# Comparing CRF with spaCy



Andrei Pruteanu
PHD

@andrei\_pruteanu

https://sites.google.com/site/andreipruteanu

#### Overview



**Custom Modeling with spaCy** 

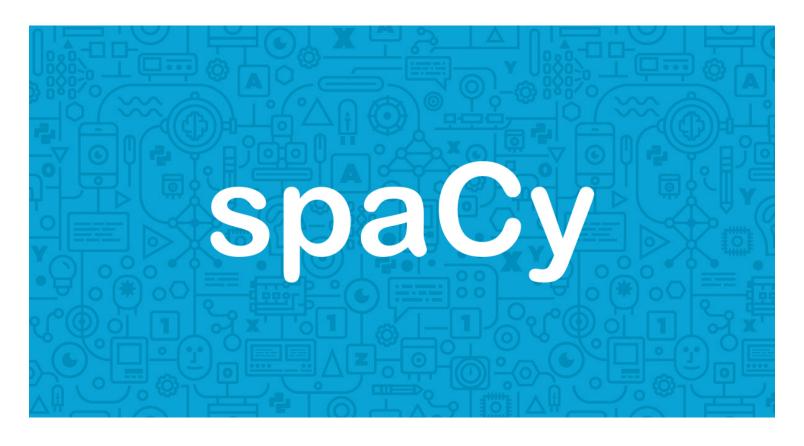
CRF VS spaCy

**Example and Visualization** 



# Custom Modeling with spaCy

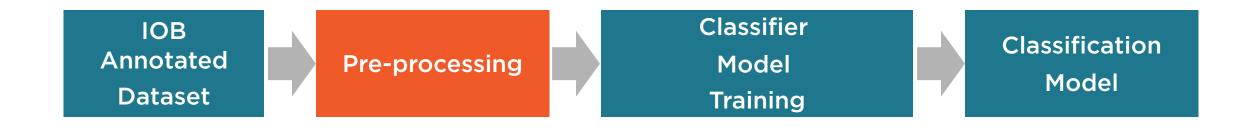




https://spacy.io/

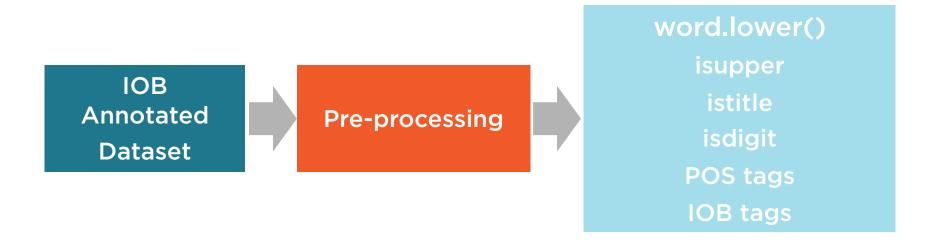


## Preprocessing



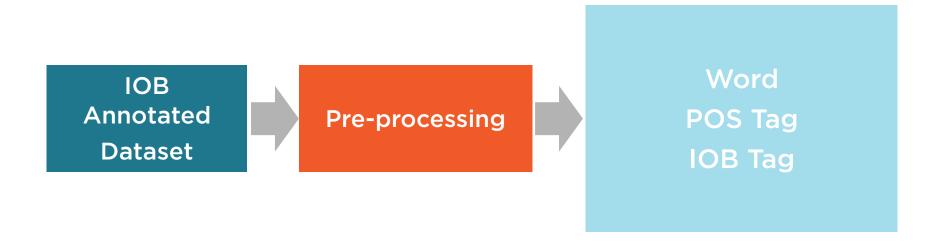


### Preprocessing - CRF





# Preprocessing - spaCy





#### Preprocessing Example

The FAO's estimate includes damage to fishing industries in Indonesia, Maldives, Somalia, Sri Lanka and Thailand.



#### Preprocessing Example

```
Tokens: [{"orth": "The", "tag": "DT", "ner": "O"},
         {"orth": "FAO". "tag": "NNP". "ner": "B-org"}.
         {"orth": "s". "tag": "POS". "ner": "O"}.
         {"orth": "estimate", "tag": "NN", "ner": "O"},
         {"orth": "includes". "tag": "VBZ". "ner": "O"}.
         {"orth": "damage", "tag": "NN", "ner": "O"},
         {"orth": "to". "tag": "TO". "ner": "O"}.
         {"orth": "fishing", "tag": "NN", "ner": "O"}.
         {"orth": "industries". "tag": "NNS". "ner": "O"}.
         {"orth": "in". "tag": "IN". "ner": "O"}.
         {"orth": "Indonesia", "tag": "NNP", "ner": "B-geo"},
         {"orth": ".". "taq": ".". "ner": "O"}.
         {"orth": "Maldives". "tag": "NNP". "ner": "B-geo"}.
         {"orth": ",", "tag": ",", "ner": "O"},
         {"orth": "Somalia". "NNP". "ner": "B-geo"}.
         {"orth": ",", "tag": ",", "ner": "O"},
         {"orth": "Sri", "tag": "NNP", "ner": "B-geo"},
         {"orth": "Lanka". "tag": "NNP". "ner": "I-geo"}.
         {"orth": "and", "tag": "CC", "ner": "O"},
         {"orth": "Thailand". "tag": "NNP". "ner": "B-geo"}.
         {"orth": ".", "tag": ".", "ner": "O"}]
```



## Demo



**Pre-processing** 

**Accuracy Comparison** 



# Custom Modeling Remarks

Custom modeling with spaCy is easy

The library has built-in preprocessing tools

Further tuning is needed for custom spaCy models



# Demo



Example

**Visualization** 



### Summary



Custom modeling with spaCy library

Preprocessing to transform CSV to JSON

Spacy modeling needs further tuning

Visualization helps developing better models

Related courses on PluralSight

Read CRFsuite and spaCy documentation

