

# **BERT downstream task analysis: Named Entity Recognition in Serbian**

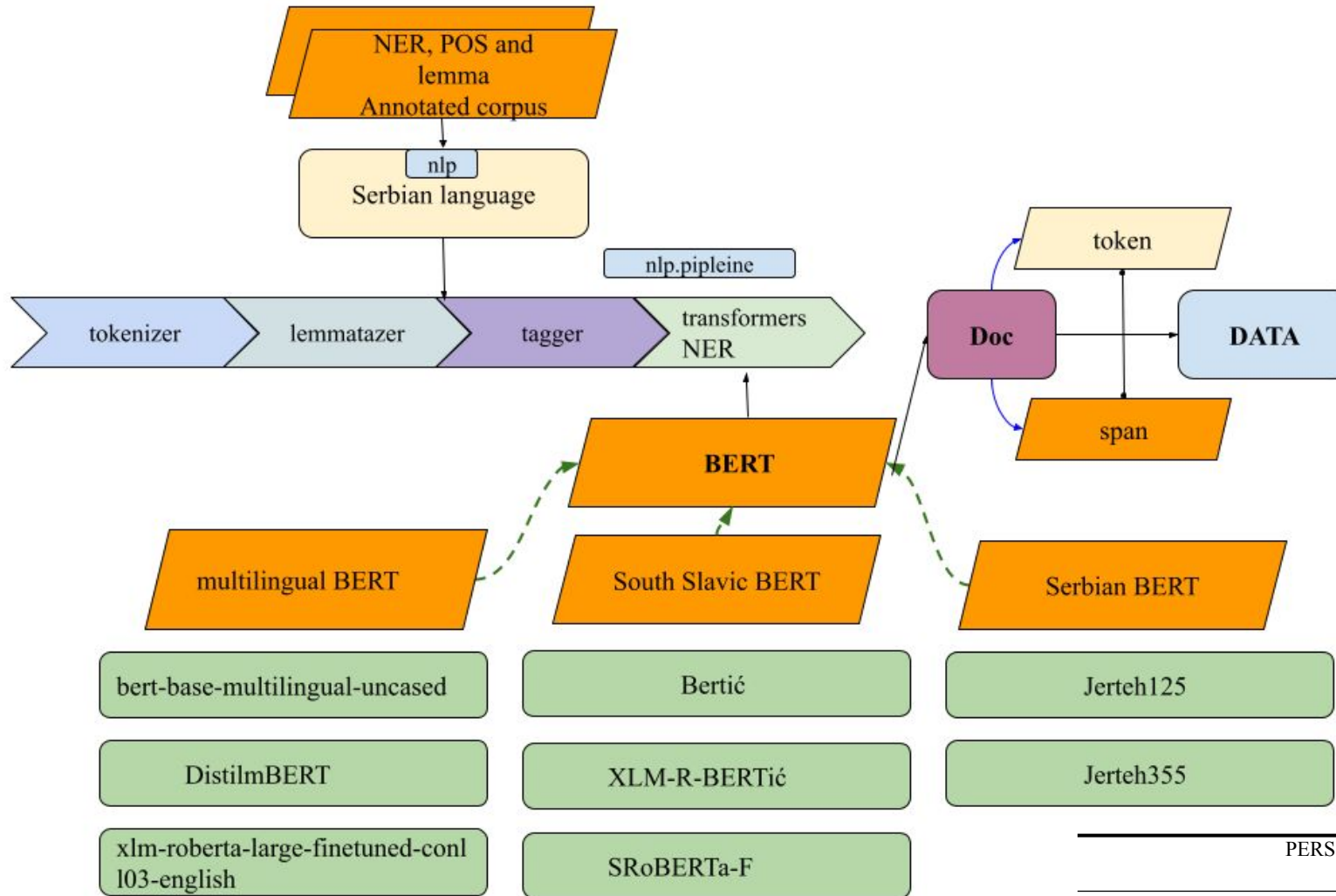
Milica Ikonić Nešić<sup>1</sup>, Saša Petalinkar<sup>2</sup>, Mihailo Škorić<sup>3</sup>, Ranka Stanković<sup>3</sup>

Serbia, University of Belgrade: <sup>1</sup>Faculty of Philology,  
<sup>2</sup>Multidisciplinary Studies, <sup>3</sup>Faculty of Mining and Geology, Serbia

# Research questions

- **RQ1.** Does the integration of BERT into the spaCy NER pipeline significantly improve model performance for Serbian language, and if so, is this improvement justifiable in terms of increased computational resources and processing time?
- **RQ2.** How can BERT be most effectively applied within the spaCy framework for Serbian NER tasks?
- **RQ3.** Is a multilingual BERT model sufficient for Serbian NER, or does a Serbian-specific BERT model offer improvements that justify the additional resources required for its development and training?

# Methodology



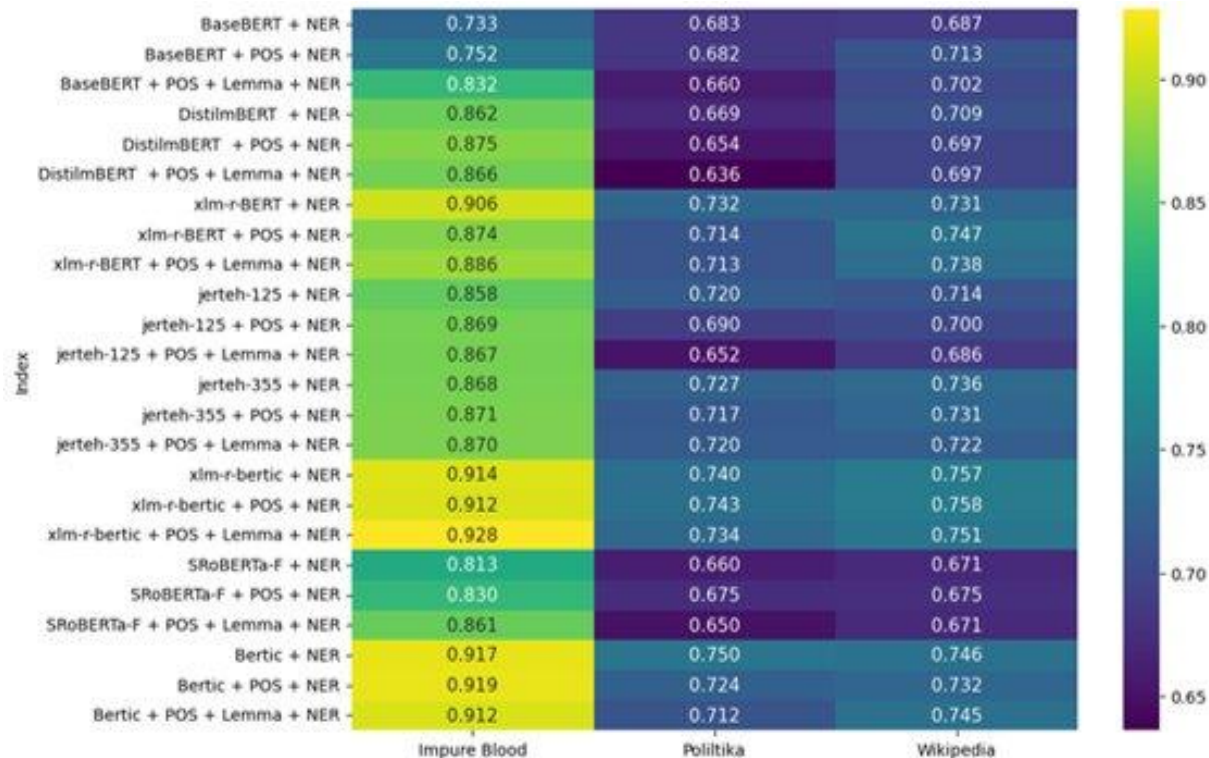
**148,819 sentences**

- Sentences generated from Wikidata and Leximirka
- srpELTeC-gold-extended
- Newspaper articles

Corpus NE distribution

	PERS	ROLE	DEMO	ORG	LOC	WORK	EVENT
<b>novels</b>	18505	11837	1844	432	2699	260	195
<b>newspaper</b>	4125	4213	4458	6012	9243	282	163
<b>generated</b>	56873	0	0	1376	27393	0	0
<b>TOTAL</b>	79503	16050	6302	7820	39335	542	358

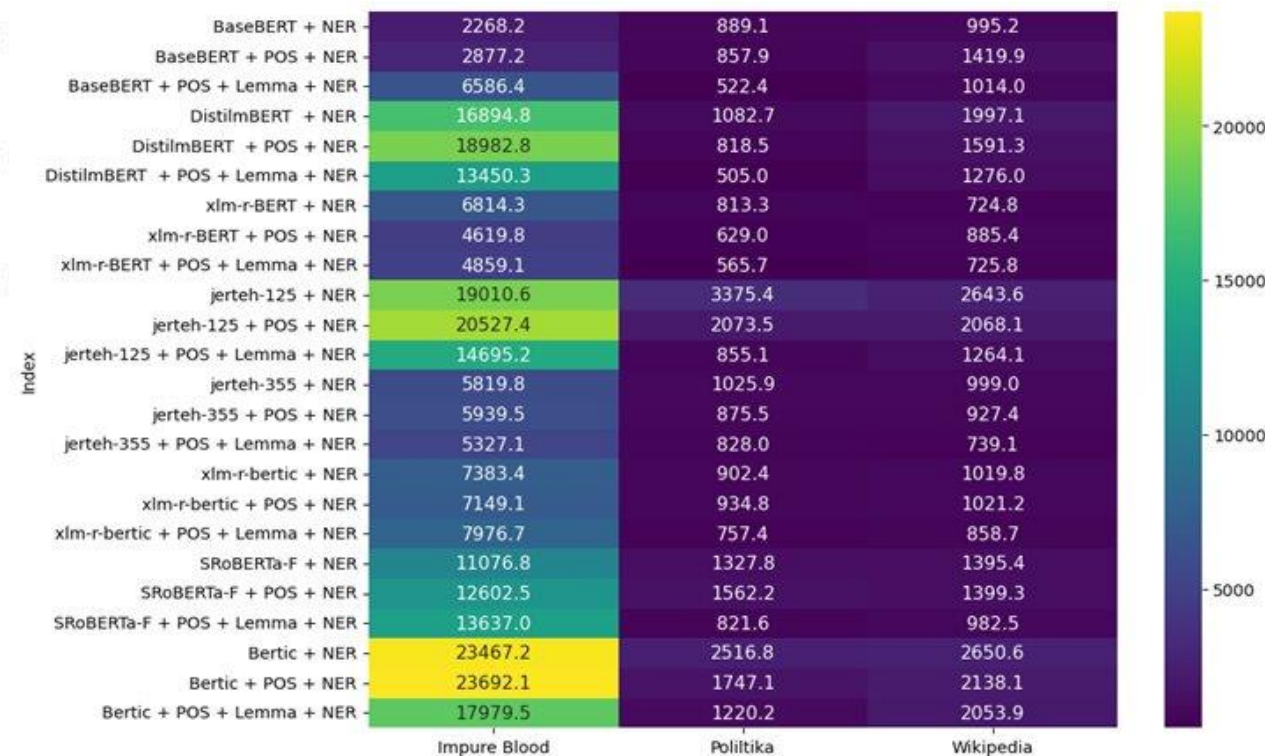
# Results



F1-score heatmap on pipelines on the evaluation tests with yellow meaning better results, and dark blue meaning worse results

NE distribution over evaluation sets

	PERS	ROLE	DEMO	ORG	LOC	WORK	EVENT	tokens
<b>Impure blood</b>	202	110	37	3	25	1	10	21876
<b>Politika</b>	495	180	67	301	458	5	16	22787
<b>Wikipedia</b>	231	129	154	127	350	21	27	16758



M-measure heatmap on the evaluation test with yellow meaning better results, and dark blue meaning worse results

# Applications

```
!pip install https://huggingface.co/Tanor/sr_pln_tesla_j125/resolve/main/sr_pln_tesla_j125-any-py3-none-any.whl
```

```
import spacy  
nlp = spacy.load("sr_pln_tesla_j125")
```

## An example of using Jerteh-355-Tesla model

[https://colab.research.google.com/drive/1XQ95DHSa\\_cn6B6L4RaRCchazbci-jtiY?usp=sharing](https://colab.research.google.com/drive/1XQ95DHSa_cn6B6L4RaRCchazbci-jtiY?usp=sharing)

Prvi svetski rat je trajao od 1914. do 1918. godine. U njemu je učestvovala većina velikih svetskih sila, grupisanih u dva sukobljena vojna saveza: **Saveznika ORG** (okupljenih oko **Trojne Antante ORG**) i **Centralnih sila ORG**. Više od 70 miliona ljudi je bilo pod oružjem, a od toga preko 60 miliona ljudi u **Evropi LOC** je bilo mobilisano u jedan od najvećih ratova u istoriji.

Posledice rata su bile da je ubijeno više od 15 miliona ljudi, 20 miliona ranjeno, a direktne učesnice rata pretrpele su i ogromna razaranja država i privreda.[4] Prvi svetski rat poznat je i pod imenima **Veliki rat EVENT** i **Svetski rat EVENT** (do izbijanja **Drugog svetskog rata EVENT**).

Prvi svetski rat su vodila dva velika saveza. Sile **Antante LOC** su na početku činile **Srbija LOC**, **Crna Gora LOC**, **Rusija LOC**, **Francuska LOC**, **Velika Britanija LOC** i **Japan LOC**. Centralne sile su pre početka rata činile **Nemačka LOC**, **Austrougarska LOC** i **Italija LOC**, koja je zbog **Londonskog ugovora WORK** od 26. aprila 1915. godine, kojim je za **Italiju LOC** predviđen deo **Dalmacije LOC**, **Istra LOC**, **Gorica LOC**, **Kvarnerska ostrva LOC** i **Dodokanezi LOC**, pristupila silama **Antante LOC**.

**Osmansko carstvo LOC** se pridružilo Centralnim silama oktobra 1914, a godinu dana kasnije to je uradila i **Bugarska LOC**. **SAD LOC** ulaze u rat 6. aprila 1917. posle objavljivanja **Cimermanovog telegrama WORK** kojim dobijaju povod. **Rumunija LOC** pristupa **Antanti LOC** 1916 kada je potpisan **Bukureški sporazum WORK**. **Grčka LOC** je na početku rata bila neutralna, ali 1917. pristupa strani **Antante LOC**. Do završetka rata, od **evropskih DEMO** zemalja **Holandija LOC**, **Švajcarska LOC**, **Španija LOC** i **skandinavске DEMO** države su ostale zvanično neutralne.

Neposredni povod za rat je bio atentat na naslednika **austrougarskog DEMO** prestola, **nadvojvodu ROLE** **Franca Ferdinanda PERS** u **Sarajevu LOC** 28. juna 1914, koga je ubio **Gavrilo Princip PERS**, **Srbin DEMO** iz **Bosne LOC**, koja je tada bila deo **Austrougarske LOC**.

Objava rata **Austrougarske LOC** **Srbiji LOC** aktivirala je niz savezništava koja su pokrenula lančanu reakciju objava rata. Do kraja avgusta 1914. veći deo **Evrope LOC** se našao u ratu.

[Tanor/sr\\_pln\\_tesla\\_j355 · Hugging Face](#)  
[Tanor/sr\\_ner\\_tesla\\_j355 · Hugging Face](#)  
[Tanor/sr\\_ner\\_tesla\\_j125 · Hugging Face](#)  
[Tanor/sr\\_pln\\_tesla\\_j125 · Hugging Face](#)  
[Tanor/sr\\_pln\\_tesla\\_bcx · Hugging Face](#)  
[Tanor/sr\\_ner\\_tesla\\_bcx · Hugging Face](#)

Thank you for your  
attention!

milica.ikonic.nesic@fil.bg.ac.rs

<https://tesla.rgf.bg.ac.rs/>

A research conducted with the support of  
the Science Fund of the Republic of  
Serbia, project number: 7276.

**Text Embeddings -  
Serbian Language Applications –**

**TESLA**

**ТЕСЛА**

