Entity Tagging & Linking

Mariana Almeida, Afonso Mendes, André Martins, Samuel Broscheit, Filipe Aleixo

mla@priberam.pt, amm@priberam.pt, afm@priberam.pt
samuel.broscheit@googlemail.com, filipe.aleixo@priberam.pt

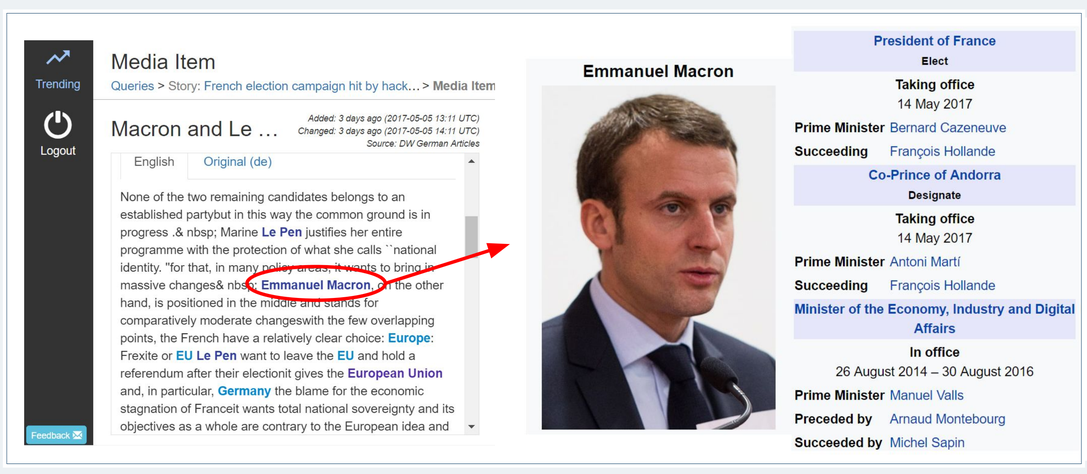
Goals

- To develop statistical models for entity recognition and linking.
- Explore multilingual approaches for the core languages of the project (English, German, Spanish) and for Portuguese.

Named Entity Recognition (NER)

- Strong industrial implementation that uses a sequence labeling model with gazetteers.
- New approaches using deep neural networks with bidirectional LSTMs and character- and word-level embeddings can achieve higher scores.

Entity Linking



- State-of-art industrial implementation, with document-level and copora-level coherence steps
- Good results for English and Spanish.
- The model has a small number of parameters, being easily portable between different languages.
- New approaches using deep neural networks are being tested.

What are the Challenges?

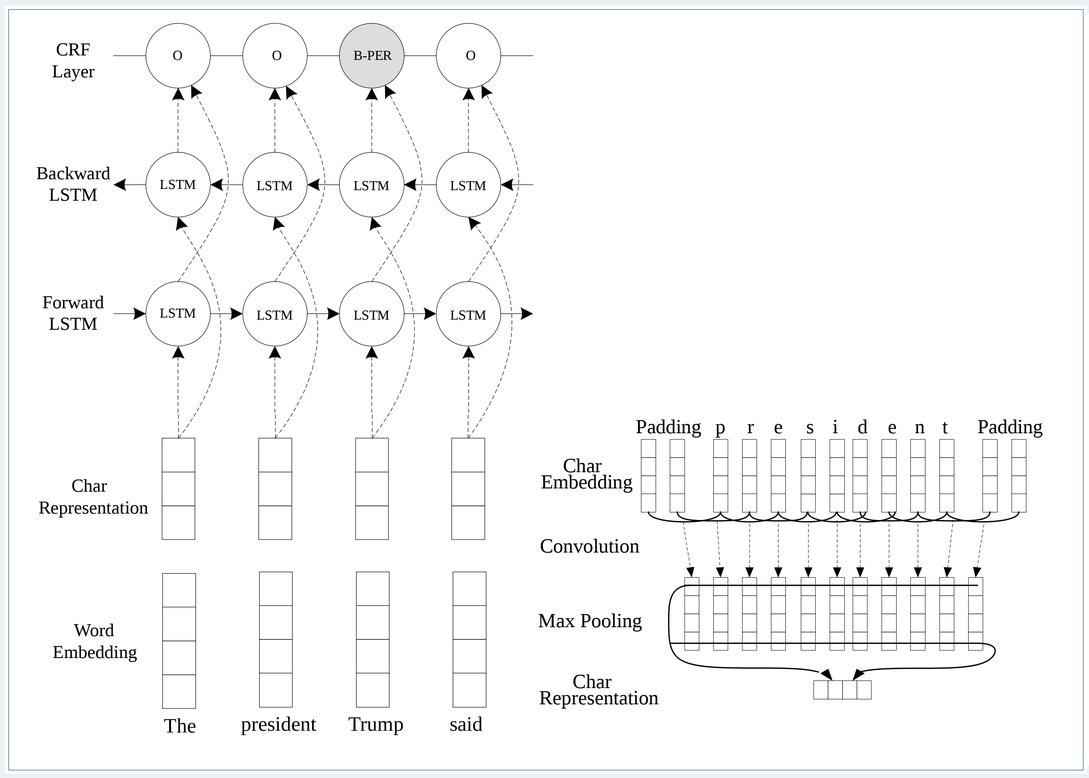
- Named entity recognition performance is directly connected with the amount of data used for training.
- Ensemble methods for named entity recognition.
- Learn a suitable cross-lingual representations for entities.
- Explore multilingual approaches.

Text Analysis Conference (TAC)

- Participation in 2016 Entity Discovery and Linking (EDL) track.
- Planning 2017 participation in the same task.
- NER was the main bottleneck of our TAC 2016's submission.

NER System	Model	Training Data	NER	NERC
Summa (TAC-2016)	Linear	TAC+OntoNotes	83.1	76.1
Summa (2017)	Linear	TAC	81.67	77.61
Summa (2017)	Linear	TAC+OntoNotes	85.68	81.23
SummaNN (2017)	Neural	TAC+OntoNotes	87.87	84.49
USTC (TAC-2016)	Neural	TAC+iFLYTEK	**90.55**	**87.79**

SUMMA NER system performance with different models and datasets.



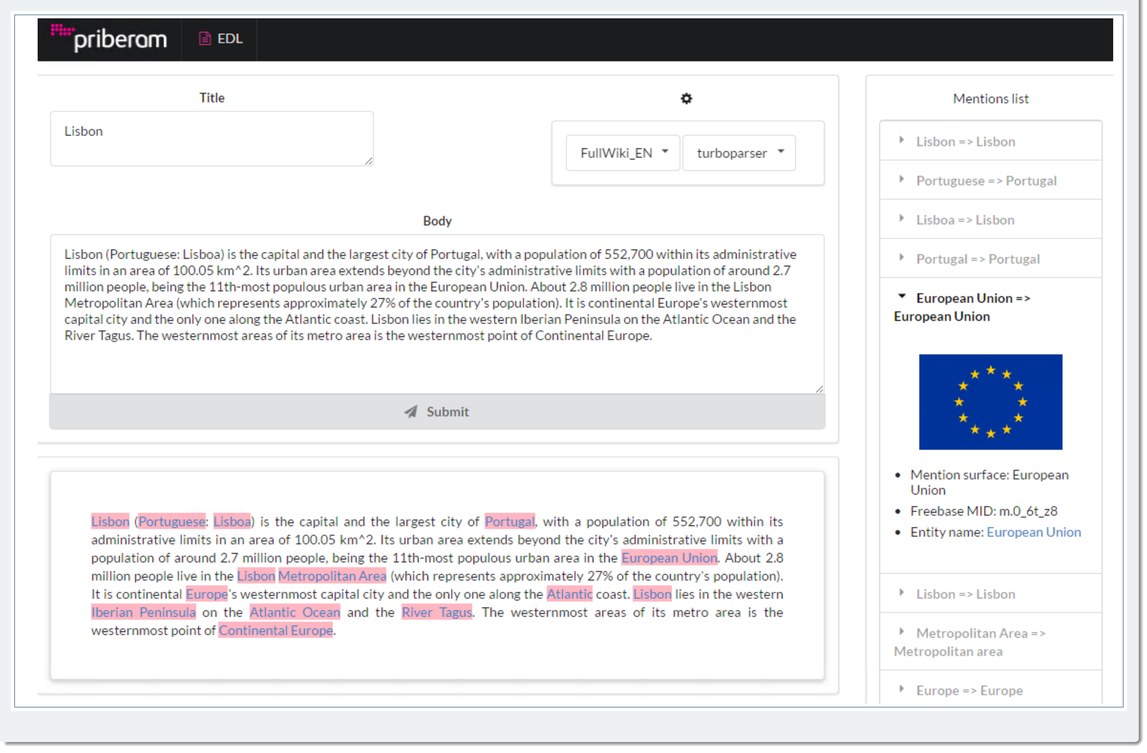
Architecture for SummaNN System.

- Our Entity linking system has state-of-art performance.

Entity Linking System	NERL	KBIDs	CEAFm
Best Team (TAC 2016)	81.1	81.1	83.2
Summa (TAC 2016)	81.8	81.0	83.3
Summa Improved	**83.4**	**81.3**	**86.1**

Entity Linking system performance using entities provided by the USTC system.

Demo



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 688139 (SUMMA).