

NER using Flair

- [flair](#) | [github](#)

```
In [1]: %pip install datasets  
%pip install evaluate  
%pip install flair
```

Requirement already satisfied: datasets in /opt/conda/lib/python3.10/site-packages (2.16.1)

Requirement already satisfied: filelock in /opt/conda/lib/python3.10/site-packages (from datasets) (3.13.1)

Requirement already satisfied: numpy>=1.17 in /opt/conda/lib/python3.10/site-packages (from datasets) (1.26.2)

Requirement already satisfied: pyarrow>=8.0.0 in /opt/conda/lib/python3.10/site-packages (from datasets) (15.0.0)

Requirement already satisfied: pyarrow-hotfix in /opt/conda/lib/python3.10/site-packages (from datasets) (0.6)

Requirement already satisfied: dill<0.3.8,>=0.3.0 in /opt/conda/lib/python3.10/site-packages (from datasets) (0.3.7)

Requirement already satisfied: pandas in /opt/conda/lib/python3.10/site-packages (from datasets) (2.2.0)

Requirement already satisfied: requests>=2.19.0 in /opt/conda/lib/python3.10/site-packages (from datasets) (2.31.0)

Requirement already satisfied: tqdm>=4.62.1 in /opt/conda/lib/python3.10/site-packages (from datasets) (4.65.0)

Requirement already satisfied: xxhash in /opt/conda/lib/python3.10/site-packages (from datasets) (3.4.1)

Requirement already satisfied: multiprocessing in /opt/conda/lib/python3.10/site-packages (from datasets) (0.70.15)

Requirement already satisfied: fsspec<=2023.10.0,>=2023.1.0 in /opt/conda/lib/python3.10/site-packages (from fsspec[http]<=2023.10.0,>=2023.1.0->datasets) (2023.10.0)

Requirement already satisfied: aiohttp in /opt/conda/lib/python3.10/site-packages (from datasets) (3.9.2)

Requirement already satisfied: huggingface-hub>=0.19.4 in /opt/conda/lib/python3.10/site-packages (from datasets) (0.20.3)

Requirement already satisfied: packaging in /opt/conda/lib/python3.10/site-packages (from datasets) (23.1)

Requirement already satisfied: pyyaml>=5.1 in /opt/conda/lib/python3.10/site-packages (from datasets) (6.0.1)

Requirement already satisfied: aiosignal>=1.1.2 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (1.3.1)

Requirement already satisfied: attrs>=17.3.0 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (23.1.0)

Requirement already satisfied: frozenlist>=1.1.1 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (1.4.1)

Requirement already satisfied: multidict<7.0,>=4.5 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (6.0.4)

Requirement already satisfied: yarl<2.0,>=1.0 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (1.9.4)

Requirement already satisfied: async-timeout<5.0,>=4.0 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets) (4.0.3)

Requirement already satisfied: typing-extensions>=3.7.4.3 in /opt/conda/lib/python3.10/site-packages (from huggingface-hub>=0.19.4->datasets) (4.7.1)

Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->datasets) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->datasets) (3.4)

Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->datasets) (1.26.18)

Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->datasets) (2023.11.17)

Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python

on3.10/site-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.10/site-packages (from pandas->datasets) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.10/site-packages (from pandas->datasets) (2023.4)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.10/site-packages (from python-dateutil>=2.8.2->pandas->datasets) (1.16.0)
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: evaluate in /opt/conda/lib/python3.10/site-packages (0.4.1)
Requirement already satisfied: datasets>=2.0.0 in /opt/conda/lib/python3.10/site-packages (from evaluate) (2.16.1)
Requirement already satisfied: numpy>=1.17 in /opt/conda/lib/python3.10/site-packages (from evaluate) (1.26.2)
Requirement already satisfied: dill in /opt/conda/lib/python3.10/site-packages (from evaluate) (0.3.7)
Requirement already satisfied: pandas in /opt/conda/lib/python3.10/site-packages (from evaluate) (2.2.0)
Requirement already satisfied: requests>=2.19.0 in /opt/conda/lib/python3.10/site-packages (from evaluate) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in /opt/conda/lib/python3.10/site-packages (from evaluate) (4.65.0)
Requirement already satisfied: xxhash in /opt/conda/lib/python3.10/site-packages (from evaluate) (3.4.1)
Requirement already satisfied: multiprocessing in /opt/conda/lib/python3.10/site-packages (from evaluate) (0.70.15)
Requirement already satisfied: fsspec>=2021.05.0 in /opt/conda/lib/python3.10/site-packages (from fsspec[http]>=2021.05.0->evaluate) (2023.10.0)
Requirement already satisfied: huggingface-hub>=0.7.0 in /opt/conda/lib/python3.10/site-packages (from evaluate) (0.20.3)
Requirement already satisfied: packaging in /opt/conda/lib/python3.10/site-packages (from evaluate) (23.1)
Requirement already satisfied: responses<0.19 in /opt/conda/lib/python3.10/site-packages (from evaluate) (0.18.0)
Requirement already satisfied: filelock in /opt/conda/lib/python3.10/site-packages (from datasets>=2.0.0->evaluate) (3.13.1)
Requirement already satisfied: pyarrow>=8.0.0 in /opt/conda/lib/python3.10/site-packages (from datasets>=2.0.0->evaluate) (15.0.0)
Requirement already satisfied: pyarrow-hotfix in /opt/conda/lib/python3.10/site-packages (from datasets>=2.0.0->evaluate) (0.6)
Requirement already satisfied: aiohttp in /opt/conda/lib/python3.10/site-packages (from datasets>=2.0.0->evaluate) (3.9.2)
Requirement already satisfied: pyyaml>=5.1 in /opt/conda/lib/python3.10/site-packages (from datasets>=2.0.0->evaluate) (6.0.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /opt/conda/lib/python3.10/site-packages (from huggingface-hub>=0.7.0->evaluate) (4.7.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->evaluate) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->evaluate) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/conda/lib/python3.10/site-packages (from requests>=2.19.0->evaluate) (1.26.18)
Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.

10/site-packages (from requests>=2.19.0->evaluate) (2023.11.17)
 Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.10/site-packages (from pandas->evaluate) (2.8.2)
 Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.10/site-packages (from pandas->evaluate) (2023.3.post1)
 Requirement already satisfied: tzdata>=2022.7 in /opt/conda/lib/python3.10/site-packages (from pandas->evaluate) (2023.4)
 Requirement already satisfied: aiosignal>=1.1.2 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (1.3.1)
 Requirement already satisfied: attrs>=17.3.0 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (23.1.0)
 Requirement already satisfied: frozenlist>=1.1.1 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (1.4.1)
 Requirement already satisfied: multidict<7.0,>=4.5 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (6.0.4)
 Requirement already satisfied: yarl<2.0,>=1.0 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (1.9.4)
 Requirement already satisfied: async-timeout<5.0,>=4.0 in /opt/conda/lib/python3.10/site-packages (from aiohttp->datasets>=2.0.0->evaluate) (4.0.3)
 Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.10/site-packages (from python-dateutil>=2.8.2->pandas->evaluate) (1.16.0)
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>
 Note: you may need to restart the kernel to use updated packages.
 Requirement already satisfied: flair in /opt/conda/lib/python3.10/site-packages (0.13.1)
 Requirement already satisfied: boto3>=1.20.27 in /opt/conda/lib/python3.10/site-packages (from flair) (1.34.29)
 Requirement already satisfied: bpemb>=0.3.2 in /opt/conda/lib/python3.10/site-packages (from flair) (0.3.4)
 Requirement already satisfied: conllu>=4.0 in /opt/conda/lib/python3.10/site-packages (from flair) (4.5.3)
 Requirement already satisfied: deprecated>=1.2.13 in /opt/conda/lib/python3.10/site-packages (from flair) (1.2.14)
 Requirement already satisfied: ftfy>=6.1.0 in /opt/conda/lib/python3.10/site-packages (from flair) (6.1.3)
 Requirement already satisfied: gdown>=4.4.0 in /opt/conda/lib/python3.10/site-packages (from flair) (5.0.1)
 Requirement already satisfied: gensim>=4.2.0 in /opt/conda/lib/python3.10/site-packages (from flair) (4.3.2)
 Requirement already satisfied: huggingface-hub>=0.10.0 in /opt/conda/lib/python3.10/site-packages (from flair) (0.20.3)
 Requirement already satisfied: janome>=0.4.2 in /opt/conda/lib/python3.10/site-packages (from flair) (0.5.0)
 Requirement already satisfied: langdetect>=1.0.9 in /opt/conda/lib/python3.10/site-packages (from flair) (1.0.9)
 Requirement already satisfied: lxml>=4.8.0 in /opt/conda/lib/python3.10/site-packages (from flair) (5.1.0)
 Requirement already satisfied: matplotlib>=2.2.3 in /opt/conda/lib/python3.10/site-packages (from flair) (3.8.2)
 Requirement already satisfied: more-itertools>=8.13.0 in /opt/conda/lib/python3.10/site-packages (from flair) (10.1.0)
 Requirement already satisfied: mpld3>=0.3 in /opt/conda/lib/python3.10/site-packages (from flair) (0.5.10)
 Requirement already satisfied: pptree>=3.1 in /opt/conda/lib/python3.10/site

```

-packages (from flair) (3.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.10/site-packages (from flair) (2.8.2)
Requirement already satisfied: pytorch-revgrad>=0.2.0 in /opt/conda/lib/python3.10/site-packages (from flair) (0.2.0)
Requirement already satisfied: regex>=2022.1.18 in /opt/conda/lib/python3.10/site-packages (from flair) (2023.12.25)
Requirement already satisfied: scikit-learn>=1.0.2 in /opt/conda/lib/python3.10/site-packages (from flair) (1.4.0)
Requirement already satisfied: segtok>=1.5.11 in /opt/conda/lib/python3.10/site-packages (from flair) (1.5.11)
Requirement already satisfied: sqlitedict>=2.0.0 in /opt/conda/lib/python3.10/site-packages (from flair) (2.1.0)
Requirement already satisfied: tabulate>=0.8.10 in /opt/conda/lib/python3.10/site-packages (from flair) (0.9.0)
Requirement already satisfied: torch!=1.8,>=1.5.0 in /opt/conda/lib/python3.10/site-packages (from flair) (2.1.2)
Requirement already satisfied: tqdm>=4.63.0 in /opt/conda/lib/python3.10/site-packages (from flair) (4.65.0)
Requirement already satisfied: transformer-smaller-training-vocab>=0.2.3 in /opt/conda/lib/python3.10/site-packages (from flair) (0.3.3)
Requirement already satisfied: transformers<5.0.0,>=4.18.0 in /opt/conda/lib/python3.10/site-packages (from transformers[sentencepiece]<5.0.0,>=4.18.0->flair) (4.37.1)
Requirement already satisfied: urllib3<2.0.0,>=1.0.0 in /opt/conda/lib/python3.10/site-packages (from flair) (1.26.18)
Requirement already satisfied: wikipedia-api>=0.5.7 in /opt/conda/lib/python3.10/site-packages (from flair) (0.6.0)
Requirement already satisfied: semver<4.0.0,>=3.0.0 in /opt/conda/lib/python3.10/site-packages (from flair) (3.0.2)
Requirement already satisfied: boto3<1.35.0,>=1.34.29 in /opt/conda/lib/python3.10/site-packages (from boto3>=1.20.27->flair) (1.34.29)
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /opt/conda/lib/python3.10/site-packages (from boto3>=1.20.27->flair) (1.0.1)
Requirement already satisfied: s3transfer<0.11.0,>=0.10.0 in /opt/conda/lib/python3.10/site-packages (from boto3>=1.20.27->flair) (0.10.0)
Requirement already satisfied: numpy in /opt/conda/lib/python3.10/site-packages (from bpemb>=0.3.2->flair) (1.26.2)
Requirement already satisfied: requests in /opt/conda/lib/python3.10/site-packages (from bpemb>=0.3.2->flair) (2.31.0)
Requirement already satisfied: sentencepiece in /opt/conda/lib/python3.10/site-packages (from bpemb>=0.3.2->flair) (0.1.99)
Requirement already satisfied: wrapt<2,>=1.10 in /opt/conda/lib/python3.10/site-packages (from deprecated>=1.2.13->flair) (1.16.0)
Requirement already satisfied: wcwidth<0.3.0,>=0.2.12 in /opt/conda/lib/python3.10/site-packages (from ftfy>=6.1.0->flair) (0.2.13)
Requirement already satisfied: beautifulsoup4 in /opt/conda/lib/python3.10/site-packages (from gdown>=4.4.0->flair) (4.12.2)
Requirement already satisfied: filelock in /opt/conda/lib/python3.10/site-packages (from gdown>=4.4.0->flair) (3.13.1)
Requirement already satisfied: scipy>=1.7.0 in /opt/conda/lib/python3.10/site-packages (from gensim>=4.2.0->flair) (1.12.0)
Requirement already satisfied: smart-open>=1.8.1 in /opt/conda/lib/python3.10/site-packages (from gensim>=4.2.0->flair) (6.4.0)
Requirement already satisfied: fsspec>=2023.5.0 in /opt/conda/lib/python3.10/site-packages (from huggingface-hub>=0.10.0->flair) (2023.10.0)

```

Requirement already satisfied: pyyaml<=5.1 in /opt/conda/lib/python3.10/site-packages (from huggingface-hub<=0.10.0->flair) (6.0.1)

Requirement already satisfied: typing-extensions<=3.7.4.3 in /opt/conda/lib/python3.10/site-packages (from huggingface-hub<=0.10.0->flair) (4.7.1)

Requirement already satisfied: packaging<=20.9 in /opt/conda/lib/python3.10/site-packages (from huggingface-hub<=0.10.0->flair) (23.1)

Requirement already satisfied: six in /opt/conda/lib/python3.10/site-packages (from langdetect<=1.0.9->flair) (1.16.0)

Requirement already satisfied: contourpy<=1.0.1 in /opt/conda/lib/python3.10/site-packages (from matplotlib<=2.2.3->flair) (1.2.0)

Requirement already satisfied: cycler<=0.10 in /opt/conda/lib/python3.10/site-packages (from matplotlib<=2.2.3->flair) (0.12.1)

Requirement already satisfied: fonttools<=4.22.0 in /opt/conda/lib/python3.10/site-packages (from matplotlib<=2.2.3->flair) (4.47.2)

Requirement already satisfied: kiwisolver<=1.3.1 in /opt/conda/lib/python3.10/site-packages (from matplotlib<=2.2.3->flair) (1.4.5)

Requirement already satisfied: pillow<=8 in /opt/conda/lib/python3.10/site-packages (from matplotlib<=2.2.3->flair) (10.0.1)

Requirement already satisfied: pyparsing<=2.3.1 in /opt/conda/lib/python3.10/site-packages (from matplotlib<=2.2.3->flair) (3.1.1)

Requirement already satisfied: jinja2 in /opt/conda/lib/python3.10/site-packages (from mpld3<=0.3->flair) (3.1.2)

Requirement already satisfied: joblib<=1.2.0 in /opt/conda/lib/python3.10/site-packages (from scikit-learn<=1.0.2->flair) (1.3.2)

Requirement already satisfied: threadpoolctl<=2.0.0 in /opt/conda/lib/python3.10/site-packages (from scikit-learn<=1.0.2->flair) (3.2.0)

Requirement already satisfied: sympy in /opt/conda/lib/python3.10/site-packages (from torch!=1.8,>=1.5.0->flair) (1.12)

Requirement already satisfied: networkx in /opt/conda/lib/python3.10/site-packages (from torch!=1.8,>=1.5.0->flair) (3.1)

Requirement already satisfied: tokenizers<0.19,>=0.14 in /opt/conda/lib/python3.10/site-packages (from transformers<5.0.0,>=4.18.0->transformers[sentencepiece]<5.0.0,>=4.18.0->flair) (0.15.1)

Requirement already satisfied: safetensors<=0.3.1 in /opt/conda/lib/python3.10/site-packages (from transformers<5.0.0,>=4.18.0->transformers[sentencepiece]<5.0.0,>=4.18.0->flair) (0.4.2)

Requirement already satisfied: protobuf in /opt/conda/lib/python3.10/site-packages (from transformers[sentencepiece]<5.0.0,>=4.18.0->flair) (4.25.2)

Requirement already satisfied: accelerate<=0.21.0 in /opt/conda/lib/python3.10/site-packages (from transformers[torch]<5.0,>=4.1->transformer-smaller-training-vocab<=0.2.3->flair) (0.26.1)

Requirement already satisfied: soupsieve>1.2 in /opt/conda/lib/python3.10/site-packages (from beautifulsoup4->gdown<=4.4.0->flair) (2.5)

Requirement already satisfied: MarkupSafe<=2.0 in /opt/conda/lib/python3.10/site-packages (from jinja2->mpld3<=0.3->flair) (2.1.1)

Requirement already satisfied: charset-normalizer<4,>=2 in /opt/conda/lib/python3.10/site-packages (from requests->bpemb<=0.3.2->flair) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-packages (from requests->bpemb<=0.3.2->flair) (3.4)

Requirement already satisfied: certifi>=2017.4.17 in /opt/conda/lib/python3.10/site-packages (from requests->bpemb<=0.3.2->flair) (2023.11.17)

Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /opt/conda/lib/python3.10/site-packages (from requests[socks]->gdown<=4.4.0->flair) (1.7.1)

Requirement already satisfied: mpmath<=0.19 in /opt/conda/lib/python3.10/site-packages (from sympy->torch!=1.8,>=1.5.0->flair) (1.3.0)

Requirement already satisfied: psutil in /opt/conda/lib/python3.10/site-pack

ages (from accelerate>=0.21.0->transformers[torch]<5.0,>=4.1->transformer-smaller-training-vocab>=0.2.3->flair) (5.9.0)

WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>

Note: you may need to restart the kernel to use updated packages.

In [2]: `from utils import *`

In [3]: `from flair.data import Sentence
from flair.nn import Classifier
from flair.models import SequenceTagger`

In [4]: `# load the 4-class NER tagger
tagger = SequenceTagger.load("flair/ner-english")
flair_tagger = Classifier.load('ner')`

2024-01-29 03:21:14,202 SequenceTagger predicts: Dictionary with 20 tags: <unk>, 0, S-ORG, S-MISC, B-PER, E-PER, S-LOC, B-ORG, E-ORG, I-PER, S-PER, B-MISC, I-MISC, E-MISC, I-ORG, B-LOC, E-LOC, I-LOC, <START>, <STOP>

In [18]: `def predict_ner_flair(flair_tagger: SequenceTagger, labeled_dataset: Dataset
 """ Run inference on the tokens using trained BERT model """
 references: list[list[str]] = []
 fl_predictions: list[list[str]] = []

 for row in tqdm(labeled_dataset, desc=str(len(labeled_dataset))):
 # add ground truth labels to references
 references.append([re.sub("^[BI]-", "", tag_names[id]) for id in row
 tokens = row["tokens"]
 # make a flair sentence object
 sentence = Sentence(" ".join(tokens))
 # run NER over sentence
 flair_tagger.predict(sentence)
 ner_results = ["0" for _ in range(len(tokens))]
 for entity in sentence.get_spans('ner'):
 tag = entity.tag
 for token in entity:
 idx = token.idx - 1
 # print(f"{entity.text} {idx} / {len(tokens)}")
 try:
 ner_results[idx] = tag
 except IndexError as err:
 break
 # translate numerical index to NER class label
 predicted_tags = ner_results
 fl_predictions.append(predicted_tags)
 return references, fl_predictions`

In [19]: `references, fl_predictions = predict_ner_flair(flair_tagger, test)`

3453: 100%|██████████| 3453/3453 [00:25<00:00, 137.96it/s]

In [20]: `results = evaluate_results(references, fl_predictions)
results`

```
Out[20]: {'ER': {'precision': 0.9126637554585153,
               'recall': 0.9047619047619048,
               'f1': 0.9086956521739129,
               'number': 1617},
          'ISC': {'precision': 0.7702503681885126,
                 'recall': 0.7546897546897547,
                 'f1': 0.762390670553936,
                 'number': 693},
          'OC': {'precision': 0.865681233933162,
                 'recall': 0.8104693140794224,
                 'f1': 0.8371659415786201,
                 'number': 1662},
          'overall_precision': 0.868421052631579,
          'overall_recall': 0.8391238670694864,
          'overall_f1': 0.8535211267605634,
          'overall_accuracy': 0.9658231937116399}
```

```
In [21]: # Save NER results to disk
         flair_results_path = os.path.join(interim_dir, "ner_results_flair.json")

         save_ner_results(flair_results_path, references, fl_predictions)

         # Load persisted NER results
         # references, fl_predictions = load_ner_results(flair_results_path)
```

Saving NER results to ../data/interim/ner_results_flair.json

```
In [22]: flair_evaluation_path = os.path.join(interim_dir, "evaluation_results_flair.json")
         save_evaluation_results(flair_evaluation_path, results)
```

Saving evaluation results to ../data/interim/evaluation_results_flair.json