

Multilingual Transformers for Named Entity Recognition

Rinalds VĪKSNA^{1,2}, Inguna SKADIŅA^{1,2}

¹ Tilde, Vienības gatve 75a, Rīga, Latvia

² University of Latvia, Raiņa bulv. 29, Rīga, Latvia

`rinalds.viksna@Tilde.lv, inguna.skadina@tilde.lv`

Abstract. Different methods for automatic named entity recognition (NER) have been researched for many years. Today, the most common technique for training named entity recognition models is a fine-tuning of large pre-trained language models. In this paper, we investigate the performance of various multilingual NER models in the state-of-the-art natural language processing framework Flair and compare them against the multilingual NER solution of the MAPA anonymization toolkit and BERT multilingual model, fine-tuned for NER. We demonstrate that in multilingual settings the best results could be achieved with fine-tuned XLM-R model, while in the case of Latvian (monolingual settings), the more targeted LitLat BERT model leads to the best results.

Keywords: named entity recognition, pre-trained language models, multilingual models, FLAIR

1 Introduction

Automatic methods for named entity recognition (NER) have been researched for many years, mostly in the context of information extraction, data mining, and entity linking.

NER is an important constituent for different natural language understanding tasks, such as semantic annotation, question answering, ontology population, and opinion mining (Marrero et al., 2013). A rather new direction that recently has raised interest in NER, is text de-identification and anonymization. The need for text data anonymization (or at least de-identification) has become essential with the changes in legislation on the protection of natural persons with regard to the processing of personal data³ and opening public service data.

The goal of the anonymization task is to anonymize text (see Fig.1), making impossible to identify the natural person mentioned in the text directly or indirectly. Text anonymization is an extremely difficult task, because even seemingly unrelated facts

³ <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

may allow an informed attacker to identify involved persons (Gianola et al., 2020; Lison et al., 2021).

De-identification is easier task, aiming to delete specific identifiers (e.g., person names, place names, etc.) which could allow identification of an individual (Chevrier et al., 2019), or to replace such identifiers with similar ones, thus producing a readable text, which does not allow to find the identity of the original entity. The traditional automatic text de-identification approach uses the NER to identify named entities and other identifiers and then hides identified entities using either data masking or pseudonymization techniques. Usually, the performance of text de-identification system is evaluated by comparing output from de-identification systems to human-annotated gold standard and calculating precision, recall, and F1-measure (Pilán et al., 2022).

- (0) Trump is the only federal officeholder in American history to have been impeached twice.
 (1) ██████ is the only ██████ in ██████ history to have been ██████ twice.
 (2) ██████ is the only federal officeholder in ██████ history to have been impeached twice.

Fig. 1. Illustration of anonymization and de-identification: Original sentence (0), anonymized version (1), and de-identified version .

The state-of-the-art technique used for named entity identification and recognition is fine-tuning a large pre-trained language model for NER task. This approach became popular with the introduction of BERT models (Devlin et al., 2019a) in 2018. Since that time different transformer models have been introduced (e.g., XLM-R (Conneau et al., 2020), ALBERT (Lan et al., 2020), BART (Lewis et al., 2020) and others).

In this work, we explore the use of multilingual Transformer models for the NER task in multilingual settings, using FLAIR (Akbik et al., 2018) and MAPA libraries. The NER model using fine-tuned mBERT pre-trained by Devlin et al. (2019a) is also included as a baseline. As we are primarily interested in Baltic languages, we also include a litlat BERT (Ulčar and Robnik-Šikonja, 2020) in our evaluation. While overall goal is to analyse applicability of different NER solution in de-identification process, in this paper we mainly focus on performance of multilingual Transformer models in NER task, treating all named entity types equally.

2 Related Work

CLARIN Resource Families⁴ of Tools for named entity recognition⁵ lists 26 tools that support named entity recognition: 17 tools support a single language (4 Dutch, 2 English, 1 Finnish, 2 German, 1 Icelandic, 1 Greek, 1 Hungarian, 1 Latvian, 3 Polish, 1 Portuguese), while the rest have a very broad multilingual scope, however, none of the tools support all European languages, including Baltic languages.

The use of transfer learning for building multilingual NER has been explored by Rahimi et al. (2019) using multilingual fasttext (Bojanowski et al., 2017) embeddings

⁴ <https://www.clarin.eu/resource-families>

⁵ <https://www.clarin.eu/resource-families/tools-named-entity-recognition>

and Wikiann NER corpus (Pan et al., 2017) for training. As the sequential tagger, a BiLSTM-CRF (Fang and Cohn, 2016) is used.

Tedeschi et al. (2021) use Wikipedia as a source for multilingual annotations and proceed to build a multilingual NER model using multilingual BERT embeddings. The dataset produced in this manner contains CoNLL entity types (PER, ORG, LOC, and MISC). As the sequential tagger, an architecture based on BiLSTM with a CRF layer (Lafferty et al., 2001) is used.

Baumann (2019) evaluated multilingual language models used to train NER models in English and German on a CONLL dataset. For English, the NER using the native BERT model outperforms the mBERT model. For German, the NER using mBERT model embeddings outperformed most other NER models except Flair (Akbik et al., 2018) embeddings, showing the importance of the embedding layer.

Besides dedicated NER tools and natural language processing pipelines that support NER, several de-identification toolkits, such as Presidio or MAPA, include solutions for NER task. MAPA toolkit contains a built-in NER component, while Presidio uses either Flair (Akbik et al., 2019a), spaCy (Honnibal and Montani, 2017), or other NER models.

The Multilingual Anonymisation Toolkit for Public Administrations (MAPA) project⁶, aims to leverage natural language processing tools for text de-identification. It relies on named entity recognition and classification techniques that aim to detect relevant entities and decide which ones must be masked. MAPA entity detection models use multilingual BERT to convert text into embeddings which are fed into a two-level Classifier used to detect entities according to MAPA two-level scheme (Ajausks et al., 2020). For comparison purposes with other NER models, in this work, we evaluate only Level 1 annotations.

Flair⁷ is a state-of-the-art natural language processing (NLP) framework designed to facilitate training and distribution of sequence labeling, text classification and language models (Akbik et al., 2019a). Flair includes text embedding libraries, NLP models, and helper classes intended for data handling. The text embedding libraries included in the Flair framework allow the use and combination of different word and document embeddings, including ELMo (Peters et al., 2018), Flair (Akbik et al., 2018) and Transformer based BERT (Devlin et al., 2019b) and XLM-R (Conneau et al., 2020) embeddings. The NLP models are exposed to the user using standardized training and hyperparameter selection routines allowing researchers to mix various models and embeddings (Akbik et al., 2019b). Flair framework allows the use of document-level features with transformer-based models achieving state-of-the-art NER results (Schweter and Akbik, 2020).

3 Datasets, Annotations and Formats

To test the performance of transformers for the NER task two different multilingual datasets (BSNLP 2021 and EUR-LEX) and two monolingual (one Latvian and one English) datasets were used. The size and languages each dataset represents differ, Table 1 summarizes the number of entities in each dataset used for analysis.

⁶ <https://gitlab.com/MAPA-EU-Project/mapa-project>

⁷ <https://github.com/flairNLP/flair>

Table 1. Number of entity mentions in selected datasets

Language	Conll-2003	FullStack-LV	BSNLP 2021	EUR-LEX
bg			25537	1305
cs			14977	1396
da				1492
de				1467
el				1527
en	35097			1442
es				1569
et				1509
fi				1473
fr				1551
hr				1295
hu				1170
it				952
lt				1343
lv		9697		1599
mt				2180
nl				1195
pl			28099	1153
pt				1468
ro				1343
ru			13810	
sk				1189
sl			11639	1307
sv				1492
uk			4733	
all	35097	9697	98795	32417

Formats of these datasets include conll-2003, conll-U, WebAnno tsv, json, jsonlines and BSNLP. In order to train baseline BERT and Flair models, all datasets were converted into a conll-like format: one token and its named entity label per line, with empty lines representing sentence and document boundaries. To train MAPA named entity recognizer datasets were converted into the jsonlines format.

The datasets also differ in respect to labeled entity types (see Table 2). Even in cases when the same-named entity types are annotated, domains and annotation guidelines could differ. In the next sub-sections we summarize information about each dataset.

3.1 The English Conll 2003 Named Entity Dataset

The English Conll-2003 named entity dataset (Tjong Kim Sang and De Meulder, 2003) consists of training, development, and test files and a file with unannotated data. Annotated data are tokenized, part-of-speech tagged, and chunked. Named entity annotation has been performed manually, mostly following MUC conventions (Chinchor and Robinson, 1998), using 4 named entity categories: location (LOC), person (PER), organization (ORG), and MISC (contains all names not included in other categories). Data is

Table 2. Number of entities in different data sets

Entity type	Conll-2003	FullStack-LV	BSNLP 2021	EUR-LEX
Person	10059	3104	20494	9522
Location	10647	677	32536	
MISC	5062			
Organization	9323	1847	25338	6992
Entity		215		
GPE		2031		
Date/Time		1227		11084
Product		293	5535	
Money		44		
Event		259	6158	
Vehicle				13
Address				4047
Amount				749
All	35091	9697	90061	32407

provided in a Conll-2003 format, with one word per line with empty lines representing sentence boundaries. Each line contains four fields: the word, its part-of-speech tag, its chunk tag, and its named entity tag. Multi-word entities are labeled using BIO (Begin, Inside, Outside) labeling schema.

3.2 Multilingual BSNLP 2021 Dataset

BSNLP dataset⁸ is annotated corpus in 6 languages: Bulgarian, Czech, Polish, Russian, Slovenian, and Ukrainian, created for the shared task of 8th BSNLP: Balto-Slavic Natural Language Processing Workshop (Piskorski et al., 2021). The dataset contains a list of five types of named entity mentions: person (PER), organization (ORG), location (LOC), event (EVT), and product (PRO). Each named entity mention has a unique identifier assigned to link the same entity across different languages and documents and a lemma or canonical form of multi-word expression, if given named entity is a multi-word entity. Documents in this corpus are represented as a text file with metadata and a corresponding annotation file. Annotation file contains an identifier of the text file and a list of tab-separated fields: mention, its base form, named entity type, and named entity Id. In order to train and test NER systems, we converted the data into conll-2003-like data format, by keeping only word and named entity label in each line.

3.3 Multilingual MAPA Dataset

MAPA de-identification package contains 12 annotated documents from Court of Justice of the European Union. The documents are translated into 24 European languages, annotated according to MAPA annotation schema and available ELRC share as 24 separate datasets⁹. MAPA annotation schema has 2 levels of entities. Top-level contains 6

⁸ <http://bsnlp.cs.helsinki.fi/shared-task.html>

⁹ <https://elrc-share.eu/repository/search/?q=MAPA>

entity classes: Person, Time, Location, Organization, Amount, and Vehicle. The second annotation level consists of attributes or subdivisions of top-level entities, e.g. entity PERSON may have second-level annotations "title", "initial name", "marital status" and so on. The second level contains 57 entity classes, however, most of them are not used in the MAPA EUR-LEX dataset. The data files are provided in WebAnno TSV 3.2 format (Eckart de Castilho et al., 2016). To train MAPA toolkit models, a conversion to JSON format is performed using the conversion script included in the MAPA project. 12 documents for each language were split into 8 for training, 2 for development, and 2 for test documents and kept translated versions of the same document together.

3.4 Latvian FullStack Corpus

FullStack-LV corpus¹⁰ (Grūzītis et al., 2019) contains hierarchical named entity annotations in Latvian. In cases when outer layer annotations contain nested named entities, nested entities are annotated as inner layer entities. Fullstack-LV corpus contains 3947 paragraphs of text with 9697 outer entities and 944 inner entities (Gruzitis et al., 2018). In contrast with the MAPA dataset, where second-level entities are parts of the first-level entity, in the FullStack-LV dataset inner entities are independent of outer entities. Because most of the tools tested are intended to detect single-level named entities, we discard inner entity annotations. Dataset has been labeled using 9 entity types: Geopolitical, Person, Time, Location, Product, Organization, Money, Event, and Entity (for all entities, which do not belong to any other classes). The data is provided in a version of Conll-2003 data format, single document per file. The data was converted to a simpler version of Conll-2003-like format keeping only word and entity labels in each line and joining all files in training, development and test sets into appropriate files.

4 Experimental Setup

To evaluate different NER models, we train a NER model using given framework, language model, and dataset. All NER models were trained for 20 epochs or until early stopping conditions were met. The training parameters (the number of epochs, batch size, and learning rate) were selected as it was recommended in documentation. The experiments were run 3 times and the results were averaged.

For a baseline, we fine-tune the multilingual BERT model in the same way as in original BERT paper (Devlin et al., 2019a) - by adding a single output classification layer and jointly fine-tuning all parameters on the NER task. Fine-tuning is performed on BERT-base Multilingual Cased model¹¹ using batch size 16, learning rate 2e-5 for 20 epochs.

Within the Flair framework, we explore multiple approaches for NER model training: multilingual BERT (mBERT), multilingual XLM RoBERTa (XLM-R). We also test the Litlat BERT (Ulčar and Robnik-Šikonja, 2020), which has been pre-trained on Latvian, Lithuanian, and English data. The "Flair Litlat", "Flair mBERT" and "Flair

¹⁰ <https://github.com/LUMII-AILab/FullStack>

¹¹ <https://github.com/google-research/bert>

”XLM-R” models were trained using the Flair Sequence labeling approach (Akbik et al., 2018) and Litlat, multilingual BERT or XLM-R language models. The ”Flert mBERT” and ”Flert XLM-R” models were trained using Flair’s ”Flert” (Schweter and Akbik, 2020) approach and using multilingual BERT or XLM-R language models. The Flert approach uses a transformer model fine-tuned using the target sentence and document-level features (64 subtokens of left and right context).

MAPA named entity hierarchy is hard-coded to detect 6+59 named entities by default, therefore, MAPA named entity recognizer was adapted to each dataset by editing appropriate labeling hierarchy. Training was performed using default settings specified in MAPA documentation.

5 Results

The overall results of the evaluation are summarised in Table 3. For comparison we also include the Conll-2003 Named Entity Recognition result (BERT-base) from original BERT paper (Devlin et al., 2019a). For both multilingual datasets, the best results are achieved by Flert XLM-R model. While for the Latvian litlat bert is the best in this evaluation, it still performs worse than monolingual LVBERT (Znotiņš and Barzdins, 2020) and none of the multilingual models exceed the F1 score of English monolingual BERT.

Table 3. Evaluation results (F-score)

Dataset	Language	BERT -base	mBERT	MAPA	Flair mBERT	Flair XLM-R	Flair litlat	Flert mBERT	Flert XLM-R
Conll-2003	English	92.4	91.27	90.35	90.59	89.36	89.60	91.58	92.37
FullStack-LV	Latvian		78.91	77.36	78.29	79.42	81.97	79.13	81.02
BSNLP 2021	multi		96.11	94.73	95.36	95.11	92.07	96.38	96.71
EUR-LEX	multi		77.25	75.82	79.37	79.73	69.80	84.32	85.10

5.1 Named Entity Recognition of Baltic Languages

In monolingual settings, when NER is trained only on Latvian language data from Fullstack-LV dataset, the best result (81.97 F1 points) is achieved with LitLat BERT (pre-trained on Latvian, Lithuanian and English data) as a language model.

We also compare fine-tuned Latvian (Viksna and Skadina, 2020) BERT model results (81.91 F1 points) to Flair NER using LitLat BERT language model. LitLat Flair NER model achieves an F1 score of 81.97, which is close to the state-of-the-art monolingual LVBERT (Znotiņš and Barzdins, 2020) results (82.6 F1 points) on the Fullstack-LV dataset. However, LitLat BERT performs poorly for other languages on which it has not been pre-trained.

In the case of the multilingual EUR-LEX dataset XLM-R model demonstrates the best result for Latvian, reaching an F-score of 75, while mBERT model’s results are the best for Lithuanian, reaching F-score of 80 (see Table4).

The reason why results achieved for Latvian are in many cases lower than for other languages could be explained by the morphological richness of Latvian, its share in large multilingual language models, as well as, the size of training data.

5.2 Named Entity Recognition of Slavic Languages

For Slavic languages, we tested the Flert mBERT model, which demonstrated similar results (96.56 F-score) to the Flert XLM-R model. Although Slavic languages are morphologically rich languages and thus complicate to process, all models demonstrate better performance for Slavic languages than for English. This could be explained by the fact that the BSNLP data belongs to specific domain - news about a few selected themes (Brexit, Nord stream, Ryanair, and other).

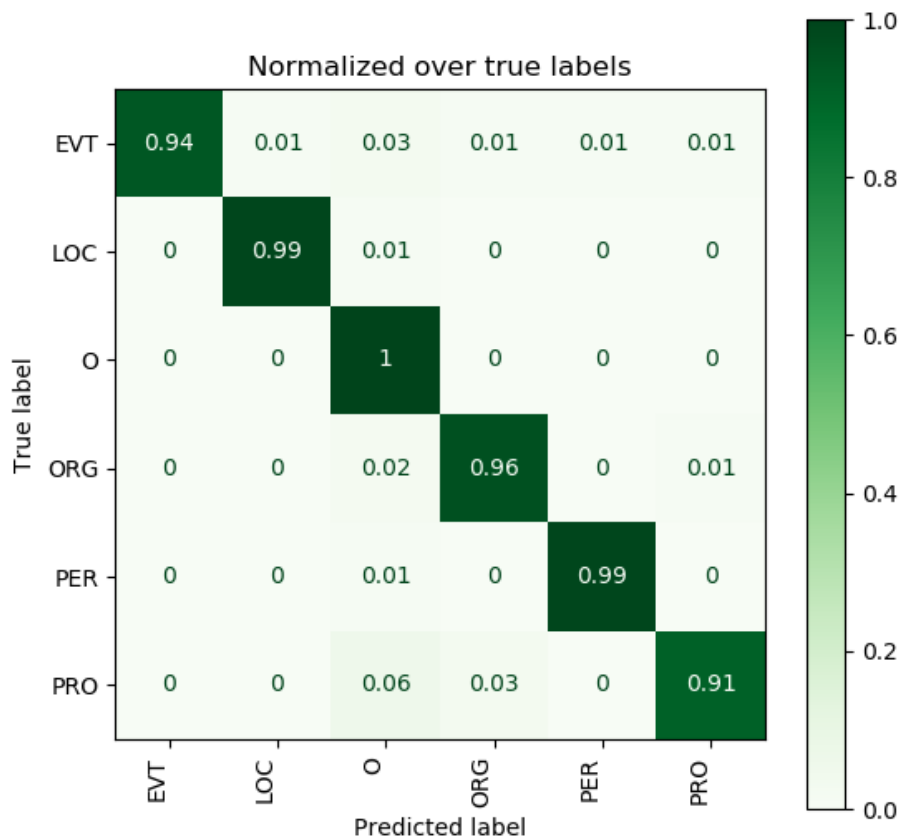


Fig. 2. Confusion matrix of Flert XLM-R model on BSNLP dataset, normalized over true labels

Figure 2 shows a confusion matrix for Flert XLM-R model on the BSNLP test set. Event class (EVT) consists of diverse set of entities: most misclassifications hap-

pen when event entity contains entity of another class, e.g. "Meeting Trump-Putin in Helsinki" (here both Person (PER) and Location (LOC) are present), or Event entity is not detected at all, e.g. "vietnam war", "All-Star" or "referendum of 2016". Entities of LOC type are most often misclassified as non-entities, e.g. "punjabi", "statue of Jesus Christ", or entities have wrong boundary as in "airport Charleroi", "airport Brussel" - airport is not detected as part of LOC entity. Similarly, most of the errors for Organisation (ORG) class are boundary errors with LOC entities, e.g. "Pakistan government" or "The Group of European Conservatives and Reformists". 47 instances of organizations are misclassified as Products (PRO), e.g. "Facebook", "YouGov", "Funke". Person class is misclassified as a non-entity in cases where it refers to groups of people ("Republican senators") or ethnicity ("pakistanis [wife]"). Product entities are mostly misclassified as non-entities, and also as organisations ("The Economist", "Bloomberg").

5.3 Transformer models in Multilingual Settings - Case of EUR-LEX

In case of EUR-LEX the Flert XLM-R model shows the best performance among multilingual models used.

Table 4. Detailed evaluation results of EUR-LEX dataset (F-score)

Language	MAPA	mBERT	Flair mBERT	Flair XLM-R	Flair litlat	Flert mBERT	Flert XLM-R
bg	78.95	79.41 ± 0.33	80 ± 2.17	75.62 ± 0.91	75.53 ± 2.32	75.62 ± 1.89	72.36 ± 1.94
cs	75.14	77.92 ± 3.03	80.58 ± 2.07	83.51 ± 0.61	72.53 ± 2.73	74.42 ± 2.53	83.5 ± 0.69
da	82.08	79.19 ± 2.01	80.6 ± 0.59	76 ± 0.84	69.11 ± 2.45	72.73 ± 2.43	82.59 ± 1.99
de	66.98	80.78 ± 2.93	74.35 ± 1.75	80 ± 1.41	77.42 ± 1.73	81.04 ± 2.15	82.17 ± 2.00
el	78	82.44 ± 1.98	81 ± 0.59	81.19 ± 0.85	76.92 ± 0.64	74.07 ± 4.29	84.65 ± 0.82
en	74.04	68.79 ± 1.40	71.07 ± 0.76	75.79 ± 0.72	70.72 ± 2.29	70.19 ± 2.52	73.1 ± 1.89
et	60.27	59.57 ± 2.81	59.59 ± 1.76	56.03 ± 0.98	41.15 ± 2.15	52.28 ± 0.94	59.26 ± 1.26
es	81.74	81.93 ± 0.24	81.24 ± 0.75	80 ± 1.56	72.3 ± 1.78	73.86 ± 1.11	79.28 ± 0.98
fi	79.44	74.67 ± 2.01	79.8 ± 0.23	81.44 ± 1.35	66.3 ± 1.48	79.25 ± 1.94	85.44 ± 3.70
fr	63.07	74.59 ± 1.15	72.65 ± 2.25	75.44 ± 0.62	64.25 ± 0.52	72.51 ± 4.25	78.11 ± 1.53
hr	75.76	81.24 ± 0.30	81.42 ± 1.19	83.41 ± 0.64	70.19 ± 2.34	80.17 ± 2.14	84.26 ± 3.14
hu	79.07	77.02 ± 1.64	77.08 ± 1.90	85.86 ± 2.41	65.26 ± 1.76	71.29 ± 4.21	86.17 ± 2.39
it	82.89	83.87 ± 2.59	76.24 ± 2.38	76.4 ± 1.86	71.01 ± 1.25	77.78 ± 2.65	77.27 ± 2.17
lt	72.63	80 ± 0.32	77.42 ± 1.36	74.73 ± 1.47	73.74 ± 2.63	80 ± 3.26	78.72 ± 2.21
lv	66.94	66.99 ± 1.37	72.45 ± 1.22	72.58 ± 2.63	66.12 ± 1.91	70.85 ± 6.35	75 ± 0.35
mt	82.42	72.95 ± 2.49	70.97 ± 0.69	66.25 ± 1.86	70.52 ± 2.81	62.27 ± 1.36	65.14 ± 2.94
nl	80.92	85.48 ± 0.78	80.87 ± 1.78	81.61 ± 1.24	80 ± 1.30	76.76 ± 3.47	85.23 ± 0.94
pl	73.43	81.08 ± 1.42	81.48 ± 1.66	80.77 ± 0.22	68.02 ± 0.70	80.37 ± 1.94	83.11 ± 3.12
pt	80.75	80.21 ± 0.33	79.17 ± 0.43	81.74 ± 0.93	76.79 ± 0.62	83.4 ± 0.87	77.69 ± 1.01
ro	73.68	73.68 ± 2.60	82.61 ± 1.82	85.39 ± 0.74	81.36 ± 1.70	76.92 ± 0.28	84.95 ± 3.17
sk	78.07	81.12 ± 2.47	73.03 ± 3.93	75.74 ± 2.90	70.24 ± 1.06	75.27 ± 1.79	71.11 ± 2.51
sl	78.12	85.54 ± 2.29	82.79 ± 1.75	86.54 ± 0.62	81.95 ± 1.19	79.65 ± 4.71	89.09 ± 0.71
sv	79.6	78.43 ± 1.02	77.45 ± 1.61	78.31 ± 1.68	72.83 ± 1.69	79.61 ± 3.28	83.08 ± 1.43

The NER model based on trilingual LitLat, trained using EUR-LEX data, is performing unexpectedly well, considering that most of the training data are in languages it has never seen before. Finally, since the MAPA EUR-LEX dataset is small, and the test set consists of around 200 entity mentions per language, thus few misclassified entities lead to large F1 score differences.

6 Conclusions

In this paper, we examined the NER performance of several toolkits which use multilingual Transformer language models. Our experiments show that multilingual Transformer models achieve NER performance comparable to monolingual Transformer models trained and fine-tuned on target language data. However, if a language model (such as LitLat) is fine-tuned for a language that is not represented in training data, the resulting model demonstrates poor performance. In our experiments, the NER using fine-tuned XLM-R (the "Flert XLM-R") model shows the best results for multilingual NER task. Such multilingual NER could be used for named entity identification in multilingual de-identification tools, allowing usage of single model, instead of separate models for each language. However, analysis of applicability of particular NER solution in de-identification pipeline, taking into account importance of different types of entities in de-identification process, is the next task of our research.

7 Acknowledgements

This work has been funded by the European Regional Development Fund within the research project "AI Assistant for Multilingual Meeting Management" No. 1.1.1.1/19/A/082. We would also like to thank reviewers for their valuable remarks on the content of the paper.

References

- Ajauskas, Ē., Arranz, V., Bié, L., Cerdà-i Cucó, A., Choukri, K., Cuadros, M., Degroote, H., Estela, A., Etchegoyhen, T., García-Martínez, M., García-Pablos, A., Herranz, M., Kohan, A., Melero, M., Rosner, M., Rozis, R., Paroubek, P., Vasiļevskis, A., Zweigenbaum, P. (2020). The multilingual anonymisation toolkit for public administrations (MAPA) project, *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, European Association for Machine Translation, Lisboa, Portugal, pp. 471–472. <https://aclanthology.org/2020.eamt-1.57>
- Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., Vollgraf, R. (2019a). Flair: An easy-to-use framework for state-of-the-art nlp, *NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 54–59.
- Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., Vollgraf, R. (2019b). FLAIR: An easy-to-use framework for state-of-the-art NLP, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, Association for Computational Linguistics, Minneapolis, Minnesota, pp. 54–59. <https://aclanthology.org/N19-4010>

- Akbik, A., Blythe, D., Vollgraf, R. (2018). Contextual string embeddings for sequence labeling, *Proceedings of the 27th International Conference on Computational Linguistics*, Association for Computational Linguistics, Santa Fe, New Mexico, USA, pp. 1638–1649.
<https://aclanthology.org/C18-1139>
- Baumann, A. (2019). Multilingual language models for named entity recognition in German and English, *Proceedings of the Student Research Workshop Associated with RANLP 2019*, INCOMA Ltd., Varna, Bulgaria, pp. 21–27.
<https://aclanthology.org/R19-2004>
- Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. (2017). Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics* **5**, 135–146.
<https://aclanthology.org/Q17-1010>
- Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., Lovis, C. (2019). Use and understanding of anonymization and de-identification in the biomedical literature: Scoping review, *J Med Internet Res* **21**(5), e13484.
<https://doi.org/10.2196/13484>
- Chinchor, N., Robinson, P. (1998). Appendix E: MUC-7 named entity task definition (version 3.5), *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference Held in Fairfax, Virginia, April 29 - May 1, 1998*.
<https://aclanthology.org/M98-1028>
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, pp. 8440–8451.
<https://aclanthology.org/2020.acl-main.747>
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (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, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 4171–4186.
<http://aclweb.org/anthology/N19-1423>
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019b). 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, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171–4186.
<https://aclanthology.org/N19-1423>
- Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., Biemann, C. (2016). A web-based tool for the integrated annotation of semantic and syntactic structures, *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, The COLING 2016 Organizing Committee, Osaka, Japan, pp. 76–84.
<https://aclanthology.org/W16-4011>
- Fang, M., Cohn, T. (2016). Learning when to trust distant supervision: An application to low-resource POS tagging using cross-lingual projection, *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, Association for Computational Linguistics, Berlin, Germany, pp. 178–186.
<https://aclanthology.org/K16-1018>
- Gianola, L., Ajauskas, E., Arranz, V., Bendahman, C., Bié, L., Borg, C., Cerdà, A., Choukri, K., Cuadros, M., de Gibert, O., Degroote, H., Edelman, E., Etchegoyhen, T., Torres, Á. F., Hernandez, M. G., García-Pablos, A., Gatt, A., Grouin, C., Herranz, M., Kohan, A., Lavergne,

- T., Melero, M., Paroubek, P., Rigault, M., Rosner, M., Rozis, R., van der Plas, L., Vīksna, R., Zweigenbaum, P. (2020). Automatic removal of identifying information in official eu languages for public administrations: The mapa project, *JURIX*.
- Grūzītis, N., Pretkalniņa, L., Saulīte, B., Rituma, L., Nešpore-Bērzkalne, G., Paikens, P., Auziņa, I., Znotiņš, A., Levāne-Petrova, K., Dargis, R. (2019). Full stack of latvian language resources for NLU. CLARIN-LV digital library at IMCS, University of Latvia. <http://hdl.handle.net/20.500.12574/5>
- Gruzitis, N., Pretkalnina, L., Saulite, B., Rituma, L., Nespore-Berzkalne, G., Znotins, A., Paikens, P. (2018). Creation of a balanced state-of-the-art multilayer corpus for NLU, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA), Miyazaki, Japan. <https://aclanthology.org/L18-1714>
- Honnibal, M., Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.
- Lafferty, J. D., McCallum, A., Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, p. 282–289.
- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R. (2020). Albert: A lite bert for self-supervised learning of language representations, *ArXiv abs/1909.11942*.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, pp. 7871–7880. <https://aclanthology.org/2020.acl-main.703>
- Lison, P., Pilán, I., Sanchez, D., Batet, M., Øvrelid, L. (2021). Anonymisation models for text data: State of the art, challenges and future directions, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Online, pp. 4188–4203. <https://aclanthology.org/2021.acl-long.323>
- Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., Gómez-Berbís, J. M. (2013). Named entity recognition: Fallacies, challenges and opportunities, *Computer Standards & Interfaces* **35**(5), 482–489. <https://www.sciencedirect.com/science/article/pii/S0920548912001080>
- Pan, X., Zhang, B., May, J., Nothman, J., Knight, K., Ji, H. (2017). Cross-lingual name tagging and linking for 282 languages, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, Vancouver, Canada, pp. 1946–1958. <https://aclanthology.org/P17-1178>
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L. (2018). Deep contextualized word representations, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Association for Computational Linguistics, New Orleans, Louisiana, pp. 2227–2237. <https://aclanthology.org/N18-1202>
- Pilán, I., Lison, P., Øvrelid, L., Papadopoulou, A., Sánchez, D., Batet, M. (2022). The text anonymization benchmark (TAB): A dedicated corpus and evaluation framework for text anonymization, *CoRR abs/2202.00443*. <https://arxiv.org/abs/2202.00443>

- Piskorski, J., Babych, B., Kancheva, Z., Kanishcheva, O., Lebedeva, M., Marcińczuk, M., Nakov, P., Osenova, P., Pivovarova, L., Pollak, S., Přibáň, P., Radev, I., Robnik-Sikonja, M., Starko, V., Steinberger, J., Yangarber, R. (2021). Slav-NER: the 3rd cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic languages, *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, Association for Computational Linguistics, Kiyv, Ukraine, pp. 122–133.
<https://aclanthology.org/2021.bsnlp-1.15>
- Rahimi, A., Li, Y., Cohn, T. (2019). Massively multilingual transfer for NER, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Florence, Italy, pp. 151–164.
<https://aclanthology.org/P19-1015>
- Schweter, S., Akbik, A. (2020). Flert: Document-level features for named entity recognition.
<https://arxiv.org/abs/2011.06993>
- Tedeschi, S., Maiorca, V., Campolungo, N., Cecconi, F., Navigli, R. (2021). WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER, *Findings of the Association for Computational Linguistics: EMNLP 2021*, Association for Computational Linguistics, Punta Cana, Dominican Republic, pp. 2521–2533.
<https://aclanthology.org/2021.findings-emnlp.215>
- Tjong Kim Sang, E. F., De Meulder, F. (2003). Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition, *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pp. 142–147.
<https://aclanthology.org/W03-0419>
- Ulčar, M., Robnik-Šikonja, M. (2020). LitLat BERT. CLARIN-LT digital library in the Republic of Lithuania.
<http://hdl.handle.net/20.500.11821/42>
- Viksna, R., Skadina, I. (2020). Large language models for latvian named entity recognition, in Andrius, U., Jurgita, V., Jolantai, K., Danguole, K. (eds), *Human Language Technologies - The Baltic Perspective - Proceedings of the Ninth International Conference Baltic HLT 2020, Kaunas, Lithuania, September 22-23, 2020*, Vol. 328 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, pp. 62–69.
<https://doi.org/10.3233/FAIA200603>
- Znotiņš, A., Barzdins, G. (2020). *LVBERT: Transformer-Based Model for Latvian Language Understanding*, IOS Press Ebooks, pp. 111–115.

Received August 20, 2022 , accepted August 27, 2022