкт-профессор, Коимбрский политехнический университет)

Подписано: Матеус Даниэль Алмейда Мендес  
Идентификационный номер: 11238412  
Дата: 2023.09.20 10:58:03+0100

Текст-перевод документа с английского языка на русский язык выполнен мною, переводчиком, Журтыбаевой Гулназ Мухамедалиевной.

Имя переводчика Журтыбаева Гулназ Мухамедалиевна

Подпись \_\_\_\_\_

Республика Казахстан, город Алматы, девятое октября две тысячи двадцать третьего года. Я, Абылкасымова Айман Турехановна, нотариус города Алматы, действующий на основании государственной лицензии № 0000189 выданной Министерством юстиции Республики Казахстан от 10.08.1998 года, свидетельствую подлинность подписи переводчика Журтыбаевой Гулназ Мухамедалиевной. Личность переводчика установлена, дееспособность и полномочия проверены.

Зарегистрировано в реестре: 2883

Взыскано: согласно Закону РК «О нотариате»

Подпись нотариуса \_\_\_\_\_

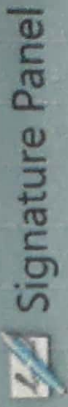


ES7201553231009154718K551967

Нотариаттық іс-әрекеттің бірегей нөмірі / Уникальный номер нотариального действия



Signed and all signatures are valid.



Signature Panel



Signatures



Validate All



**Rev. 1: Signed by MATEUS DANIEL ALMEIDA MENDES**

Signature is valid:

Document has not been modified since this signature was applied

Signer's identity is valid

Signing time is from the clock on the signer's computer.

Signature is not LTV enabled and will expire after 2030/02/11 15:56:49 +06'00'



Signature Details

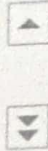
Last Checked: 2023.09.22 07:56:29 +06'00'

Field: Signature3149346453 on page 1

[Click to view this version](#)

Signed and all signatures are valid.

## Signatures



Validate All

**Rev. 1: Signed by MATEUS DANIEL ALI**

Signature is valid:

Document has not been modified since

Signer's identity is valid

Signing time is from the clock on the

Signature is not LTV enabled and will

Signature Details

Last Checked: 2023.09.22 07:56:29 +06

Field: Signature3149346453 on page 1

[Click to view this version](#)

Signature Panel



PhD thesis review report

Coimbra, Portugal, 28 September 2023

Review report for the PhD thesis, by Bekaryntynlyz Alabayeva, from Soboyev University, Almaty, Kazakhstan, for the PhD in Management information systems

The thesis analyses, discusses and proposes models automatic speech recognition, in Kazakh and other agglutinative languages. The document reviews the different language models, different models for automatic speech recognition, as well as the problems and challenges associated with agglutinative languages and lack of resources for training large machine learning models.

The state of the art, presented in Chapter 1, is fair enough, both in length and depth, giving a sufficiently clear picture of the latest developments in the field, along with their limitations and advantages.

The theoretical foundations, presented in Chapter 2, are also clear and comprehensive, both in terms of topics covered and depth of the analysis. Older and newer models are described, including the working principles and sometimes even important implementation details are named.

Regarding the experimental work, described in Chapter 3, it was very extensive and approaches the topic from a wide variety of angles. The PhD candidate not only searched adequate datasets but also collected additional data. The data were also pre-processed and transformed as necessary to fit the desired goals and suit the machine learning models being used. The choice of models was adequate for the goals of the project. The models were also trained and adapted in a way that was adequate to the purpose. The PhD candidate also showed adequate mastery of the research methodology, correctly following the steps of the scientific method and data analysis project, while still re-planning when necessary.

The results obtained are in line with what could be expected for the project and in many aspects the student's proposals achieved results above the state of the art, therefore adding valid contributions to the scientific community.

In general the document reads well and is well organized. 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 were described. Those techniques were chosen wisely and adequately used for the purpose.

As a second general comment, the work is specially difficult and relevant in a context where the tools for text analysis and classification in Kazakh language are still in the initial stages of development - even though they are more mature for other languages, application in Kazakh as an agglutinative language resource is not straightforward, as highlighted in the document. In this context, my opinion is that the work has merit, considering the novelty, the quality and diversity of experiments and the utility as precursor of future work in Kazakh natural language processing. Therefore, I recommend the candidate Bekaryntynlyz Alabayeva for a PhD degree in the field of 8206103 - Management information systems.

Mateus Mendes

(Adjunct Professor, Polytechnic University of Coimbra)

Almaty (e-mail): MATEUS.DANIEL.ALI@FEUC.PT

MATEUS

Name: Mateus Mendes

ID: 20230928103000000000

DATE: 2023/09/28 10:56:29

FEUC

FEUC