Comparison of Named Entity Recognition Tools Applied to News Articles

Sergey Vychegzhanin, Evgeny Kotelnikov
Vyatka State University, Kirov, Russia

Named entity recognition task

- Named entity recognition (NER) is a task that seeks to locate and classify named entity mentions in unstructured text into pre-defined categories such as the names of persons, organizations, locations and others
- NER serves as the basis for text summarization, machine translation, topic detection, etc.
- Example:

$$\underbrace{\textbf{Bob Ross}}_{\text{person}} \text{ lived in } \underbrace{\textbf{Florida}}_{\text{location}} \text{ in } \underbrace{\textbf{2006}}_{\text{time}}$$

Goal

- The goal of our research – a comparative study of well-known tools for named entity recognition in relation to news articles
- We evaluated the precision of named entity recognition tools for news articles in the English and the Russian languages
- We evaluated the processing time for tools
- We highlighted the general and distinctive features of the considered tools

Criteria for choosing the NER tools

- Free license
- Existence of desktop version
- Independence from targeting domain
- Ability to recognize basic entity types:
  - person (PER)
  - organization (ORG)
  - location (LOC)
  - time indicators (TIM)
- Support for the English or the Russian languages

Selected NER tools

1. Stanford NER
   (Stanford Natural Language Processing Group)
2. spaCy (Explosion AI)
3. NLTK (University of Pennsylvania)
4. Polyglot (Rami Al-Rfou)
5. GATE (University of Sheffield)
6. Flair (Zalando Research team)
7. DeepPavlov
   (Laboratory of Neural Systems and Deep Learning
   at Moscow Institute of Physics and Technology)

Characteristics of NER Tools

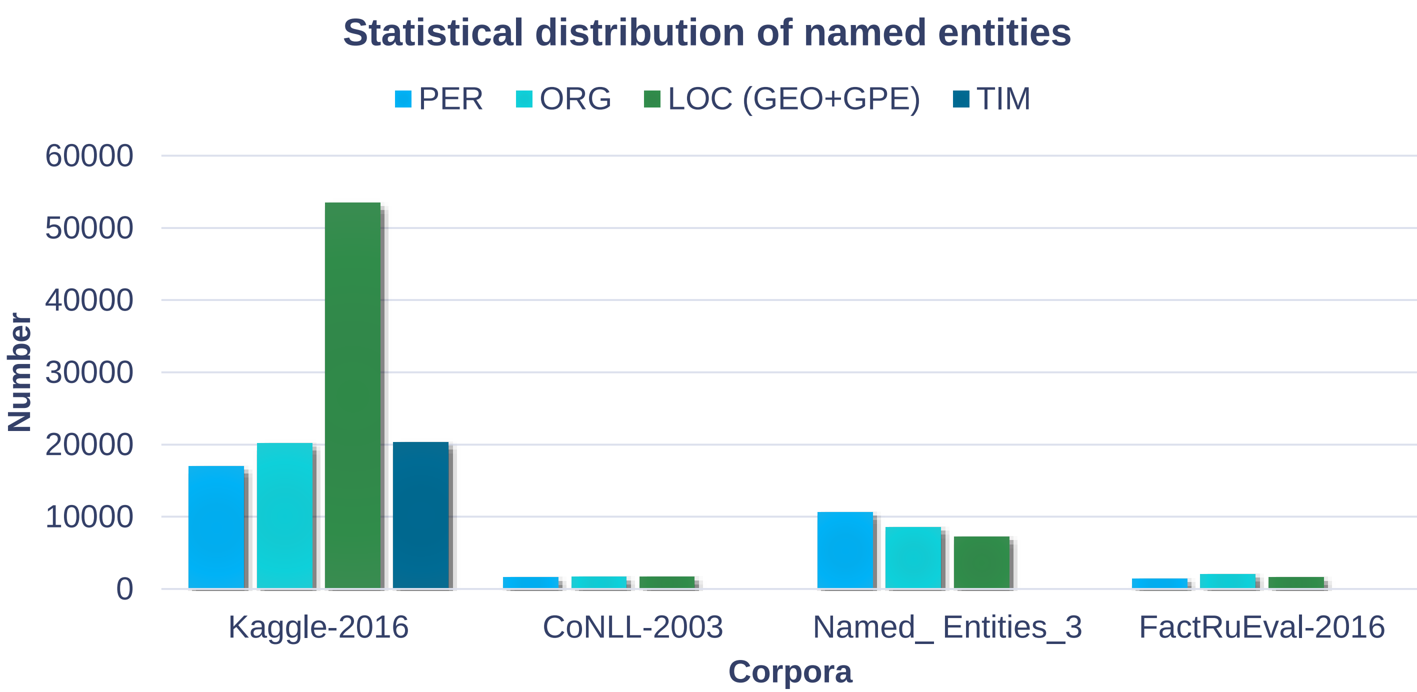
Tool	Programming language	License	Method	Model	Training corpus
Stanford NER	Java	GPL	Conditional Random Field	english_conll_4class	CoNLL-2003
				english_muc_7class	MUC-6, MUC-7
spaCy	Python	MIT	Bloom embeddings and a residual convolutional neural network	en_core_web_sm	OntoNotes
				en_core_web_md	OntoNotes, Common Crawl
				en_core_web_lg	OntoNotes, Common Crawl
				xx_ent_wiki_sm	WikiNER
				ru2	–
NLTK	Python	Apache License v2.0	Maximum Entropy	–	ACE
Polyglot	Python	GPLv3	Feedforward neural network	–	Wikipedia
Flair	Python	MIT	BiLSTM-CRF	–	CoNLL-2003
GATE	Java	LGPL	Finite state machines and rules in the Jape language	–	–
DeepPavlov	Python	Apache License v2.0	BERT	ner_conll2003_bert	CoNLL-2003
				ner_rus_bert	Wikipedia, news data
				ner_ontonotes_bert_mult	OntoNotes

Text corpora

1. Kaggle-2016 – English-language corpus annotated for Named Entity Recognition using GMB (Groningen Meaning Bank) corpus
2. CoNLL-2003 – collection of English-language news articles from the Reuters Corpus used in 2003 at the Conference on Computational Natural Language Learning (CoNLL)
3. Named_Entities_3 – Russian-language corpus based on Person-1000 collection, created by Artificial Intelligence Research Center of the Institute of Program Systems of the Russian Academy of Sciences
4. FactRuEval-2016 – Russian-language corpus which was used in the named entity recognition and fact extraction competition at the conference Dialogue

Characteristics of text corpora

Corpus	Text language	Types of named entities	Number of texts	Average text length, tokens
Kaggle-2016	English	PER, ORG, GEO, GPE, TIM, ART, EVE, NAT	47,959	23
CoNLL-2003	English	PER, ORG, LOC, MISC	1,627	30
Named_Entities_3	Russian	PER, ORG, LOC	1,000	273
FactRuEval-2016	Russian	PER, ORG, LOC	132	463



Evaluation conditions

1. Exact matching of boundaries and types of predicted and true entities
2. Partial matching of the predicted and true entities

Evaluation metrics

- $\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ (Precision)
- $\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ (Recall)
- $\mathrm{F1} = \frac{2 \cdot \mathrm{P} \cdot \mathrm{R}}{\mathrm{P+R}}$ (F1-score)

Entity belongs to class C		**True**	
		Yes	No
Predict	Yes	TP	FP
	No	FN	TN

Results of experiments for English

Tool	Model	F1-score					
		Exact matching			Partial matching		
		Kaggle-2016		CoNLL-2003	Kaggle-2016		CoNLL-2003
		3 types	4 types	3 types	3 types	4 types	3 types
Stanford NER	english_conll_4class	0.554	–	0.860	0.663	–	0.886
	english_muc_7class	0.511	**0.486**	0.611	0.650	0.613	0.683
spaCy	en_core_web_sm	0.483	0.452	0.521	0.649	0.619	0.608
	en_core_web_md	0.503	0.468	0.570	0.665	0.631	0.659
	en_core_web_lg	0.496	0.463	0.597	0.660	0.629	0.697
	xx_ent_wiki_sm	0.501	–	0.597	0.637	–	0.695
NLTK	–	0.476	–	0.467	0.616	–	0.555
Polyglot	–	0.476	–	0.467	0.650	–	0.595
Flair	–	**0.584**	–	**0.887**	**0.691**	–	**0.904**
GATE	–	0.460	0.448	0.528	0.575	0.554	0.598
DeepPavlov	ner_conll2003_bert	0.576	–	0.860	**0.691**	–	0.901
	ner_ontonotes_bert_mult	0.523	0.475	0.687	0.685	**0.637**	0.741

Analysis

1. F1-score under the condition of exact matching is lower than with partial matching the following reasons:
   - The presence of words preceding a named entity:
     – prime minister John Howard
     – the New York Times
     – in November
   - Excessive or incomplete extraction of the entity:
     – Luxembourg-based Court of First Instance
     – the Neolithic period
     – Danilovsky District of Moscow
   - Extracting extra non-entity characters such as ‘.’, ‘-’, ‘(’, etc.

Analysis

2. The most difficult to recognize are the types ORG and TIM, the simplest is the type PER
   - For DeepPavlov on the Kaggle-2016 corpus:
     - $F1_{PER} = 0.813$
     - $F1_{ORG} = 0.540$
     - $F1_{LOC} = 0.703$
     - $F1_{TIM} = 0.493$
3. GATE has the lowest F1-score across all classes, especially in ORG. This tool uses a dictionary in which many organizations are missing
   - $F1_{ORG} = 0.433$ on Kaggle-2016
   - $F1_{ORG} = 0.382$ on CoNLL-2003

Results of experiments for Russian

Tool	Model	F1-score			
		Exact matching		Partial matching	
		Named_ Entities_3	FactRuEval-2016	Named_ Entities_3	FactRuEval-2016
		3 types		3 types	
spaCy	xx_ent_wiki_sm	0.454	0.418	0.681	0.559
	ru2	0.214	0.210	0.361	0.307
Polyglot	–	0.499	0.429	0.674	0.589
GATE	–	0.299	0.268	0.370	0.342
DeepPavlov	ner_rus_bert	**0.945**	**0.622**	**0.973**	**0.752**
	ner_ontonotes_bert_mult	0.688	0.556	0.816	0.679

Named entity recognition processing time

Tool	Model	Processing time, sec	
		Kaggle-2016	Named_Entities_3
Stanford NER	english_conll_4class	76,935	–
	english_muc_7class	75,723	–
spaCy	en_core_web_sm	298	–
	en_core_web_md	319	–
	en_core_web_lg	327	–
	xx_ent_wiki_sm	164	**13**
	ru2	–	136
NLTK	–	475	–
Polyglot	–	107	113
Flair	–	31,711	–
GATE	–	**87**	21
DeepPavlov	ner_conll2003_bert	2,793	–
	ner_ontonotes_bert_mult	2,759	497
	ner_rus_bert	–	465

Conclusion

- We compared the performance of well-known named entity recognition tools: Stanford NER, spaCy, NLTK, Polyglot, Flair, GATE and DeepPavlov
- Flair allowed to get the best performance for the English language and DeepPavlov for the Russian language
- GATE, Polyglot and spaCy turned out to be the fastest tools and Stanford NER – the slowest tool

Thank you for your attention!