KyrgyzNER: The First NER Dataset for the Kyrgyz Language

Anonymous ACL submission

Abstract

We introduce KyrgyzNER, the first manually annotated named entity recognition dataset for the Kyrgyz language. Comprising 1,499 news articles from the 24.KG news portal, the dataset contains 10,900 sentences and 39,075 entity mentions across 27 named entity classes. We show our annotation scheme, discuss the challenges encountered in the annotation process, and present comprehensive corpus statistics. Our experiments with several NER models, including classical CRF-based approaches and state of the art pretrained multilingual models fine-tuned on KyrgyzNER, demonstrate that while all approaches struggle with underrepresented classes, models such as XLM-RoBERTa achieve a promising balance between precision and recall. These results highlight both the challenges and the potential of leveraging multilingual pretraining for low-resource languages; we note that while XLM-RoBERTa was best, all multilingual models achieved similar scores, indicating that further investigation and experiments with modified, "more atomic" annotation schemes might provide better insight model comparison for Kyrgyz language processing.

1 Introduction

Recent advances in machine learning and natural language processing (NLP) have been fueled by the availability of large, high-quality annotated datasets. However, the bulk of these resources have been developed for high-resource languages that have abundant linguistic data. *Less-resourced languages* (LRL, low-resource languages) are those spoken in the world but with fewer linguistic resources for language technologies (Cieri et al., 2016); they are significantly underrepresented in NLP research, and this imbalance not only limits the development of robust language technologies for low-resource language communities but also perpetuates inequalities in the access to cutting-edge NLP tools.

The emergence of multilingual language models such as BERT (Devlin et al., 2018) and XLM-RoBERTa (Xue et al., 2021a), trained on languages with varying amounts of resources, offers new possibilities for NLP in low-resource languages. These models allow knowledge transfer from well-resourced languages to underrepresented ones, providing a solid foundation for tasks such as *named entity recognition* (NER).

Kyrgyz is a prime example of a less-resourced language, with only a limited number of tools and datasets available for its processing. Developing manually annotated datasets for Kyrgyz is essential for evaluating and improving language models. While multilingual models continue to evolve, creating human-validated datasets remains a fundamental step in building reliable language resources. Even in the era of modern NLP, which is increasingly moving towards universal models such as LLMs (Brown et al., 2020; Achiam et al., 2023) and others, it is still difficult to envision progress without, at the very least, access to the evaluation data prepared and/or validated by humans.

NER is a core NLP task that involves identifying and classifying specific entities in text, such as names of people, locations, or organizations (Jurafsky and Martin, 2008), or other predefined domain-specific categories (Miftahutdinov et al., 2020). It is a fundamental component for applications such as information extraction, question answering, and conversational systems, playing a crucial role in extracting structured information from unstructured text data. Despite its importance, there are no high-quality manually annotated datasets for NER in Kyrgyz, which severely limits progress in this area.

In this work, we address this gap by presenting the first manually annotated Kyrgyz NER dataset, based on news articles from the 24.KG portal[1]. Our

[1]We use the data scraped from https://24.kg/ with the permission of the agency's editors to use the data for research purposes.

contributions are as follows: (1) we introduce a new dataset with 10,900 sentences and 39,075 entity mentions classified into 27 categories; (2) we evaluate multiple baseline models, including classical and mainstream modern NER approaches; (3) we establish a benchmark for Kyrgyz NER, providing a foundation for future research. The rest of the paper is organized as follows: Section 2 reviews related work, Section 3 presents the dataset and its annotation process, Section 4 reports corpus-related statistics, Section 5 describes the baseline models, our experimental setup, and experimental results, Section 6 investigates reoccurring mistake patterns, and Section 7 concludes the paper.

2 Related Work

2.1 Named Entity Recognition

Named entity recognition (NER) has evolved significantly over the years, with early approaches categorized into rule-based and machine learning-based methods. Rule-based methods (Hanisch et al., 2005; Quimbaya et al., 2016) rely on manually crafted rules and lexicons with entities extracted by substring matching; they offer high precision but suffer from limited adaptability and extensive manual labor required. Notable examples include systems such as FASTUS (Appelt et al., 1995), Lasie-II (Humphreys et al., 1998), NetOw1 (Krupka and Hausman, 1998), Facile (Black et al., 1998) and others. Machine learning-based methods treat NER as a sequence labeling problem, where large annotated corpora are used to train models that predict tags for each token in a text (Liu et al., 2022). Classical NER models employed hidden Markov models (HMM) (Eddy, 1996), maximum entropy models (Kapur, 1989), and conditional random fields (CRF) (Lafferty et al., 2001), which led to better generalization abilities compared to rule-based systems.

In recent years, deep learning architectures such as BiLSTM-CRF models (Vajjala and Balasubramaniam, 2022) have become a standard tool for NER, with bidirectional long short-term memory (LSTM) architectures capturing the context better (Huang et al., 2015; Ma and Hovy, 2016; Hochreiter and Schmidhuber, 1997) and CRF providing additional improvements in decoding the most probable label sequence (Qi et al., 2018). Transformer-based models, such as BERT (Devlin et al., 2019a) and its multilingual variants, have further advanced the field by providing contextual word representations that significantly improve performance across multiple languages. Recent innovations include GPT-based NER that reframes NER as a text generation task, achieving competitive results in few-shot settings (Wang et al., 2023); these models, however, lead to new challenges such as hallucinations and require the development of self-verification strategies.

Nested NER (Finkel and Manning, 2009) refers to the identification of entities that are hierarchically structured or can overlap within a single text, such as recognizing "The Chinese embassy in France" as both a facility entity and geographical entities like "Chinese" and "France". While Nested NER finds numerous applications in different domains (Wang et al., 2020; Loukachevitch et al., 2023) and represents a significant NLP challenge due to its complex annotation requirements, it falls outside the scope of this work, where we focus on "flat" NER for the Kyrgyz language.

2.2 Custom Datasets for Less-Resourced Languages

Custom datasets are crucial for addressing the unique linguistic features of less-resourced languages, and similar techniques have been used in other domains for similar problems. For instance, in the biomedical domain, specialized corpora for disease and drug entity recognition tailored to the unique terminologies and nomenclatures of the field have driven advances in biomedical text mining (Zhang and Wu, 2021). This approach of building domain-specific datasets mirrors the development of a custom Kyrgyz NER dataset, which addresses the lack of language resources by capturing the specific linguistic characteristics of Kyrgyz.

Several widely used datasets have significantly advanced NER research and applications. *CoNLL-2003*, a landmark dataset for NER tasks, features English and German texts with annotations for entities such as *Person*, *Location*, *Organization*, and *Miscellaneous*; it was derived from Reuters news stories and has been extensively utilized in NLP research. CoNLL-2003 is available via TensorFlow Datasets and Papers With Code (Tjong Kim Sang and De Meulder, 2003). *OntoNotes 5.0* encompasses annotations across multiple domains such as newswire, broadcast conversations, and telephone conversations; this dataset supports 18 entity types and provides an essential resource for diverse NLP applications (Pradhan et al., 2013). WNUT 2016/2017 was designed for entity recogni-

Dataset	Lang.	# of Sent.	# of Entity Classes	Sources
KyrgyzNER	Kyrgyz	10,900	27	News (24.KG)
Uzbek NER	Uzbek	1,160	6	News and social media
KazNERD	Kazakh	112,702	25	Wiki and news
Turkish Wiki NER	Turkish	20,000	3	Wiki
WikiANN	Multiple	Variable	3	Wiki (282 languages)

Table 1: Comparison of Turkic and related NER datasets.

tion in noisy user-generated text, particularly from social media platforms, making it indispensable for real-world NLP applications (Strauss et al., 2016). *WikiANN* is a large multilingual dataset annotated for named entities in over 282 languages, including support for less-resourced languages such as Uzbek, Turkish, Tatar, and Kazakh; it is a critical resource for low-resource NLP tasks, accessible via platforms such as Hugging Face and GitHub (Rahimi et al., 2019).

2.3 NER Datasets for Turkic Languages

Several datasets have been developed specifically for Turkic languages, including:

- *Uzbek NER Dataset* with 1,160 sentences annotated for parts of speech and named entities in the Uzbek language (Mengliev et al., 2024);
- *Kazakh NER Dataset* (KazNERD), an open-source dataset with 112,702 sentences and 136,333 annotations across 25 entity classes; it is available on GitHub and offers robust support for NER in Kazakh (Yeshpanov et al., 2022);
- *Turkish Wiki NER Dataset* with 20,000 sentences sampled and re-annotated from the Kuzgunlar NER dataset; this resource focuses on entity types such as *Person*, *Location*, and *Organization*; it is hosted on GitHub and Hugging Face (Altinok, 2023).

The only NER model currently available in the Kyrgyz language is M. Jumashev's model[2] trained on *WikiANN*, which is, according to the author himself, "not usable".

[2]Available at https://huggingface.co/murat/kyrgyz_language_NER

2.4 State of the Art in NER Approaches

NER research continues to evolve rapidly, driven by advancements in deep learning, with larger multilingual corpora and pre-trained models being made available. Transformer-based architectures such as RoBERTa and XLM-RoBERTa provide state-of-the-art performance by capturing semantic nuances and linguistic patterns across different languages. Recent models incorporate more sophisticated attention mechanisms and larger training corpora, leading to state-of-the-art results on various NER benchmarks (Lample et al., 2024).

Recent innovations include integrating external knowledge bases into neural models, enabling better recognition of rare or unseen entities; this hybrid approach has shown promising results in domains such as biomedical NER, where new domain-specific entities frequently appear (Liu et al., 2024). Zero-shot and few-shot learning techniques have also gained popularity, allowing models to generalize to new domains or languages with minimal training data and without extensive retraining at all (Brown et al., 2024).

Cross-lingual and multilingual NER models have demonstrated significant potential in addressing low-resource challenges. NER models are increasingly being adapted to less-resourced languages and domains with specific terminologies such as medical or legal texts. Researchers are increasingly using transfer learning, domain adaptation, and few-shot learning to enable models to generalize better from high-resource to low-resource settings. This has resulted in substantial improvements in the performance of NER systems for languages with limited annotated data, such as Kyrgyz, by leveraging knowledge from more widely spoken languages such as English or Russian (Peters et al., 2024).

Finally, multilingual and cross-lingual NER models are also increasingly able to exploit shared linguistic features across languages to improve recognition accuracy for underrepresented languages such as Kyrgyz. By using shared subword tokenization and multilingual embeddings, these models have improved the accuracy and robustness of NER in languages with limited resources and varied orthographic systems (Conneau and Lample, 2024). Overall, state of the art NER approaches in 2025 focus on advanced neural architectures, integrating external knowledge, and enhancing model adaptability to diverse languages

and domains. This has resulted in more accurate, efficient, and scalable NER systems capable of operating effectively in multilingual and low-resource environments.

2.5 Kyrgyz language processing

Research on the Kyrgyz language has primarily focused on linguistic studies rather than computational approaches. A substantial body of linguistic research has already been devoted to various aspects of the Kyrgyz language; see a recent survey of Kyrgyz NLP research by Alekseev and Turatali (2024). While projects promoting the Kyrgyz language, both commercial (Kan, 2024) and non-commercial (UNESCO-IITE, 2022), have increased recently, there remains a severe lack of annotated datasets for NLP tasks. With this work, we aim to address this gap by providing the first manually annotated dataset for Kyrgyz NER and offering baseline results to support future research.

3 Kyrgyz NER corpus

In this section, we introduce the first manually annotated dataset for named entity recognition (NER) in Kyrgyz. Given the lack of pre-existing high-performance NER models for Kyrgyz, we explored two approaches to address this gap. The first approach involved building an NER model from scratch, while the second focused on adapting existing multilingual models by fine-tuning them on Kyrgyz data.

3.1 Dataset

The dataset consists of 1,499 news articles in Kyrgyz, collected from the 24.KG news portal with permission from the agency's editors for research purposes. These articles, dating from May 2017 to October 2022, were manually annotated to identify named entities using an annotation scheme adapted from the GROBID NER project (GRO, 2008–2023). The dataset contains 10,900 sentences and 39,075 entity mentions across 27 classes, making it the most comprehensive resource for Kyrgyz NER. For annotation, we used the open-source tool *Doccano* (Nakayama et al., 2018), which provided a user-friendly interface for managing the annotation process (see Figure 1).

3.2 Annotation support tool

The annotation process for named entities is labor-intensive and demands both linguistic expertise and consistent attention to detail. Annotation tools aim to improve the efficiency of this process while minimizing human error. After evaluating several options, we selected *Doccano* as the primary tool for annotation due to its flexible interface and support for sequence labeling tasks. The annotation guidelines were adapted from GROBID (GeneRation Of BIbliographic Data) and customized to fit the specific requirements of the Kyrgyz language.

3.3 Annotation Guidelines

We developed detailed annotation guidelines to ensure consistent and accurate labeling of entity mentions in the Kyrgyz NER dataset based on the guidelines from the GROBID project (GRO, 2008–2023). Our tagset covers both broad and specific categories, capturing the diversity of named entities in Kyrgyz texts. Annotators were provided with practical examples and case studies to help them resolve common ambiguities. One of the most challenging cases involves context-dependent named entities. For instance, the word "Президент" (President) can be labeled as either a TITLE or a PERSON, depending on the context:

(1) ⟨Президент Сооронбай Жээнбеков⟩ бүгүн премьер-министр Сапар Исаковду кабыл алды.
⟨*Prezident Sooronbay Jeenbekov*⟩ *bügün premer-ministr Sapar İsakovdu kabıl aldı.*
⟨President Sooronbai Jeenbekov⟩ received Prime Minister Sapar Isakov today.
In this case, the words "Президент Сооронбай Жээнбеков" is a *Person* named entity.

(2) ⟨Президент⟩ бүгүн премьер-министрди кабыл алды.
⟨*Prezident*⟩ *bügün premer-ministrdi kabıl aldı.*
⟨President⟩ received the Prime Minister today.
In this case, "Президент" is a *Title*.

Our guidelines follow the "largest entity mention" principle inherited from the GROBID project: in cases of nested entities, only the encompassing entity is annotated. For example, the token "Кыргыз" (Kyrgyz) can be classified as *National* (when referring to nationality), *Person Type* (when referring to the Kyrgyz people), or *Concept* (when referring to the Kyrgyz language), depending on the context. For more details and examples, we refer to the "largest entity mention" section in the original guidelines[3]. The complete instructions of

[3]As of early 2025, available at https://grobid-ner.readthedocs.io/en/latest/largest-entity-mention/

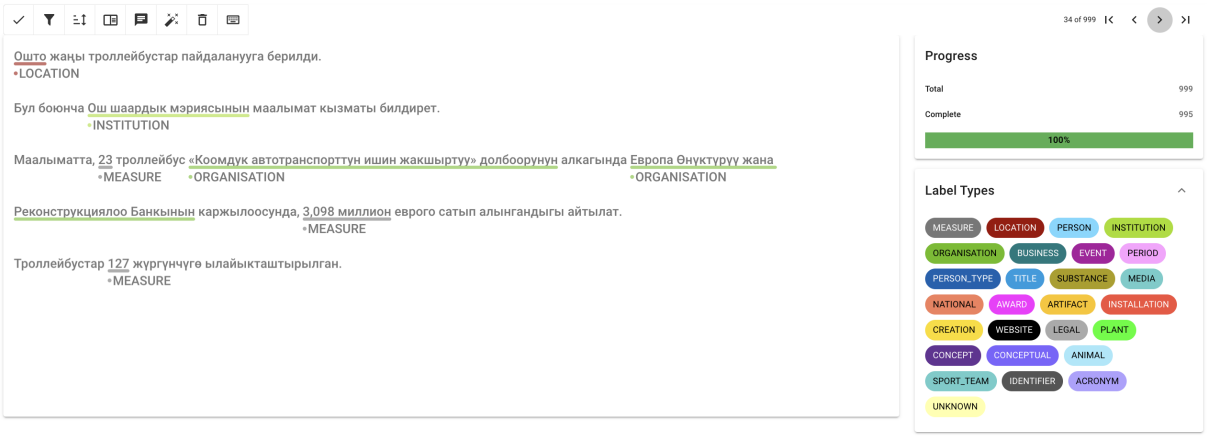

Figure 1: Doccano-based annotation example.

our own annotation scheme are provided in the Appendix.

3.4 Annotators

To ensure the reliability of the dataset, it was essential for annotators to be properly trained. We recruited 59 native Kyrgyz speakers (volunteers and students aged 18-20 who participated in a summer academic practice program) with relevant linguistic expertise and trained them as annotators. Student annotators received academic credit for their participation as part of their curriculum requirements. They followed our annotation guidelines and participated in regular discussions to resolve complex cases. An online communication channel was established to facilitate collaboration between annotators and the annotation process manager (one of the authors). We implemented several quality control measures to maintain the dataset's consistency and accuracy. Each document was annotated by multiple annotators, and a final validation step involved comparing different versions and selecting the most accurate annotations. This process was supervised by domain experts who acted as final approvers.

3.5 Annotation process

The annotation workflow was designed following the MATTER (Model, Annotate, Train, Test, Evaluate, and Revise) schema (Stubbs, 2013) and other related work (Herman Bernardim Andrade et al., 2024; Otto et al., 2023; Naraki et al., 2024). The workflow consisted of five steps, as illustrated in Figure 2:

(1) *data preparation*: the data (news articles) was prepared and uploaded to the annotation system, in our case a *Doccano* instance (Nakayama et al., 2018);

(2) *annotation*: a human annotator can select a document and manually add, remove, or modify each entity based on the instructions from the guidelines; once a document was fully annotated, it was marked as "ready for validation";

(3) *validation/curation*: annotations from different users for a given document are validated and merged into a final annotation; a domain expert ("annotation approver") can compare different annotated versions and select the best combination of annotations or add new ones; this step ensures that the annotations are cross-checked and that the document is validated by domain experts;

(4) *consistency checks and statistical analysis*: this step focuses on identifying obvious errors such as mislabeled data or incorrect linkages; a sequence labeling model is trained and evaluated using 10-fold cross-validation, providing precision, recall, and F-score metrics for each label; on the next iteration, the model is used to automatically generate annotated data, and its predictions serve as a foundation for further refinement and analysis;

(5) *review*: retrospective analysis of the iteration, where unclear cases are discussed and documented in the annotation guidelines.

To inspect and further improve data quality, after the second update we trained a *Deeppavlov NER* model on the annotated data to see whether our data can be used to train the model and identify

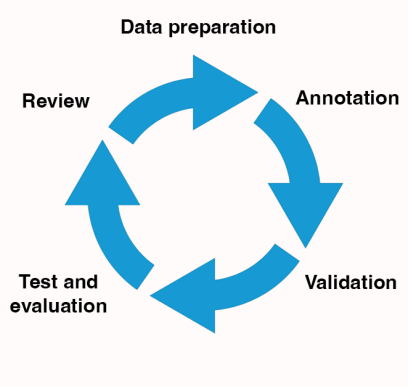

Figure 2: Annotation workflow

Апрель	-	-	B-PERIOD
айынана	-	-	O
баштап	-	-	O
мамлекеттик	-	-	O
жана	-	-	O
муниципалдык	-	-	O
кызматкерлер	-	-	O
кыргыз	-	-	B-CONCEPT
тилинен	-	-	I-CONCEPT
сынак	-	-	O
тапшырат	-	-	O
.	-	-	O

Table 2: The dataset in the CoNLL format.

annotation errors. We split the dataset into 1,000 training texts and 500 test texts, achieving an F1 score of 66.16% on the test set. This feedback loop allowed us to filter out additional errors and improve the dataset through multiple correction sessions.

We use the Cohen's Kappa agreement score to benchmark the reliability of our dataset. We computed the inter-annotator agreement using 30 sampled texts composed of 2,773 tokens. The agreement score is $\kappa = 0.89$. This rather high agreement score serves as evidence that we have obtained a high-quality dataset.

3.6 Data Format

Our dataset is presented in the CoNLL-2003 format, as shown in Table 2. Word boundaries were established using a tokenizer from the *Apertium-Kir* tool. This format ensures compatibility with standard NER evaluation frameworks and facilitates the adoption of our dataset for future research.

4 Data Statistics

In this section, we present an overview of the statistics and distribution of classes in the Kyrgyz NER dataset. The dataset comprises 1,499 documents, totaling 140,366 tokens and 10,900 sentences. The annotated dataset includes 39,075 named entity mentions distributed across 27 entity types. Table 3 provides dataset statistics and details its breakdown into training and test sets.

Items	Train	Test	Total
Documents	999	500	1,499
Sentences	7,033	3,867	10,900
Tokens	89,248	51,118	140,366
Mentions	24,949	14,126	39,075

Table 3: Data statistics and train/test split.

One of the key challenges for this dataset is the uneven distribution of entity classes. As shown in Figure 3, the top four most frequent classes account for approximately 70% of all mentions, while many other classes have only a limited number of examples. This class imbalance poses a significant challenge for training models, particularly for underrepresented classes. The histogram in Figure 3 shows the frequency distribution of entity mentions for each class. Classes such as *Person*, *Location*, *Measure*, and *Institution* are among the most common, while others like *Animal*, *Award*, or *Substance* are much rarer. The scarcity of examples for these underrepresented classes affects model performance and increases the likelihood of false negatives, and addressing class imbalance may be an important direction for future work.

5 Experimental Setup and Results

In this section, we describe the experimental setup and results of our baseline models on the Kyrgyz NER dataset. The primary goals of these experiments are: (1) to assess the performance of well-established NER techniques on the Kyrgyz dataset and identify the most suitable baseline for future research, (2) to analyze how class imbalance affects model performance, given that the top four most frequent classes account for 70% of mentions (see Fig. 3), and (3) to compare the performance of existing models on our Kyrgyz dataset with results from similar experiments on high-resource languages such as English.

5.1 Baselines and Their Hyperparameters

We split the dataset into training and validation subsets, using 20% of the training set as a validation set. We have trained a wide variety of models on

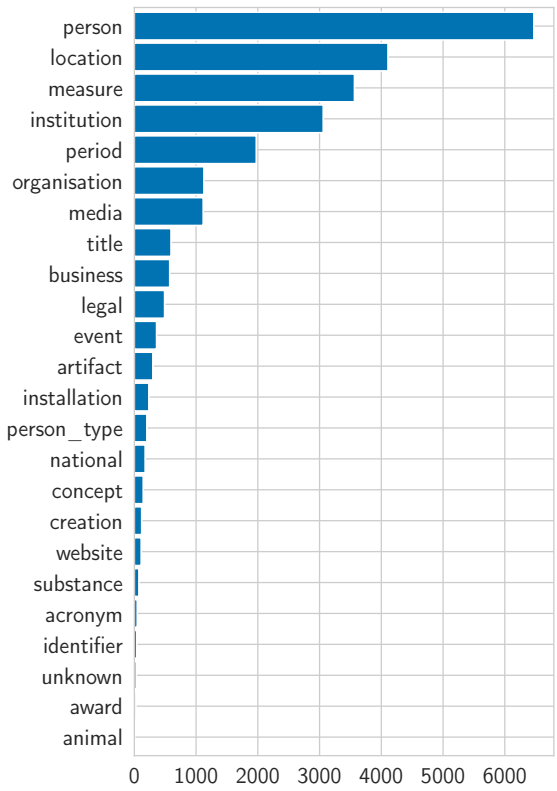

Figure 3: Distribution of label categories.

Model	Prec	Rec	F1
Bert+CRF	0.67	0.63	0.65
CRF	0.70	0.55	0.62
Pretrained mT5	0.70	0.68	0.69
Pretrained Bert	0.68	0.68	0.68
Pretrained XLMR	0.74	0.71	**0.73**

Table 4: Experimental results of NER models.

Label	BERT +CRF	CRF	mT5	BERT	XLMR	Support
Measure	0.82	0.76	0.85	0.82	0.86	1046
Person	0.80	0.66	0.82	0.77	0.82	989
Location	0.74	0.66	0.76	0.73	0.76	900
Institution	0.59	0.54	0.63	0.60	0.64	660
Period	0.54	0.64	0.68	0.69	0.69	545
Plant	0.00	0.00	0.00	0.00	0.00	10
Award	0.00	0.00	0.00	0.18	0.00	7
Conceptual	0.00	0.00	0.00	0.00	0.00	5
Identifier	0.00	0.00	0.00	0.42	0.67	4
Animal	0.00	0.00	0.00	0.00	0.00	3

Table 5: Per-class F1 score.

the dataset, ranging from classical CRF-based models to state-of-the-art transformer-based models.

Classical baseline: CRF. We utilized a Conditional Random Fields (CRF) model as the classical baseline for our experiments, training the model from scratch with the *sklearn-crfsuite* library (Okazaki, 2007) library. We used the L-BFGS optimization algorithm (*algorithm='lbfgs'*) with a maximum of 5000 iterations (*max_iterations=5000*) to ensure convergence. To balance the trade-off between precision and generalization, we set the L1 regularization coefficient to 0.3 (*c1=0.3*) and the L2 regularization coefficient to 0.6 (*c2=0.6*). Additionally, we enabled all possible state transitions (*all_possible_transitions=True*) to capture complex dependencies between labels.

BERT+CRF. We used the `DeepPavlov` library[4] to train a model from scratch with the ner_ontonotes_bert_mult[5] configuration.

BERT. We used pretrained the BERT multilingual base (cased) version (Devlin et al., 2019b) to fine-tune the NER model for Kyrgyz language. Fine-tuning was performed with the `bert-base-multilingual-cased` checkpoint, with a batch size of 128 and a learning rate of $5 \cdot 10^{-5}$. To prevent overfitting, we applied a weight decay of 0.0001. The model was trained for 8 epochs, with 3000 warmup steps to stabilize the learning process.

XLM-RoBERTa (XLMR). We used the XMLR (Conneau et al., 2020) base model to fine-tune the NER model for the Kyrgyz language. The fine-tuning process employed the xlm-roberta-base[6] checkpoint, with a batch size of 8 and a learning rate of 10^{-5}. To mitigate overfitting, we used a weight decay of 0.01, training the model over 10 epochs with 8000 warmup steps.

mT5. We fine-tuned the mT5-small model (Xue et al., 2021b) to obtain a NER model for the Kyrgyz language. Fine-tuning used the google/mt5-small[7] checkpoint, with a batch size of 16 and a learning rate of 10^{-5}. To prevent overfitting, we applied a weight decay of 0.001. Training was conducted over 10 epochs, with 800 warmup steps to stabilize the learning process. The maximum token length was set to 64.

[4] https://docs.deeppavlov.ai/en/master/features/models/NER.html

[5] https://docs.deeppavlov.ai/en/master/features/models/NER.html#3.-Models-list

[6] https://huggingface.co/FacebookAI/xlm-roberta-base

[7] https://huggingface.co/google/mt5-small

5.2 Experimental Results

The results of our experiments are summarized in Table 4. As expected, the CRF model achieved high precision but struggled with recall, leading to a relatively low F1 score even after extensive hyperparameter tuning. Transformer-based models consistently outperformed the CRF baseline, with XLM-RoBERTa delivering the best overall performance (F1 score of 0.70). The mT5 model also performed well, indicating the potential of text-to-text Transformer-based models for low-resource NER tasks. We note that in a recent study on Kyrgyz texts classification (Alekseev et al., 2023), the "Large" modification of XLM-RoBERTa also yielded the best results, while multilingual BERT was far behind. In this case, the performance gap was relatively small, so all Transformer-based models represent strong baselines for future studies on Kyrgyz tagging tasks.

Table 5 shows a breakdown of the F1 score for several popular and rare entity classes; as expected, classes with very low support are virtually unrecognizable for the models, while larger amount of training data leads to significantly improved results for all models. Note the difference in results between *Measure*/*Person* and *Institution*/*Period* classes: as support drops from 900-1000 examples to 550-650, the F1 scores go down from 0.8-0.85 to 0.6-0.65.

6 Detailed Error Analysis

There were two stages of error analysis in our work. First, we conducted a detailed error analysis on the predictions made by the BERT+CRF model to identify recurring mistake patterns and understand the challenges in the dataset. The findings highlight several critical issues, including ambiguous entity mentions and the scarcity of training examples for certain classes. This analysis helped refine our annotation guidelines, and it was part of the annotation process (see Section 3). As for the baseline models trained on the resulting KyrgyzNER dataset, we identify two major sources of errors.

Ambiguous entity mentions: one of the main sources of errors was context-dependent entity mentions that could belong to multiple classes. For example, the word "British" can be labeled as *National* ("a British newspaper reported"), *Person Type* ("a British journalist reported"), or *Concept* ("a journalist reported in British English"), depending on the context. The model struggled to disambiguate these cases, frequently misclassifying them or producing false negatives.

Scarcity of training samples: another major challenge was the lack of sufficient training samples for certain classes, including *Acronym*, *Animal*, *Artifact*, *Award*, *Concept*, *Event*, *Identifier*, *Installation*, *Legal*, *Plant*, and *Substance*. As a result, models often failed to predict any labels for these classes, leading to a high rate of false negatives. To address these challenges, we propose to either revise and filter the set of classes or extend the training data with either synthetic sentences or upsampling.

7 Conclusion

In this work, we present KyrgyzNER, the first manually annotated named entity recognition (NER) dataset for the Kyrgyz language. The dataset comprises 1,499 documents with 10,900 sentences and 39,075 entity mentions across 27 categories. This resource addresses the lack of available language datasets for Kyrgyz and aims to provide a foundation for further research in low-resource language processing and exploring the effects of an unbalanced label distribution.

We conducted baseline experiments with NER models ranging from classical CRF to modern Transformer-based approaches such as multilingual BERT, XLM-RoBERTa, and mT5. Our results showed that Transformer-based models significantly outperform classical methods, with XLM-RoBERTa achieving the highest F1 score. However, class imbalance remains a major challenge, especially for rare entity types. Interestingly, the scores achieved with mBERT (base), XLM-RoBERTa (base), and mT5 are very close to each other (F1 score about 70%), so we suggest that all of them should be used as baselines for future research.

Our error analysis revealed two key issues: ambiguity in entity mentions and the scarcity of training samples for certain classes. To mitigate these challenges, we suggest refining annotation guidelines, generating synthetic data, and applying upsampling techniques in future work. These steps could improve model performance and provide a more robust benchmark for Kyrgyz NER.

We hope that the publication of this dataset will encourage the NLP community to include Kyrgyz in multilingual benchmarks and promote further research in low-resource languages. By expanding the scope of language resources, we move closer to reducing the disparity between high-resource and under-resourced languages in NLP.

Limitations

Despite its practical value, the KyrgyzNER dataset has several limitations that should be addressed in future work. First, the dataset is derived entirely from the 24.KG news portal, limiting its domain coverage. This may affect the model's ability to generalize to other genres or registers of Kyrgyz, such as social media, conversational text, or legal documents. Second, class imbalance is a major challenge. While some entity types such as PERSON, LOCATION, and MEASURE are well-represented, many others—such as ACRONYM, ANIMAL, AWARD, or LEGAL—have relatively few examples. This makes it difficult for models to learn robust distinctions for these classes and increases the likelihood of false negatives. Expanding the dataset with more diverse sources could help alleviate this issue. Third, our annotation scheme focuses on "flat" named entities and does not address nested entities, so complex linguistic constructions involving overlapping entities remain unannotated, which could limit the dataset's utility for certain applications. Finally, while the dataset underwent rigorous quality control and achieved a high inter-annotator agreement score (Cohen's kappa of 0.89), there is always a risk of missed or inconsistent labels due to the complexity and nuance of the Kyrgyz language. Further refinement of the annotation guidelines and additional rounds of validation could help improve consistency and reduce residual errors.

References

2008–2023. Grobid. https://github.com/kermitt2/grobid.

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

A. Alekseev, Sergey I. Nikolenko, and Gulnara Kabaeva. 2023. Benchmarking multilabel topic classification in the kyrgyz language. *Preprint*, arXiv:2308.15952.

Anton Alekseev and Timur Turatali. 2024. KyrgyzNLP: Challenges, Progress, and Future. *Preprint*, arXiv:xxxx.xxxx.

Duygu Altinok. 2023. A diverse set of freely available linguistic resources for Turkish. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13739–13750, Toronto, Canada. Association for Computational Linguistics.

Douglas E. Appelt, Jerry R. Hobbs, John Bear, David Israel, Megumi Kameyama, David Martin, Karen Myers, and Mabry Tyson. 1995. SRI International FASTUS system: MUC-6 test results and analysis. In *Proceedings of the 6th Conference on Message Understanding*, MUC6 '95, USA. Association for Computational Linguistics.

William J Black, Fabio Rinaldi, and David Mowatt. 1998. FACILE: Description of the NE system used for MUC-7. In *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2024. Language models are few-shot learners. In *Proceedings of the 2024 Annual Conference on Neural Information Processing Systems (NeurIPS)*. Curran Associates, Inc.

Christopher Cieri, Mike Maxwell, Stephanie Strassel, and Jennifer Tracey. 2016. Selection criteria for low resource language programs. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4543–4549.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. *Preprint*, arXiv:1911.02116.

Alexis Conneau and Guillaume Lample. 2024. Cross-lingual language model pretraining. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019a. BERT: Pre-training of

deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019b. Bert: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arXiv:1810.04805.

Sean R Eddy. 1996. Hidden markov models. *Current opinion in structural biology*, 6(3):361–365.

Jenny Rose Finkel and Christopher D Manning. 2009. Nested named entity recognition. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 141–150.

Daniel Hanisch, Katrin Fundel, Heinz-Theodor Mevissen, Ralf Zimmer, and Juliane Fluck. 2005. Prominer: rule-based protein and gene entity recognition. *BMC bioinformatics*, 6:1–9.

Gabriel Herman Bernardim Andrade, Shuntaro Yada, and Eiji Aramaki. 2024. Is boundary annotation necessary? evaluating boundary-free approaches to improve clinical named entity annotation efficiency: Case study. *JMIR Med Inform*, 12:e59680.

S. Hochreiter and J. Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780. Based on TR FKI-207-95, TUM (1995).

Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

K. Humphreys, R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, H. Cunningham, and Y. Wilks. 1998. University of Sheffield: Description of the LaSIE-II system as used for MUC-7. In *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998*.

Daniel Jurafsky and James H Martin. 2008. Speech and language processing. prentice hall. *Upper Saddle River, NJ*.

Anastasia Kan. 2024. Akylai smart speaker: Artificial intelligence speaking kyrgyz language (june 18th, 2024). https://web.archive.org/web/20240619010036/https://24.kg/english/296874_AkylAI_smart_speaker_Artificial_intelligence_speaking_Kyrgyz_language/. Accessed: 2024-09-14.

J.N. Kapur. 1989. *Maximum-entropy Models in Science and Engineering*. Wiley.

George R. Krupka and Kevin Hausman. 1998. IsoQuest inc.: Description of the NetOwl^{TM} extractor system as used for MUC-7. In *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998*.

John Lafferty, Andrew Mccallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2024. Unsupervised machine translation using monolingual corpora only. In *Proceedings of the 2024 Annual Conference on Neural Information Processing Systems (NeurIPS)*. Curran Associates, Inc.

Pan Liu, Yanming Guo, Fenglei Wang, and Guohui Li. 2022. Chinese named entity recognition: The state of the art. *Neurocomputing*, 473:37–53.

Tianyu Liu, Jingfei Du, and John S. Platt. 2024. Towards improving neural named entity recognition with gazetteers. *Journal of Machine Learning Research*, 25:1–29.

Natalia Loukachevitch, Suresh Manandhar, Elina Baral, Igor Rozhkov, Pavel Braslavski, Vladimir Ivanov, Tatiana Batura, and Elena Tutubalina. 2023. Nerel-bio: a dataset of biomedical abstracts annotated with nested named entities. *Bioinformatics*, 39(4):btad161.

Xuezhe Ma and Eduard Hovy. 2016. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1:1064–1074.

Andrew McCallum and Wei Li. 2003. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. pages 188–191.

Paul McNamee and James Mayfield. 2002. Entity extraction without language-specific resources. In *Conference on Computational Natural Language Learning*.

Davlatyor Mengliev, Vladimir Barakhnin, Nilufar Abdurakhmonova, and Mukhriddin Eshkulov. 2024. Developing named entity recognition algorithms for uzbek: Dataset insights and implementation. *Data Brief*, 54(110413):110413.

Zulfat Miftahutdinov, Ilseyar Alimova, and Elena Tutubalina. 2020. On biomedical named entity recognition: experiments in interlingual transfer for clinical and social media texts. In *European Conference on Information Retrieval*, pages 281–288. Springer.

Jamshidbek Mirzakhalov, Anoop Babu, Aigiz Kunafin, Ahsan Wahab, Bekhzodbek Moydinboyev, Sardana Ivanova, Mokhiyakhon Uzokova, Shaxnoza Pulatova, Duygu Ataman, Julia Kreutzer, et al. 2021. Evaluating multiway multilingual nmt in the turkic languages. In *Proceedings of the Sixth Conference on Machine Translation*, pages 518–530.

Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human. Software available from https://github.com/doccano/doccano.

Yuji Naraki, Ryosuke Yamaki, Yoshikazu Ikeda, Takafumi Horie, Kotaro Yoshida, Ryotaro Shimizu, and Hiroki Naganuma. 2024. Augmenting ner datasets with llms: Towards automated and refined annotation. *Preprint*, arXiv:2404.01334.

Naoaki Okazaki. 2007. Crfsuite: a fast implementation of conditional random fields (crfs).

Wolfgang Otto, Matthäus Zloch, Lu Gan, Saurav Karmakar, and Stefan Dietze. 2023. Gsap-ner: A novel task, corpus, and baseline for scholarly entity extraction focused on machine learning models and datasets. *Preprint*, arXiv:2311.09860.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2024. Semi-supervised sequence tagging with bidirectional language models. *Transactions of the Association for Computational Linguistics*, 12:193–208.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. 2013. Towards robust linguistic analysis using OntoNotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152, Sofia, Bulgaria. Association for Computational Linguistics.

Peng Qi, Timothy Dozat, Yuhao Zhang, and Christopher D. Manning. 2018. Universal Dependency parsing from scratch. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 160–170, Brussels, Belgium. Association for Computational Linguistics.

Alexandra Pomares Quimbaya, Alejandro Sierra Múnera, Rafael Andrés González Rivera, Julián Camilo Daza Rodríguez, Oscar Mauricio Muñoz Velandia, Angel Alberto Garcia Peña, and Cyril Labbé. 2016. Named entity recognition over electronic health records through a combined dictionary-based approach. *Procedia Computer Science*, 100:55–61. International Conference on ENTERprise Information Systems/International Conference on Project MANagement/International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN / HCist 2016.

Afshin Rahimi, Yuan Li, and Trevor Cohn. 2019. Massively multilingual transfer for ner. *Preprint*, arXiv:1902.00193.

Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe, and Wei Xu. 2016. Results of the WNUT16 named entity recognition shared task. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, pages 138–144, Osaka, Japan. The COLING 2016 Organizing Committee.

Amber C. Stubbs. 2013. *A methodology for using professional knowledge in corpus annotation*. Ph.D. thesis, USA. AAI3558548.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.

UNESCO-IITE. 2022. A chatbot for teenagers about puberty, relationships, and health launched in kyrgyzstan (may 24th, 2022). `http://web.archive.org/web/20240525072322/https://iite.unesco.org/highlights/oilo-chatbot-sex-ed-kyrgyzstan-en/`. Accessed: 2024-09-14.

Sowmya Vajjala and Ramya Balasubramaniam. 2022. What do we really know about state of the art ner? *Preprint*, arXiv:2205.00034.

A. et al. Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Jue Wang, Lidan Shou, Ke Chen, and Gang Chen. 2020. Pyramid: A layered model for nested named entity recognition. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 5918–5928.

Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, and Guoyin Wang. 2023. Gpt-ner: Named entity recognition via large language models. *Preprint*, arXiv:2304.10428.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021a. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021b. mt5: A massively multilingual pre-trained text-to-text transformer. *Preprint*, arXiv:2010.11934.

Rustem Yeshpanov, Yerbolat Khassanov, and Huseyin Atakan Varol. 2022. KazNERD: Kazakh named entity recognition dataset. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 417–426, Marseille, France. European Language Resources Association.

L. Zhang and H. Wu. 2021. Medical text entity recognition based on deep learning. *Journal of Physics: Conference Series*, 1744(4):042209.

A Appendix: Annotation Instructions

The original instructions were prepared in Russian. Below is the translated version.

A.1 Key Principles

The ultimate goal of such methods is to extract some useful information from the text (not necessarily commercial, for example, for distant reading of the national epic by philologists); data should be annotated with this in mind.

Example: "The policeman saved the boy." — there's nothing to extract here, as we can't match any specific person to either the policeman or the boy in this text. However, if the unique title such as the "Mayor of Bishkek" is mentioned, we can do this.

The same goes for time intervals. "Yesterday" tells us nothing, but "in January" already allows us to make some intelligent guesses about the mentioned time span.

A.1.1 The Principle of the Largest Entity

The main rule: if there are nested entities, one should always take the largest one covering (including) everything (largest entity mention).

Consider the word "British." Depending on the context, "British" may correspond to classes (labels):

- NATIONAL (when referring to something related to the UK),
- PERSON_TYPE (when referring to the British people),
- CONCEPT (when it means British English).

However, "British Brexit referendum" should be marked as an EVENT because "British" is a part of a larger entity.

B TITLE and PERSON

If a title (position) is followed by a name, mark it as a single PERSON tag.

Other examples:

- While attending the May 2012 NATO summit meeting (EVENT),
- Obama (PERSON) was the US President (TITLE),
- She is the CEO of IBM (TITLE),
- The President of Argentina (TITLE) said no,
- German South-West Africa (LOCATION),
- American Jewish Holocaust survivors (PERSON_TYPE),
- chairman of the Central Committee of the World Sephardi Federation and a member of the Knesset (TITLE).

IMPORTANT: Entities should NOT be marked by clicking, to avoid space characters being included. The word should be selected FULLY; PARTIAL word selection is NOT allowed.

B.1 Labels, Their Meaning, Examples

B.2 Frequently Asked Questions

- Do quotes around an organization or sports team name get included? Yes, the quotes should be marked along with the entity.
- Can an entity be split into parts? No, it cannot.
- Is a road, an interchange, or a ring road considered INSTALLATION or LOCATION? We'll consider it as LOCATION.
- What label should be used for a specific hospital, prison, school, theater, or border crossing? These should be marked as INSTITUTIONs.
- What about a state-owned enterprise, like a specific factory? This should be marked as BUSINESS.
- If I encounter a difficult case and I am unsure what to do? Gather a small batch of questions and post them in the speacial annotation-related chat [link].

B.3 Clarifications

Specific Cases

- "President Almazbek Atambayev" is a single PERSON segment. If there is no name after "President", it should be labeled as a TITLE.

Label Name	Description	Examples EN	Examples KG
PERSON	Names, surnames, nicknames, and callsigns of real and fictional PEOPLE	John Smith	Исхак Раззаков, Чыңгыз Айтматов
PERSON_TYPE	Type of person, societal role, often based on group membership	African-American, Asian, Conservative, Liberal, Jews	кыргыз, татар, түрк
ANIMAL	Animal names	Hachikō, Jappeloup	Хатико
TITLE	Title, professional address, or position	Mr., Dr., General, President	Мырза, Президент
ORGANISATION	Organized group of people, with some form of legal entity and concrete membership	Alcoholics Anonymous, Jewish resistance, Polish underground	Кыргыз кино
INSTITUTION	Organization of people and a location or structure that shares the same name	Yale University, European Patent Office, the British government	КГТУ, Политех, Кыргызпатент
BUSINESS	A company or commercial organization	Air Canada, Microsoft	Шоро
SPORT_TEAM	Organization or group associated with sports	The Yankees, Leicestershire	ФК Дордой-динамо
MEDIA	Media, publishing organization, or name of the publication itself	Le Monde, The New York Times	Кактус медиа, Клоп, 24.KG
WEBSITE	Website name or link. For news companies, mark as MEDIA.	Wikipedia, http://www.inria.fr	Instagram, Facebook
IDENTIFIER	Identifier like phone number, email, ISBN	2081396505, weirdturtle@gmail.com	+996312000001
LOCATION	A specific place; includes planets and galaxies	Los Angeles, Earth	Кыргызстан, Бишкек, Чолпон
NATIONAL	Pertaining to location or nationality	a British newspaper, a British historian	кыргызстандык, орусиялык
INSTALLATION	A structure built by people	Strasbourg Cathedral, Auschwitz camp	Бурана
ARTIFACT	A man-made object, including software products	FIAT 634, Microsoft Word	«Гиннес китеби»
CREATION	A work of art or entertainment, such as a song, movie, or book	Mona Lisa, Kitchen Nightmares	Дочь Советской Киргизии
EVENT	A specific event	World War 2, Brexit referendum	Экинчи дуйнолук согуш
AWARD	Award for achievements in various fields	Ballon d'Or, Nobel Prize	Нобель сыйлыгы
PERIOD	Date or historical period, time intervals	January, 1985-1989	15 март, дүйшөмбү
LEGAL	Legal references like laws, conventions, court cases	European Patent Convention, Roe v. Wade	
MEASURE	Numerical or ordinal value	1,500, 72%	
PLANT	Name of a plant	Ficus religiosa	
SUBSTANCE	Substance	HCN, gold	алтын, күмүш
CONCEPT	Abstract concept, not included in any other class	Communism, FTSE 100	Коммунизм, кыргыз тили
CONCEPTUAL	Entity associated with a concept	Greek myths, eurosceptic doctrine	
ACRONYM	Acronyms and abbreviations	DIY, BYOD, IMHO	
UNKNOWN	Entity that does not fit into any listed classes	Plan Marshall, Horizon 2020	Маршалл планы

Table 6: Description of labels.

- MEASURE
  "9 миллион 500 миң" (nine million five hundred thousand) should be marked as a single MEASURE segment (since it's a single number).
  - Examples like "10 сом" or "5 apples" — only the number should be marked.

B.4 Iteration #1 (June 5-8, 2023)

GENERAL RULE: If an entity becomes too long, please check whether it has been formed

correctly; ensure that no sequences of entities are included into the spans, and no spaces or verbs are included into them.

If the same entity appears multiple times in the text, it should be labeled each time. Words should not be split into parts by selected spans: “70тен”

B.5 Iteration #2 (Questions and Answers After the First Month of Annotation Process)

- Muftiate — INSTITUTION.
- Kyrgyz language — should mark only “‘Kyrgyz”’ as a CONCEPT.
- Central Mosque in Bishkek — INSTALLATION.
- Religious Affairs Directorate — INSTITUTION.
- Religious Affairs Information and Counseling Center — INSTITUTION.
- Markets (e.g., Osh Bazaar, Dordoi Bazaar) — BUSINESS.
- Media abbreviation (ЖКМК) — ACRONYM.
- Uzbekistan Gymnastics Federation — ORGANISATION.
- Rogun HPP (Hydroelectric Power Plant) — INSTALLATION.
- Makarov-type pistol — ARTIFACT.
- World Bank — INSTITUTION.
- “SDPK Party” — ORGANISATION.
- “Kyrgyzstan” political party — ORGANISATION.
- Central Mosque in Bishkek — INSTALLATION.
- Manas (airport) — INSTALLATION.
- Tokmok city — only mark “TOKMOK” part as a LOCATION.