The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2023

Volume 26, Pages 633-640

IConTES 2023: International Conference on Technology, Engineering and Science

Using RDF Models to Create Knowledge Bases in the Kazakh Language: Comparison with Other Methods

Assel Mukanova
Astana International University

Gulnazym Abdikalyk
Astana International University

Aizhan Nazyrova
Astana International University

Assem Dauletkaliyeva
Astana International University

Abstract: Currently, there is a rapid development of information technologies, the amount of information on the Internet is growing very fast and it is becoming increasingly difficult to find the necessary information. A search using keywords does not give results adequate to the meaning of the information sought. Therefore, the creation of a technology for designing intelligent question answering systems in the Kazakh language based on the presentation, processing and extraction of knowledge is a very actual problem, since it is in such a system that the linguistic and semantic relationships between the texts of the request and the answer can be taken into account. This research paper focuses on the integration of the Resource Description Framework (RDF) model, a semantic web technology, and provides a detailed evaluation of data mining techniques in Kazakh. The paper examines many Kazakh language data collection methods such as online scraping, community collaboration and translation. It also explores the function of RDF models in organizing knowledge, connecting data points and adding semantic richness to datasets. The paper discusses linguistic features and challenges unique to the Kazakh language and emphasizes the need to address these challenges with domain-specific data. The need for thorough cleaning, annotation and data quality assurance is emphasized to guarantee the reliability and use of the collected datasets. Within global communications and technology, the study emphasizes the importance of languages other than English and examines how semantic web technologies can improve data representation and knowledge retrieval. The study lays the groundwork for future initiatives to address the shortage of datasets in languages with fewer resources and to create semantic web technologies for language diversity.

Keywords: Resource description framework (RDF) model, Question-Answering system, Ontology model, Knowledge base

Introduction

Recent advances in machine learning and deep learning have led to significant progress in natural language processing, or NLP. The availability of high-quality datasets that serve as the basis for training and evaluating language models and applications is essential for the development of natural language processing (NLP) (Tang et al., 2021). Underrepresented languages such as Kazakh often face a lack of data, while dominant languages such as English, Chinese, and Spanish benefit from huge databases and linguistic resources (Zhubanov, 2018).

In order to shed light on the opportunities and challenges faced when collecting Kazakh language data sets, this study aims to explore the different approaches used in this process. In particular, it is about integrating structured data representation and semantic enrichment using the Resource Description Framework (RDF) paradigm, a semantic web technology (Rubiera et al., 2012). A variety of datasets are needed in linguistic research, machine learning, and NLP applications for model training, language understanding, and cross-linguistic research. However, due to data limitations, lack of digital material, and specific linguistic features, underrepresented languages such as Kazakh face particular challenges.

This study reviews the methods currently used to collect Kazakh datasets, such as online scraping, community collaboration and translation, and critically analyzes the function of RDF models in providing semantic enrichment and interconnectivity of data points. A more efficient knowledge organization is made possible by the structured representation of data in RDF, which also enables linking data points and sophisticated information retrieval (Decker et al., 2000).

To address the problem of data scarcity in under-represented languages such as Kazakh, this paper explores the use of representational density models (RDF). This paper emphasizes the importance of linguistic diversity and language representations beyond English in communication and technology. It lays the groundwork for future NLP research that will focus on structured and semantically enriched datasets. As NLP research develops, it will also contribute to linguistic diversity.

Research Methods

Modeling the Semantic Web Using RDF

In the field of semantic web, the Resource Description Framework (RDF) is a fundamental technology. Its ability to simplify the representation of structured data has had a great impact on many different areas, such as the development of structured databases for underrepresented languages such as Kazakh. In this part, we will look at the features, principles and applications of RDF, highlighting how important this technology is for data organization and language diversity.

RDF: A Framework for Representing Structured Data

At its core, RDF is an adaptable and extensible framework for structuring and semantically meaningful expression of information on the Web. RDF, unlike conventional databases, allows data to be represented as linked triples, each consisting of three elements: object, predicate and subject. By establishing associations between things, these triples allow data to be conceptualized semantically (Heath & Bizer, 2022). A resource is represented by the subject, its relation or attribute is labeled by the predicate, and its value or target resource is identified by the object. RDF is an effective tool for encoding, connecting and analyzing data in a way that conveys not only the information itself, but also its meaning and context because it is based on this triple structure, which serves as the basis for the semantic representation of RDF (Prud'hommeaux & Seaborne, 2008).

Interoperability and Semantic Structure

The strength of RDF lies in its ability to provide semantic context to data. RDF builds a semantic network in which things are connected by clear links by storing information in the form of trills. This semantic structure provides several advantages for data representation.

- Interoperability: RDF allows easy integration of data from different sources. The overall structure of RDF allows the integration and searching of different datasets, each represented within it. Since interoperability offers some structure for data integration, it is particularly useful when dealing with multilingual datasets (Bizer et al., 2023).
- Semantic enrichment: RDF allows the addition of meanings to data. RDF data can be enriched with meaning through the use of ontologies and controlled vocabularies, allowing for more precise and context-aware searches. Semantic enrichment can be used to reflect the subtleties and cultural features of the Kazakh language in the context of Kazakh datasets.

• Structured representation: RDF offers a structured data representation that allows semantic analysis and cross-lingual comparisons, making it useful for Kazakh datasets in Natural Language Processing (NLP) tasks due to its ability to provide a semantically meaningful representation.

RDF Regarding Kazakh Datasets

RDF models greatly facilitate NLP research using Kazakh language datasets. These models reflect the subtle cultural and linguistic aspects of the Kazakh language, improving data quality and facilitating interaction with other resources. By allowing NLP applications to access structured data, RDF models promote linguistic diversity. RDF preserves the semantic richness of data through the use of controlled vocabularies and domain-specific ontologies, allowing for more accurate and context-aware processing. The semantic web-based RDF model is a powerful tool that can be used to create structured datasets in Kazakh, to provide context-rich and structured data representation, and to promote linguistic diversity in NLP applications and research.

The Value of Non-English QA Datasets

High quality datasets play an important role in the development of Natural Language Processing (NLP) technology, which has advanced significantly in recent years. However, linguistic diversity and representation is becoming increasingly important for the development and relevance of this field. One language that is still underrepresented in the global digital arena is Kazakh, which is the official language of Kazakhstan and spoken by millions of people throughout Central Asia. Due to the lack of high-quality datasets, the Kazakh language is still neglected by NLP researchers, despite its importance to the region and its cultural heritage.

Beyond English: Kazakh QA Data is Needed

The importance of non-English QA data, like Kazakh, is in highlighting the world's linguistic diversity. Millions of people speak Kazakh, a Turkic language that is widely spoken and the official language of Kazakhstan. Kazakh language datasets are created with consideration for the language's variety as well as its distinct cultural, historical, and contextual elements. The development of Kazakh QA datasets contributes to a more comprehensive, inclusive and integrated natural language processing (NLP) ecosystem by acting as a model for the development of comparable resources in other non-English languages. The NLP community's dedication to fairly serve people worldwide is demonstrated by the development of multilingual QA datasets, particularly in Kazakh. These databases foster inclusion and creativity in addition to maintaining language.

Lack of Up-To-Date Data: A Major Challenge

While it is a laudable goal to create excellent Kazakh-language question answering (QA) datasets, there are a number of difficulties. The lack of available data in Kazakh language is one of the main obstacles faced by researchers in this field (Hao & Agichtein, 2012). Kazakh language does not have a data corpus similar to English, which is rich in textual data including QA pairs, making it unsuitable for creating QA data. The importance of this issue cannot be overemphasized, as the process of creating reliable Kazakh QA data is complicated by the lack of available data.

Another major challenge in creating Kazakh QA datasets is the translation of existing English-language QA datasets into Kazakh. Although translation is a common method of extending QA sets into non-English languages, it is not without drawbacks. In the process of translating English QA pairs into Kazakh, linguistic nuances, cultural context and contextual variations present in the original language may be inadvertently excluded, reducing the overall quality and efficiency of the dataset (Diab & Resnik, 2002).

Overview of Data Collection Methods

This section reviews various methods of data collection in Kazakh language, in particular web scraping, manual translation, crowdsourcing and innovative methods. It presents an in-depth analysis of the advantages and limitations of each method, emphasizing the importance of linguistic diversity and supporting research and applications in the field of natural language processing (NLP) in Kazakh.

Inclusivity

One of the frequently used methods for data collection in Kazakh language is web scraping. With this method, text and information are automatically extracted from various internet sources such as blogs, news portals, forums and social networks. Considering Kazakh data sets, some advantages and disadvantages of online scraping should be noted (Finkel et al., 2005).

Advantages
• Rich and up-to-date material: Web scraping provides access to a large amount of relevant and diverse material, which keeps datasets up-to-date. This is particularly useful for applications such as sentiment analysis and event monitoring that need up-to-date data (Callison-Burch et al., 2011).
• Large scale data collection: Automated web scraping allows for the rapid collection of huge amounts of data, making it a useful tool for creating the huge data sets needed to train large NLP models (Bird et al., 2006).

Limitations
• Legal and ethical issues: When it comes to copyrighted or proprietary material, both legal and ethical issues can arise in web scraping. Respecting the rights and limitations of sources, the terms of service and intellectual property rights must be carefully considered (Spalka, 2015).
• Data Quality: Due to unstable formatting and unstructured content, raw data obtained through web scraping often needs careful cleaning and preparation. It is not easy to ensure the accuracy and high quality of the collected data (Baeza-Yates & Ribeiro-Neto, 2011).

Manual Translation

Manual translation involves the use of human translators to create new data entirely in Kazakh or to translate existing information into that language. When it comes to Kazakh data sets, manual translation has a unique combination of advantages and disadvantages (Lewis, 1992).

Advantages
• Accuracy and quality: The use of human translators ensures accurate translations that convey the subtleties and unique cultural characteristics of the Kazakh language. This is particularly useful for initiatives that need to be culturally sensitive and accurate in wording (Bird, 2006).

Limitations
• Time and resource costs: Manual text translation can be laborious and resource intensive, especially when dealing with large data sets. It increases the cost and duration of data collection as it requires skilled translators who are fluent in both the source and target language (Snow et al., 2008).

Online Auctions

Crowdsourcing is a method of creating or translating materials that utilizes the collaborative efforts of many people, sometimes with different language backgrounds. There are some advantages and disadvantages of using crowdsourcing when dealing with Kazakh datasets (Ipeirotis, 2010).

Advantages
• Scalability: Crowdsourcing is a very scalable technology that allows for the rapid creation of large datasets. This is especially effective for initiatives where large amounts of data need to be collected in a short period of time.

Limitations
• Semantic Correctness: The semantic correctness of crowd-sourced translations may be inferior to translations produced by expert translators. Ensuring that crowd-sourced translations adequately convey the desired meaning is an ongoing challenge that requires careful supervision and quality control procedures (Callison-Burch et al., 2011).

Other Innovative Techniques

To overcome the limitations of the lack of linguistic resources, new approaches including data augmentation, transfer learning and parallel corpora are being used to collect Kazakh language data. Data augmentation is the process of adding new data to existing datasets through transformations such as data synthesis, text expansion and paraphrasing (Wei et al., 2017). Transfer learning improves NLP tasks in underrepresented languages, such as Kazakh, by utilizing already existing models and information from well-resourced languages (Peters et al., 2018). Parallel corpora in similar languages such as Turkish or Uzbek can be used to improve Kazakh datasets (Bekarystankyzy et al., 2023). This would allow cross-lingual analysis and adaptation of NLP models to the Kazakh language.

Related Works

Interest in collecting data sets in the Kazakh language is growing, and researchers are considering a number of strategies to overcome the difficulties associated with resource creation and data collection. The authors of the article are engaged in the field of text processing in the Kazakh language. In (Mukanova et al., 2014) (Yergesh et al., 2014) a semantic hypergraph was used to describe ontological models of morphological rules of the Kazakh language. In (Yelibayeva et al., 2020) describes the metalanguage of morphological concepts of the Turkic languages for the creation of knowledge bases. In articles (Sharipbay et al., 2019) (Yelibayeva et al., 2020) (Yelibayeva et al., 2022) describes the methods of syntactic analysis of the text in the Kazakh language based on the ontological model.

Other related publications provide insights into the various approaches, challenges and uses associated with data collection in the Kazakh language. One of them builds on a BERT-like concept to create an extractive question-answer system using Google Cloud Translation API and a Kazakh question-answer dataset (KazQA). The ability of the system to generate Kazakh language question-answer systems with few resources is demonstrated using ALBERT and multilingual BERT as base models (Shymbayev & Alimzhanov, 2023). The following paper proposes a modular approach to question analysis using rule-based methods and Hidden Markov Models (HMM) to generate a question-answering (QA) system in Kazakh language. Using dependency relations between words, the method integrates question classifiers, focus extraction and system classification (Rakhimova et al., 2021).

There are methods using seq2seq approach to collect 60,000 corpora in Kazakh language (Rakhimova et al., 2022). There is a large number of works devoted to the creation of QA data in Kazakh language. But despite this, the Kazakh language with its rich history and multiple uses is widely spoken and studied, which makes it an invaluable resource for scholars and specialists.

Results and Discussion

One of the main parameters of text analysis for understanding the meaning of a sentence is to determine the properties of words in the text and the relationships between them. In our project, ontology was used to represent domain knowledge. It reflects the concepts of the subject area and the semantic relations between them. The semantic relation is binary.

The OWL language is used to describe ontological knowledge. This language is designed not only to represent domain knowledge, but also for applications that allow you to process this knowledge. This language uses RDF resources as its main elements. An RDF resource can be expressed as a triplet: subject – predicate - object." Then if we take the vertices of the ontology as subject and object, the relationship or property between them will be predicates.

In the Kazakh language, phrases have 4 types of communication (Kiysu (Kiysu), Mengeru (MengEt), Matasu (MatEs), Kabysu (KabEs)).These structures are created by combining two or more words. Phrases can have a head word (head) and a dependent word (dep). Words in a phrase can be a concept of a subject area, a relation or a property between them. Therefore, using formal rules (Yelibayeva . et al., 2019) constructing phrases in the Kazakh language, consider the fragment rules for converting words in the text into concepts and knowledge base relationships in the form of RDF (Table1). Using the SPARQL language to query in our knowledge base, it is possible to extract answers to questions and analyze data in Kazakh through these rules.

Table 1. Fragment rules for converting words in the text into concepts and knowledge base relationships in the form of RDF

Types of communication	Dependent word dep	Head word head	Subject S	Predicate P	Object O
KabEs	Adj	Dom	dom	? (hasType)	dep
KabEs	Adj	Dom	dom	? (hasClass)	dep
KabEs	Adj	Dom	dom	? (hasProperty)	dep
KabEs	Num	Dom	dom	? (hasNuber)	dep
KabEs	Num	dom	dom	? (hasOrder)	dep
MatEs	dep	dom	dep	dom	?
Kiysu	dep	dom	dep	?	dom
Kiysu	dep	V	dep	dom	?
MengEt	dep	V	dep	dom	?

Conclusion

This study examines different approaches to the collection of Kazakh datasets with a focus on the use of RDF models. The article discusses the methods of applying semantic models of RDF triplets depending on the type of phrases in the sentence, developed formal rules. Based on this, the question and answer are compared in accordance with the knowledge available in the knowledge base. The study sheds light on the evolution of linguistic diversity and highlights the revolutionary potential of structured data in Natural Language Processing (NLP) research and applications. Linked Data principles and the implementation of the RDF model pave the way for a more diverse and international NLP community. The study also emphasizes the need to promote underrepresented languages to guarantee their participation in the digital age. This study reminds us that language diversity is a critical resource as NLP technologies evolve. We must work together to protect and preserve these multiple language ecosystems so that no language is lost in the digital age. The findings of the study and the structured data it collects contribute to the goal of making NLP a truly inclusive and global field. We hope that this study will serve as a springboard for further research, partnerships and projects to support linguistic diversity and the inclusion of all languages in the global semantic network of the future.

Future Recommendations

The study focuses on RDF models in exploring different approaches to collecting Kazakh language datasets. It emphasizes the need for further recommendations and makes suggestions for future approaches to maintain the sustainability of structured data in underrepresented languages such as Kazakh and improve linguistic diversity. Domain-specific Kazakh language ontologies are needed to organize data in specialized industries such as healthcare, banking, and law. Creating strong ontologies requires the collaborative efforts of experts, linguists and engineers. Modeling RDF in Kazakh requires standardization at the community level to ensure that data representations conform to common practices and rigorous data standards.

Adding Kazakh datasets to the global Linked Data ecosystem can foster the development of multilingual applications, facilitate cross-language information retrieval, and increase the availability of Kazakh-language material on the Semantic Web. Crowdsourcing methods can be used to accelerate the collection and enrichment of Kazakh data, especially in areas where specialized knowledge is required. In addition to solving the problem of data scarcity, this can facilitate community engagement. The burden of human quality control can be reduced by using machine learning and artificial intelligence to clean and validate data. The goal of this effort is to improve the visibility of Kazakh language information in the Semantic Web and to create a more complete, connected knowledge network.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Acknowledgements

* This article was presented as an oral presentation at the International Conference on Technology, Engineering and Science (www.icontes.net) held in Antalya/Turkey on November 16-19, 2023.

* This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19577922)

References

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval* (Vol. 463, No. 1999). New York, NY: ACM Press.

Bekarystankyzy, A., Mamyrbayev, O., Mendes, M., Oralbekova, D., Zhumazhanov, B., & Fazylzhanova, A. (2023). Automatic speech recognition improvement for Kazakh language with enhanced language model. In *Asian Conference on Intelligent Information and Database Systems* (pp. 538-545). Cham: Springer Nature.

Bird, S. (2006). NLTK.: The natural language toolkit. In *Proceedings of the COLING/ACL on Interactive Presentation Sessions*, 69-72.

Bird, S., Loper, E., & Klein, E. (2009). *Natural language processing with Python.* Sebastopol, USA: O'Reilly Media Inc

Bizer, C., Heath, T., & Berners-Lee, T. (2023). Linked data-the story so *far. In Linking the World's Information: Essays on Tim Berners-Lee's Invention of the World Wide Web,* 115-143.

Callison-Burch, C., Koehn, P., Monz, C., & Schroeder, J. (2011). Findings of the 2011 workshop on statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation* (WMT *'11)* (pp.22-64).

Callison-Burch, C., Koehn, P., Monz, C., & Schroeder, J. (2011). Findings of the 2011 workshop on statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation* (WMT *'11)* (pp.22-64).

Decker, S., Mitra, P., & Melnik, S. (2000). Framework for the semantic Web: An RDF tutorial. *IEEE Internet Computing*, *4*(6), pp. 68-73.

Diab, M., & Resnik, P. (2002). An unsupervised method for word sense tagging using parallel corpora. In Proceedings of *the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 255-262).

Finkel, J. R., Grenager, T., & Manning, C. D. (2005, June). Incorporating non-local information into information extraction systems by gibbs sampling. *In Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)* (pp. 363-370).

Hao, T., & Agichtein, E. (2012). Finding similar questions in collaborative question answering archives: toward bootstrapping-based equivalent pattern learning. *Information Retrieval*, *15*, 332-353.

Heath, T., & Bizer, C. (2022). *Linked data: Evolving the web into a global data space* (pp.1-136). Springer Nature.

Ipeirotis, P. G. (2010). Analyzing the amazon mechanical Turk marketplace. *XRDS: Crossroads, The ACM magazine for students*, *17*(2), 16-21.

Lewis, D. D. (1992, June). An evaluation of phrasal and clustered representations on a text categorization task. In *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 37-50).

Mukanova, A., Yergesh, B., Bekmanova, G., Razakhova, B., & Sharipbay, A. (2014). Formal models of nouns in the Kazakh language. *Leonardo Electronic Journal of Practices and Technologies*, *13*(25), 264-273.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American *Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)* (pp. 2227-2237).

Prud'hommeaux, E., & Seaborne, A. (2008). SPARQL query language for RDF. *W3C Recommendation*.

Rakhimova D., Suleimenov Y. and Akhmet G. (2022). The task of synthesizing the Kazakh language based on the seq2seq approach for a question-answer system. *7th International Conference on Computer Science and Engineering (UBMK)* (pp. 289-293). IEEE.

Rakhimova, D., Khairova, N., Kassymova, D., & Janibekovich, K. (2021). Development of a system of questions and answers for the Kazakh language based on rule-based and HMM: development of a system of questions and answers for the kazakh language based on rule-based and HMM. *Advanced Technologies and Computer Science,* (2), 34–44.

Rubiera, E., Polo, L., Berrueta, D., & El Ghali, A. (2012, May). TELIX: An RDF-based model for linguistic annotation. *In Extended Semantic Web Conference* (pp. 195-209). Berlin, Heidelberg: Springer.

Sharipbay, A., Yergesh, B., Razakhova, B., Yelibayeva, G., & Mukanova A. (2019). Syntax parsing model of Kazakh simple sentences. *ACM International Conference on Data Science, E- Learning and Information Systems* (pp.1-5).

Shymbayev M., & Alimzhanov Y (2023). Extractive question answering for Kazakh language. *IEEE International Conference on Smart Information Systems and Technologies* (SIST) (pp. 401-405).

Snow, R., O'connor, B., Jurafsky, D., & Ng, A. Y. (2008, October). Cheap and fast–but is it good? evaluating non-expert annotations for natural language tasks. In *Proceedings of the 2008 conference on empirical methods in natural language processing* (pp. 254-263).

Spalka, A. (2015). Challenges of Web Scraping. Strohmaier, R., Schuetz, M., & Vannuccini, S. (2019). A systemic perspective on socioeconomic transformation in the digital age. *Journal of Industrial and Business Economics*, *46*(3), 361-378.

Tang, S. X., Kriz, R., Cho, S., Park, S. J., Harowitz, J., Gur, R. E., ... & Liberman, M. Y. (2021). Natural language processing methods are sensitive to sub-clinical linguistic differences in schizophrenia spectrum disorders. *NPJ Schizophrenia*, *7*(1), 25.

Wei, J., Zhao, Y., & Li, M. (2017). Neural machine translation with external phrase memory. In *Proceedings of* the *55th Annual Meeting of the Association for Computational Linguistics* (Vol.1, pp. 2181-2190).

Yelibayeva G., Mukanova A., Zhulkhazhav A, Razakhova B., & Sharipbay A. (2019). Combined morphological analyzer of the Kazakh language based on ontological modeling. *Human Language Technologies as a Challenge for Computer Science and Linguistics*, 241-244.

Yelibayeva, G., Mukanova, A., Sharipbay, A., Zulkhazhav, A., Yergesh, B., & Bekmanova G. (2019). Metalanguage and knowledgebase for Kazakh morphology. In *Computational Science and Its Applications- ICCSA 2019: 19 th International Conference, Saint Petersburg, Russia* (pp. 693-706). Springer International Publishing.

Yelibayeva, G., Razakhova, B., Yergesh, B., Sharipbay, A., Bekmanova, G., & Mukanova, A. (2022) Modeling of verb phrases of the Kazakh language. In *2022 International Conference on Engineering and MIS, ICEMIS 2022*. (pp.1-3). IEEE.

Yelibayeva, G., Sharipbay, A., Mukanova, A., & Razakhova, B. (2020). Applied ontology for the automatic classification of simple sentences of the Kazakh language *5th International Conference on Computer Science and Engineering (UBMK)* (pp.13-18). IEEE

Yergesh, B., Mukanova, A., Sharipbay A., Bekmanova G., Razakhova B. (2014). Semantic hyper-graph based representation of nouns in the Kazakh language. *Computacion y Sistemas, 18*(3), 627-635.

Zhubanov A.K. (2018). For digital Kazakhstan creation of the speaker corps of the Kazakh language is a topical problem. *Tiltanym, (4),* 45-50.

Author Information

Assel Mukanova
Astana International University
Kabanbay Batyra Avenue 8, Astana, Kazakhstan
Contact e-mail: *asiserikovna@gmail.com*

Gulnazym Abdikalyk
Astana International University
Kabanbay Batyra Avenue 8, Astana, Kazakhstan

Aizhan Nazyrova
Astana International University
Kabanbay Batyra Avenue 8, Astana, Kazakhstan

Assem Dauletkaliyeva
Astana International University
Kabanbay Batyra Avenue 8, Astana, Kazakhstan

To cite this article:

Mukanova, A., Abdikalyk, G., Nazyrova, A., & Dauletkaliyeva, A. (2023). Using RDF models to create knowledge bases in the kazakh language: comparison with other methods. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 26*, 633-640.