

T-NER

An All-Round Python Library for Transformer-based Named Entity Recognition

Asahi Ushio

Jose Camacho-Collados



<https://github.com/asahi417/tner>



<https://pypi.org/project/tner>

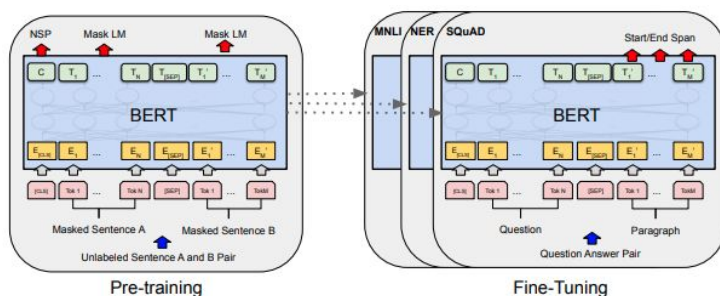
CARDIFF
UNIVERSITY

Cardiff University
School of Computer Science and Informatics

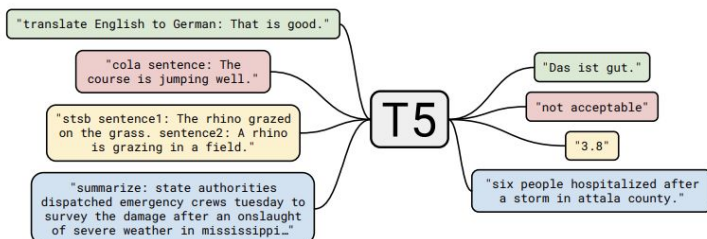
PRIFYSGOL
CAERDYDD

Presented at EACL 2021

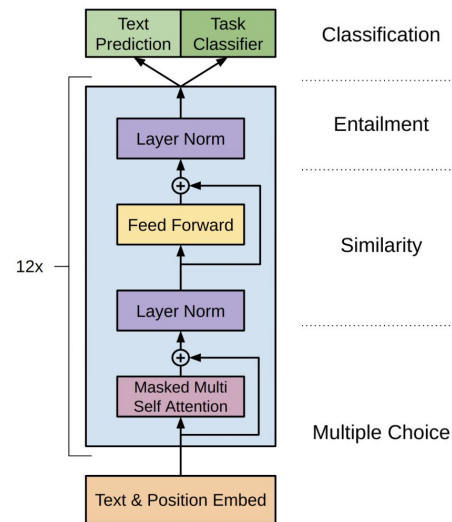
Language Model Pretraining & Finetuning



BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (Devlin, Jacob, et al., 2018)



Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer (Raffel, Colin, et al. 2020)



Improving language understanding by generative pre-training (Radford, Alec, et al., 2018)

Named Entity Recognition

Person

Jacob Collier

is an

Location

English

artist.

Named Entity Recognition

Jacob Collier is an English artist.

Named Entity Recognition

Jacob Collier is an English artist.

Tokenization

Jacob

Collier

is

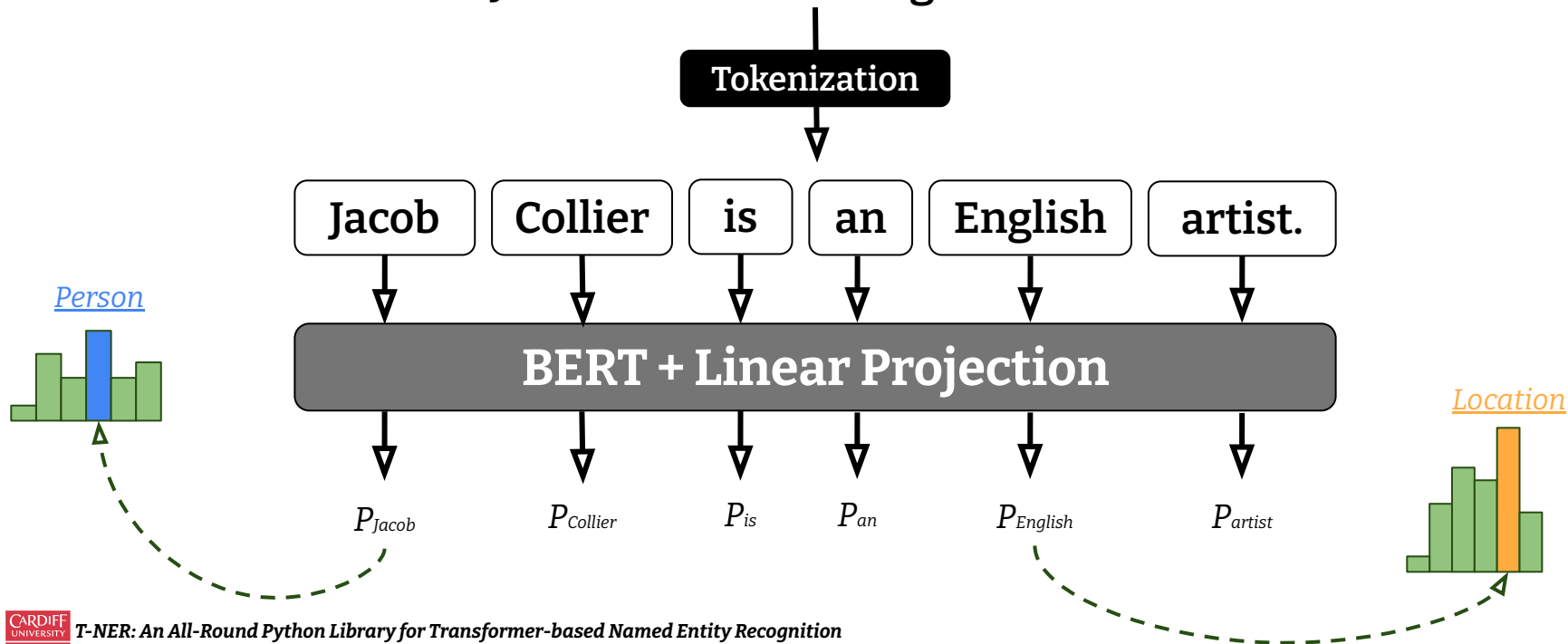
an

English

artist.

Named Entity Recognition

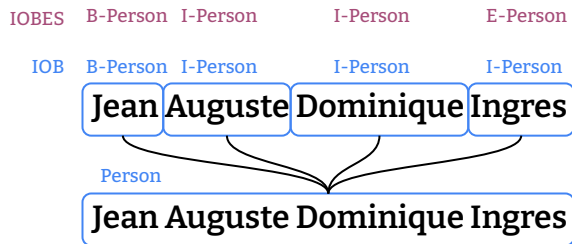
Jacob Collier is an English artist.



Implement NER System

Unify Tagging Scheme

- IOB, IOB2, IOBES, etc



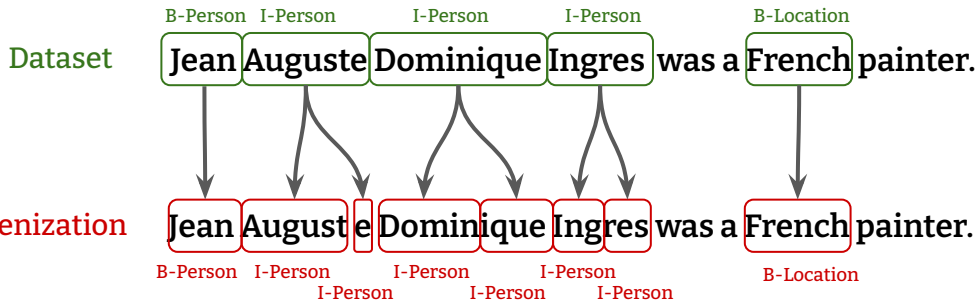
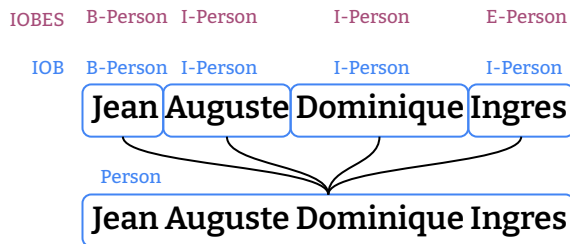
Implement NER System

Unify Tagging Scheme

- IOB, IOB2, IOBES, etc

Fix Sequence Mismatch

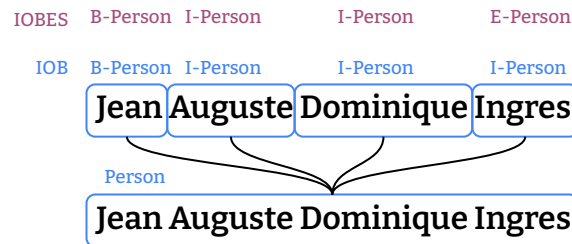
- Align label sequence to model tokenization



Implement NER System

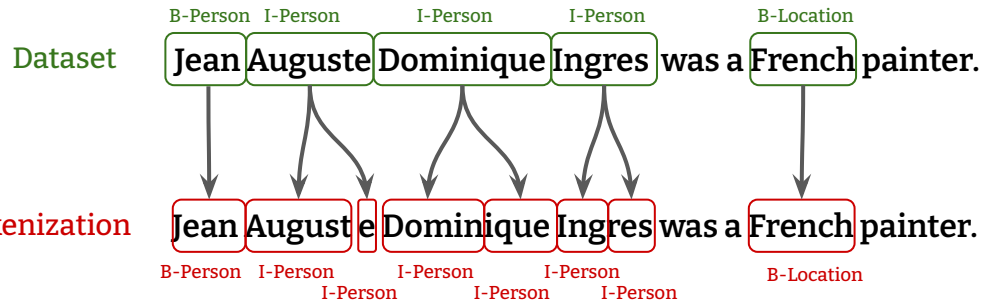
Unify Tagging Scheme

- IOB, IOB2, IOBES, etc



Fix Sequence Mismatch

- Align label sequence to model tokenization



Evaluate in Cross-domain

- Dataset specific entity definition

BioNLP2004

- Protein
- Cell type
- RNA

WNUT2017

- Person
- Corporation
- Creative work

NLP Open Source Softwares

spaCy



 PyTorch



 Transformers

AllenNLP

T-NER

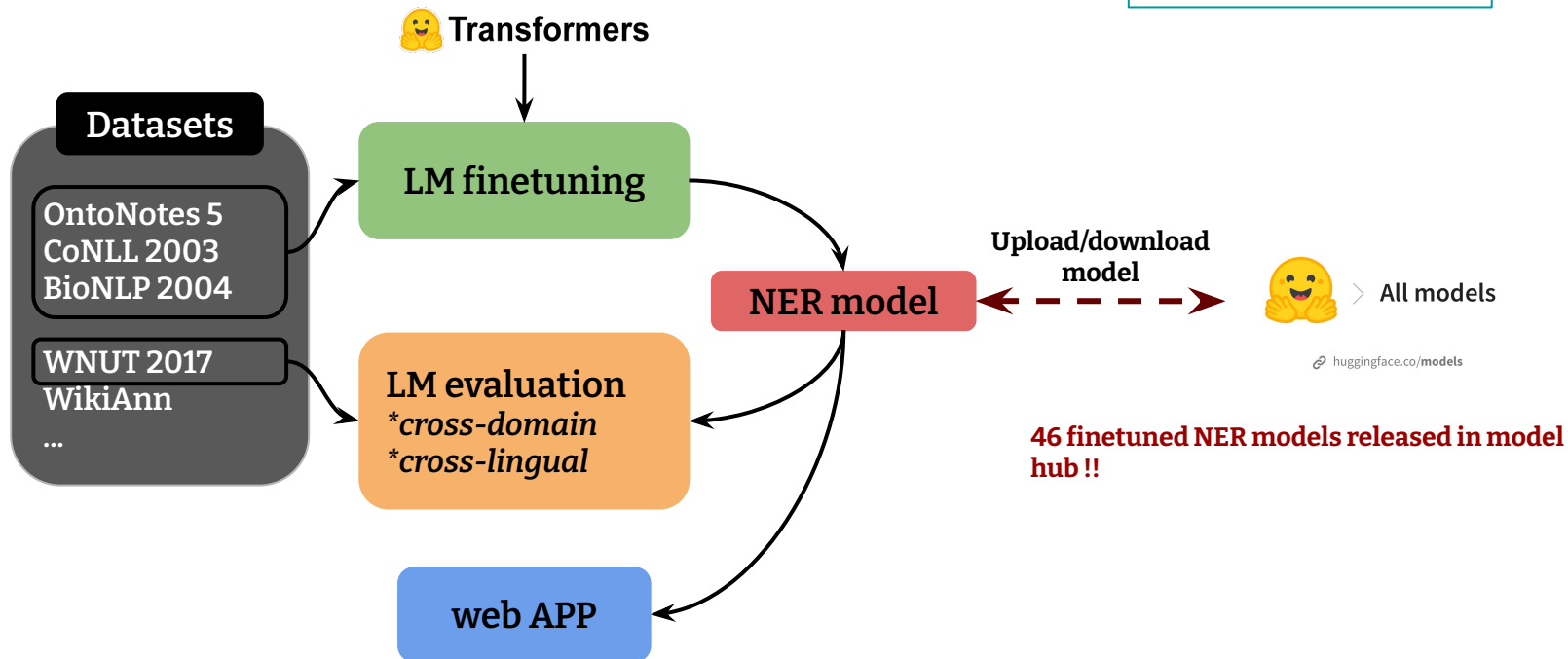


Overall T-NER Design

Notebook link

- [Finetuning](#)
- [Evaluation](#)
- [Model prediction](#)
- [Multilingual NER](#)

- IOB format
- Sequence mismatch fixed



Web Application

```
# SETUP
>>> git clone https://github.com/asahi417/tner
>>> cd tner
>>> pip install .

# RUN APPLICATION at http://0.0.0.0:8000/
>>> export NER_MODEL='asahi417/tner-xlm-roberta-large-ontonotes5'
>>> uvicorn app:app --reload --log-level debug --host 0.0.0.0 --port 8000
```

model checkpoint: asahi417/tner-xlm-roberta-large-ontonotes5

Insert a text to get prediction

スタジオジブリは日本のアニメ制作会社である。

Max sequence length: 128



Run



Result

Input sentence:

スタジオジブリは日本のアニメ制作会社である。

Entities:

- * 1. スタジオジブリ: organization
- * 2. 日本の: group

Experimental Results

train \ test	ontonotes	conll	wnut	wiki	bionlp	bc5cdr	fin	avg
ontonotes	91.6	65.4	53.6	47.5	0.0	0.0	18.3	40.8
conll	62.2	96.0	69.1	61.7	0.0	0.0	22.7	35.1
wnut	41.8	85.7	68.3	54.5	0.0	0.0	20.0	31.7
wiki	32.8	73.3	53.6	93.4	0.0	0.0	12.2	29.6
bionlp	0.0	0.0	0.0	0.0	79.0	0.0	0.0	8.7
bc5cdr	0.0	0.0	0.0	0.0	0.0	88.8	0.0	9.8
fin	48.2	73.2	60.9	58.9	0.0	0.0	82.0	38.1
all	90.9	93.8	60.9	91.3	78.3	84.6	75.5	81.7

train	test					
	en	ja	ru	ko	es	ar
en	84.0	46.3	73.1	58.1	71.4	53.2
ja	53.0	86.5	45.7	57.1	74.5	55.4
ru	60.4	53.3	90.0	68.1	76.8	54.9
ko	57.8	62.0	68.6	89.6	66.2	57.2
es	70.5	50.6	75.8	61.8	92.1	62.1
ar	60.1	55.7	55.7	70.7	79.7	90.3



Thank you!

