Large Language Models for Multilingual Slavic Named Entity Linking

Rinalds Vīksna[1,2] and **Inguna Skadiņa**[1,2] and **Daiga Deksne**[1,2] and **Roberts Rozis**[1]
[1] Tilde, Vienības gatve 75a, Riga, Latvia
[2] Faculty of Computing, University of Latvia, Raiņa bulv. 29, Riga, Latvia
{Firstname.Lastname}@tilde.lv

Abstract

This paper describes our submission for the 4th Shared Task on SlavNER on three Slavic languages - Czech, Polish, and Russian. We use pre-trained multilingual XLM-R Language Model and fine-tune it for three Slavic languages using datasets provided by organizers. Our multilingual NER model achieves a 0.896 F-score on all corpora, with the best result for Czech (0.914) and the worst for Russian (0.880). Our cross-language entity linking module achieves an F-score of 0.669 in the official SlavNER 2023 evaluation.

1 Introduction

The 4th edition of Shared Task address three Slavic languages: Czech, Polish, and Russian, and five types of named entities (persons, locations, organizations, events, and products). All languages are highly inflective and have a rather free word order. Thus named entity normalization task faces an additional challenge in the case of the normalization of multi-word expressions (MWE).

In our submission, we continue experiments with XLM-R Language Model (Conneau et al., 2020) which has demonstrated the best result in previous shared task (Ferreira et al., 2021). We also elaborate on the normalization step for MWEs by applying syntax-based noun phrase normalization tool to reach higher accuracy in named entity (NE) normalization and linking tasks. Finally, we also improve entity linking by better algorithms for linking entity variants on a document level using string similarity, proximity, and type attributes.

The paper is organized as follows. We start with an overview of the data preparation step (Section 3) and the overall architecture of the system (Section 4). Then, we present each step in our workflow - mention detection, entity normalization, and entity linking. We conclude the paper with a subset of results and a discussion (Section 8).

2 Related Work

The shared task on Slavic multilingual named entity recognition, normalization, and linking (SlavNER) has been organized since 2017 (Piskorski et al., 2017). Only two systems were submitted for the First SlavNER. The best result for NER was achieved for Polish (F-score of 66.6), while for cross-lingual entity matching only 9 F1 points were reached (Mayfield et al., 2017). Authors of this system annotated parallel English-target language datasets using an English NER and projected annotations to the target language. A target language tagger was then trained using inferred datasets.

Seven teams submitted systems to the 2nd SlavNER (Piskorski et al., 2019). The three best systems (RIS (Arkhipov et al., 2019), CogComp (Tsygankova et al., 2019) and IIUWR.PL (Piskorski et al., 2019)) used BERT for the NER task. The best model, CogComp, yields an F-measure of 91% according to the shared task organizers. The cross-lingual entity linking results also have improved significantly: the best-performing model, IIUWR.PL yields the F-measure of 45%.

Six teams submitted their systems to the 3rd SlavNER (Piskorski et al., 2021). Overall NER task results were lower when compared to the 2nd SlavNER. The best system, Priberam (Ferreira et al., 2021), achieved F-measure of 85.7% for the relaxed partial evaluation. Priberam used XLM-R Large model, a character-level embedding model, and a biaffine classifier for NER task. For cross-lingual entity linking, the best-performing model, TLD (Vīksna and Skadina, 2021), achieved an F-measure of 50.4% using LaBSE (Feng et al., 2022) embeddings to align entities according to pre-defined thresholds.

3 Data Preparation

The data provided by the SlavNER task organizers contains annotations for five classes of entities:

Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023), pages 172–178
May 6, 2023 ©2023 Association for Computational Linguistics

event (EVT), location (LOC), person (PER), organization (ORG), and product (PRO). For NER system training we convert data into a conll2003-like format. We do not use the data from the BSNLP2017 shared task (Piskorski et al., 2017), as it has 4 named entity classes which are inconsistent with the rest of SlavNER data (Prelevikj and Zitnik, 2021), and has shown to hurt the performance of NER models for this task (Ferreira et al., 2021). In addition to the dataset provided by the SlavNER task organizers(Piskorski et al., 2019, 2021), we use the following datasets in our experiments:

KPWr (Oleksy et al., 2019) contains Polish texts labeled using 82 classes of entities, which we map to the 5 classes used in the SlavNER task.

NKJP (Przepiórkowski, 2012) is National Corpus of Polish, tagged with fine-grained NEs. We use entity types PER ('forename', 'surname'), LOC (placeName, geogName), and ORG (orgName).

poleval2018 (Ogrodniczuk and Łukasz Kobyliński, 2018) is POLEVAL 2018 NER task gold dataset, labeled using the same guidelines as NKJP.

FiNER (Ruokolainen et al., 2019) is a Finnish dataset that contains the same NE types as SlavNER, thus useful to train NER for EVT and PRO classes.

CNEC (Ševčíková et al., 2014) Czech Named Entity Corpus 2.0 is labeled according to a two-level hierarchy of 46 named entities. It was mapped to the corresponding 4 classes of the SlavNER task: ORG, PER, LOC, and PRO.

FactRU (Starostin et al., 2016) is a Russian dataset, labeled with 4 classes of entities (Org, LocOrg, Location, and Person), which can be mapped to 3 classes of the SlavNER task: ORG, LOC, PER.

conll2002 (Tjong Kim Sang, 2002) is a Spanish NER dataset labeled with PER, LOC, ORG and MISC classes.

conll2003 (Tjong Kim Sang and De Meulder, 2003) is an English NER dataset labeled using PER, LOC, ORG and MISC classes.

4 Architecture and systems

The architecture of our solution is modular: the modules roughly correspond to the data processing steps necessary to reach the objectives of different SlavNER Shared Tasks: mention detection, lemmatization, and linking (Figure 1).

We submitted five systems to the SlavNER task. Table 1 provides an overview of our systems. In the following sections, we provide more details of our solutions.

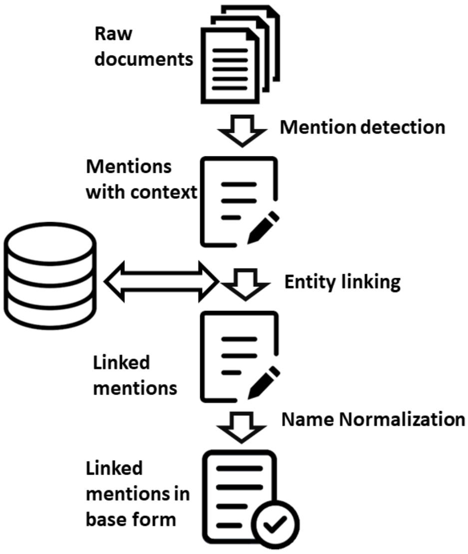

Figure 1: Overall System Architecture

	NER	Linker
1	XLM-R Base Ensemble	C
2	XLM-R Large	C
3	XLM-R Large	D and C
4	XLM-R Large plus KPWr data	D and C
5	XLM-R Base additionaly pre-trained plus KPWr data	D and C

Table 1: System overview (C-corpus level, D- document level)

5 Two Approaches to Mention Detection: traditional and ensemble

We consider the Named Entity Mention Detection and Classification task as the NER task. We use the Flair library (Schweter and Akbik, 2020) to perform NER. Flair library allows fine-tuning a Transformer (Vaswani et al., 2017) model with custom data. Multilingual XLM-R has demonstrated the best result in previous shared task (Ferreira et al., 2021) and is used as a basis for our NER models. The XLM-R is available in XLM-R Base (L= 12, H = 768, A = 12, 270M params) and XLM-R Large(L = 24, H = 1024, A = 16, 550M params) variants.

We use XLM-R Large model fine-tuned on the dataset provided by the Shared Task organizers as a NER model for our System-2 and System-3.

For System-4 we fine-tune a XLM-R Large model on the dataset given by the Shared Task organizers combined together with KPWR-NER dataset (Marcińczuk, 2020).

Although multiple NER datasets for Czech, Polish, and Russian are available, most of them could not be directly used due to differences in tagsets. However, even if the set of labeled classes is incompatible with the SlavNER labeling schema, it is still possible to use this data for training a NER system to recognize a single class that has compatible labeling. This is done by keeping only a single label in a dataset and deleting all other labels.

Using single-label datasets, we train a NER system by combing SlavNER dataset with this dataset and evaluate against the SlavNER test split. If a system achieves better results than the baseline system trained on SlavNER data, we consider this dataset as compatible with SlavNER and select to train the final NER model for a given label. Datasets used to fine-tune each single-label NER model are summarised in Table 2.

Model	Datasets used for training
EVT	SlavNER, KPWr
LOC	SlavNER, CNEC, KPWr, conll2002, conll2003, FactRu, finer, NKJP
ORG	SlavNER, CNEC, KPWr, FactRu, NKJP, Poleval
PER	SlavNER, Poleval,
PRO	SlavNER, CNEC, KPWr, finer

Table 2: Datasets used to train single-label models

Due to performance and time restrictions, the XLM-R Base model is used to fine-tune ensemble models. During the evaluation, all five NER models are run sequentially. The overlapping labels are resolved, first by selecting the longest labeled entity and then, if there is an exact overlap, by selecting the highest score returned by NER. This ensemble approach is used by our System-1.

Since the XLM-R models were created more than two years ago, and thus outdated with respect to current events, we crawled 2.6 GB of the latest Czech, Polish, and Russian news articles[1] to perform additional pretraining of XLM-R base model. Due to the time restrictions, additional pretraining was done using huggingface/transformers example script[2] with batch size 512, for 7000 steps. This additionally pre-trained XLM-R model was fine-tuned using the SlavNER dataset and the KPWr dataset for NER of System-5.

6 Entity Normalization

We use several strategies for entity normalization. In case of the Czech language we apply a simple word-level lemmatization strategy. We use Stanza (Qi et al., 2020) Czech language lemmatizer for this task.

For entity normalization in Polish and Russian, we use a language-specific noun phrase generator. It allows us to transform the noun phrase into the corresponding base form taking into account the grammar rules of the specific language.

The normalization workflow includes several steps: tokenization, morphological analysis, syntactic parsing, morphological transfer, and morphological synthesis of the base form. Morphological analysis and synthesis are performed with help of a language-specific finite state transducer (FST). This FST solution was initially developed for the Latvian language (Deksne, 2013) and recently extended to many other European languages - Lithuanian, Polish, Finnish, Swedish, Spanish, French, German, and English. For the syntactic parsing Cocke-Younger-Kasami (CYK) algorithm (Younger, 1967) is employed by adapting the corresponding Latvian tool (Deksne et al., 2014).

When analysing output of the normalisation tool, we identified several reasons for errors:

- A word in a phrase is unrecognized acronym.
- In the case of homographs, if a word has some identical singular and plural forms, the normalisation tool preserves the number of original phrases (singular or plural). As result in some cases the number of the base form of a particular NE is singular instead of plural or vice versa.
- For the multi-word expressions, the normalisation tool can create several base forms that comply with syntactical rules. As there is no disambiguation component that would take into account the semantics of the particular phrase, the first result from the result list is assumed as the correct one.

7 Entity Linking

The goal of the entity linking task is to associate entity mentions found in a text with corresponding

[1]News were collected from the news portals: www.idnes.cz, niezalezna.pl and censor.net Russian section

[2]https://github.com/huggingface/transformers/blob/main/examples/pytorch/language-modeling/run_mlm.py

entries in a Knowledge Base (KB) (Zheng et al., 2010). Traditional entity linking pipeline consists of three steps: mention detection, candidate selection, and disambiguation (Balog, 2018).

The mention detection step is described in the Section 5. Due to the small expected size of our cross-lingual knowledge base (the actual maximum number of KB entries produced in this Shared task by our systems was 939), we skip the candidate selection step. Instead, a simple consistency check is applied to filter out mentions which do not have the same type (Khosla and Rose, 2020). As result, the candidates are all entries in the Knowledge Base which have the same type as the entity mention, which we are attempting to link.

The candidate selection and disambiguation usually include a three-step process of candidate generation, candidate ranking, and unlinkable mention prediction (Shen et al., 2015). In our submission, the candidates are ranked using a mention-ranking model (Rahman and Ng, 2009) to decide whether an active mention is co-referent with a candidate antecedent. We follow the algorithm proposed by (Vīksna and Skadina, 2021): at first, we use LaBSE to obtain entity mention embeddings and then we apply cosine similarity to calculate the similarity between obtained embeddings and those in the Knowledge Base. The similarity threshold for early stopping is set at 0.95 - if the similarity is above the threshold, the process links the entity mention to the candidate mention and returns the candidate mention ID. If none of the candidate mentions has a similarity score above 0.6, the entity mention is considered not found in the Knowledge Base and is added as a new entry to the Knowledge Base. For entities with similarity scores between 0.6 and 0.95, the candidate with the highest score is selected for linking.

Usually, at the beginning of the text entities are introduced (named) carefully, e.g. with a full name (and acronym), while later in a text, when it is clear from the context what they refer to, entities are often used in the shortened form (Rychlikowski et al., 2021). For such cases, we introduce an additional linking step at the document level: for each entity mention, we check whether its name is part of another entity, e.g., encountering the name "Asia", it could be matched as part of "Asia Bibi". We perform this step before attempting to link an entity to the Knowledge Base.

We also check for organization and person name abbreviations and translations. At first, we identify entities that are surrounded by brackets (optionally, quoted). Then, if the entity immediately preceding it belongs to the same type, both entities are linked together as aliases.

8 Results

Table 3 summarizes the performance of our five systems. The best results in the entity recognition task have been achieved by System-3. System-3 does not use any additional datasets for NER training. However, the overall results differ very little, and may not be statistically significant.

	NER	Norm	Link cross-lang	Link document
System-1	0.890	0.587	0.644	0.716
System-2	**0.896**	**0.595**	0.668	0.712
System-3	**0.896**	**0.595**	**0.669**	**0.755**
System-4	0.885	0.584	0.668	0.727
System-5	0.881	0.587	0.666	0.702

Table 3: NER (Recognition, relaxed partial matching), normalization and linking (cross-language level and document level) results of Tilde systems, F scores

System-4 shows noticeable improvement for EVT detection (Table 4), which could be explained by additional XLM-R pretraining on recent news data. The performance of System-5, which was fine-tuned using additional KPWr data, is very poor in PER class. Our hypothesis is that the annotation guidelines for PER class differ significantly between KPWr and SlavNER datasets. This drawback is addressed by our ensemble System-1, which, despite being fine-tuned with XLM-R Base, achieves an overall F-score of 0.89, and for the LOC class shows better performance than System-3 (achieving an F-score of 0.944).

	S1	S2, S3	S4	S5
All	0.890	**0.896**	0.885	0.881
PER	**0.969**	**0.971**	0.930	0.906
LOC	**0.944**	0.934	0.932	0.938
ORG	0.843	0.848	**0.854**	**0.853**
PRO	0.689	**0.823**	0.761	0.796
EVT	0.273	0.300	**0.375**	0.267

Table 4: Entity recognition results evaluated on SlavNER test data (Relaxed partial matching, All 5 systems: S1 = System-1, S2 = System-2, ...), F scores

The ensemble system shows good overall performance but performs poorly on PRO class. Although the separate NER systems, fine-tuned to detect PRO entities on CNEC, KPWr, and FiNER data, performed better than the baseline on our test setup, the final system, trained on the combined dataset, did not generalize well.

The NER results vary slightly between languages (Table 5), with better scores for languages using Latin script.

	Recall	Precision	F score
cs (all)	0.885	0.945	0.914
ru (all)	0.878	0.884	0.880
pl (all)	0.869	0.932	0.899

Table 5: System-3 entity recognition results evaluated on SlavNER test data (Relaxed partial matching) by language

All our systems use the same normalization tool, therefore any differences in normalization results between our systems depend on the previous entity recognition step. The normalization results for our best-performing System-3 are summarized in Table 6. The normalization tool demonstrates good results for the Russian language (F-score 0.70), while for Polish (F-score 0.54) results are similar to Stanza, used for Czech language normalization.

	Recall	Precision	F score
PER	0.488	0.496	0.492
LOC	0.731	0.746	0.739
ORG	0.298	0.393	0.339
PRO	0.459	0.436	0.447
EVT	0.011	0.045	0.018
All corpora	0.566	0.627	**0.595**
cs (all)	0.561	0.522	0.541
ru (all)	0.692	0.716	0.704
pl (all)	0.474	0.623	0.539

Table 6: Entity normalization results evaluated on SlavNER test data (System-3)

The best results in entity linking task (in all tasks - document level, single- and cross-language) achieved System-3. Evaluation results for this system are summarized in Table 7. Since this task depends on mention detection task, results for the EVT class are poor. Our entity linking system is based on embeddings and in the case of organizations, it often fails to separate similar yet completely different organizations, e.g. our model considers ORG-Gazprom and ORG-Gazprombank as the same entity. When the output of System-2 and System-3 is compared (Table 3), we can see that document-level linking improves entity linking performance on the document level significantly (F-scores 0.712 and 0.755), while on the cross-lingual level its effects are negligible.

	Recall	Precision	F score
PER	0.713	0.764	0.738
LOC	0.813	0.787	0.800
ORG	0.422	0.416	0.419
PRO	0.428	0.615	0.505
EVT	0.102	0.241	0.144
All	0.660	0.677	**0.669**

Table 7: Entity linking results evaluated on SlavNER test data (Cross-language level, System-3)

9 Conclusions

In this paper, we presented a modular architecture for the Recognition, Normalization, Classification, and Cross-lingual linking of Named Entities in Slavic Languages. Each module (NER, normalization tool, and NE linker) is self-contained and could be improved independently from others. While none of the systems fine-tuned on additional datasets surpassed the XLM-R Large system fine-tuned on SlavNER data, the ensemble system seems promising and could be retrained again using the XLM-R Large model instead of XLM-R Base in order to obtain better results.

Limitations

Our best-performing systems use very large language models or are ensemble systems, resources required to train and run such systems are considerable.

Entity linking module performs an embedding comparison with all entities of a matching type found in the knowledge base. While the KB is small such an approach works fast, however, as the knowledge base grows, each additional entity adds to the search time. For large knowledge bases, some form of candidate selection method would be necessary.

Acknowlegements

This research has been supported by the European Regional Development Fund within the research project “AI Assistant for Multilingual Meeting Management” No. 1.1.1.1/ 19/A/082.

References

Mikhail Arkhipov, Maria Trofimova, Yuri Kuratov, and Alexey Sorokin. 2019. Tuning multilingual transformers for language-specific named entity recognition. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 89–93, Florence, Italy. Association for Computational Linguistics.

Krisztian Balog. 2018. *Entity Linking*, pages 147–188. Springer International Publishing, Cham.

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. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Daiga Deksne. 2013. Finite state morphology tool for Latvian. In *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing*, pages 49–53, St Andrews, Scotland. Association for Computational Linguistics.

Daiga Deksne, Inguna Skadina, and Raivis Skadins. 2014. Extended CFG formalism for grammar checker and parser development. In *Computational Linguistics and Intelligent Text Processing - 15th International Conference, CICLing 2014, Kathmandu, Nepal, April 6-12, 2014, Proceedings, Part I*, volume 8403 of *Lecture Notes in Computer Science*, pages 237–249. Springer.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Pedro Ferreira, Ruben Cardoso, and Afonso Mendes. 2021. Priberam labs at the 3rd shared task on SlavNER. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 86–92, Kiyv, Ukraine. Association for Computational Linguistics.

Sopan Khosla and Carolyn Rose. 2020. Using type information to improve entity coreference resolution. In *Proceedings of the First Workshop on Computational Approaches to Discourse*, pages 20–31, Online. Association for Computational Linguistics.

Michał Marcińczuk. 2020. KPWr n82 NER model (on polish RoBERTa base). CLARIN-PL digital repository.

James Mayfield, Paul McNamee, and Cash Costello. 2017. Language-independent named entity analysis using parallel projection and rule-based disambiguation. In *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, pages 92–96, Valencia, Spain. Association for Computational Linguistics.

Maciej Ogrodniczuk and Łukasz Kobyliński, editors. 2018. *Proceedings of the PolEval 2018 Workshop*. Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.

Marcin Oleksy, Michał Marcińczuk, Marek Maziarz, Tomasz Bernaś, Jan Wieczorek, Agnieszka Turek, Dominika Fikus, Michał Wolski, Marek Pustowaruk, Jan Kocoń, and Paweł Kędzia. 2019. Polish corpus of wrocław university of technology 1.3. CLARIN-PL digital repository.

Jakub Piskorski, Bogdan Babych, Zara Kancheva, Olga Kanishcheva, Maria Lebedeva, Michał Marcińczuk, Preslav Nakov, Petya Osenova, Lidia Pivovarova, Senja Pollak, Pavel Přibáň, Ivaylo Radev, Marko Robnik-Sikonja, Vasyl Starko, Josef Steinberger, and Roman Yangarber. 2021. Slav-NER: the 3rd cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic languages. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 122–133, Kiyv, Ukraine. Association for Computational Linguistics.

Jakub Piskorski, Laska Laskova, Michał Marcińczuk, Lidia Pivovarova, Pavel Přibáň, Josef Steinberger, and Roman Yangarber. 2019. The second cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic languages. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 63–74, Florence, Italy. Association for Computational Linguistics.

Jakub Piskorski, Lidia Pivovarova, Jan Šnajder, Josef Steinberger, and Roman Yangarber. 2017. The first cross-lingual challenge on recognition, normalization, and matching of named entities in Slavic languages. In *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, pages 76–85, Valencia, Spain. Association for Computational Linguistics.

Marko Prelevikj and Slavko Zitnik. 2021. Multilingual named entity recognition and matching using BERT and dedupe for Slavic languages. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 80–85, Kiyv, Ukraine. Association for Computational Linguistics.

Adam Przepiórkowski. 2012. *Narodowy korpus języka polskiego*. Naukowe PWN.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

Altaf Rahman and Vincent Ng. 2009. Supervised models for coreference resolution. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 968–977, Singapore. Association for Computational Linguistics.

Teemu Ruokolainen, Pekka Kauppinen, Miikka Silfverberg, and Krister Lindén. 2019. A finnish news corpus for named entity recognition. *Language Resources and Evaluation*, pages 1–26.

Paweł Rychlikowski, Bartłomiej Najdecki, Adrian Lancucki, and Adam Kaczmarek. 2021. Named entity recognition and linking augmented with large-scale structured data. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 115–121, Kiyv, Ukraine. Association for Computational Linguistics.

Stefan Schweter and Alan Akbik. 2020. FLERT: Document-level features for named entity recognition.

Magda Ševčíková, Zdeněk Žabokrtský, Jana Straková, and Milan Straka. 2014. Czech named entity corpus 2.0. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

W. Shen, J. Wang, and J. Han. 2015. Entity linking with a knowledge base: Issues, techniques, and solutions. *IEEE Transactions on Knowledge and Data Engineering*, 27(2):443–460.

Anatoli Starostin, Victor Bocharov, Svetlana Alexeeva, A. A. Bodrova, Alexander Chuchunkov, Sh. Sh. Dzhumaev, Irina Efimenko, D V Granovsky, Vladimir F. Khoroshevsky, Irina V. Krylova, Maria Nikolaeva, Ivan Smurov, and Svetlana Toldova. 2016. Factrueval 2016: Evaluation of named entity recognition and fact extraction systems for russian. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016"*.

Erik F. Tjong Kim Sang. 2002. Introduction to the CoNLL-2002 shared task: Language-independent named entity recognition. In *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*.

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.

Tatiana Tsygankova, Stephen Mayhew, and Dan Roth. 2019. BSNLP2019 shared task submission: Multisource neural NER transfer. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 75–82, Florence, Italy. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *CoRR*, abs/1706.03762.

Rinalds Vīksna and Inguna Skadina. 2021. Multilingual Slavic named entity recognition. In *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, pages 93–97, Kiyv, Ukraine. Association for Computational Linguistics.

Daniel H Younger. 1967. Recognition and parsing of context-free languages in time n3. *Information and control*, 10(2):189–208.

Zhicheng Zheng, Fangtao Li, Minlie Huang, and Xiaoyan Zhu. 2010. Learning to link entities with knowledge base. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 483–491, Los Angeles, California. Association for Computational Linguistics.