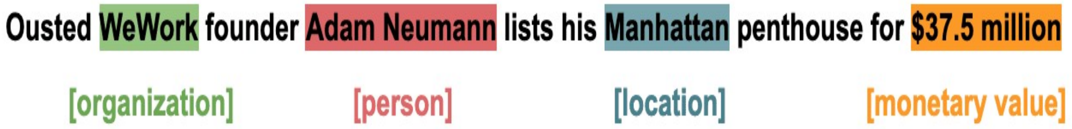
Name Entity Recognition (NER)

SAMPLE HUGGINGFACE MODELS WITH MULTINERD-DATA IN ENGLISH.



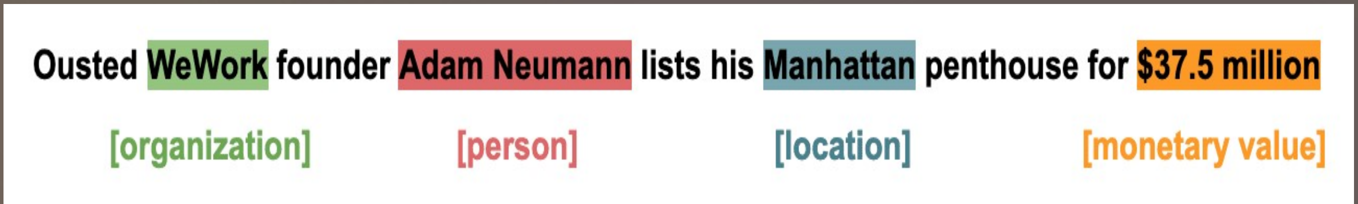
Ousted WeWork founder Adam Neumann lists his Manhattan penthouse for $37.5 million

[organization] [person] [location] [monetary value]

Image source: https://monkeylearn.com/blog/named-entity-recognition/

Name Entity Recognition (NER)

Named entity recognition (NER) – also called entity identification or entity extraction – is a natural language processing (NLP) technique that automatically identifies named entities in a text and classifies them into predefined categories. Entities can be names of people, organizations, locations, times, quantities, monetary values, percentages, and more.

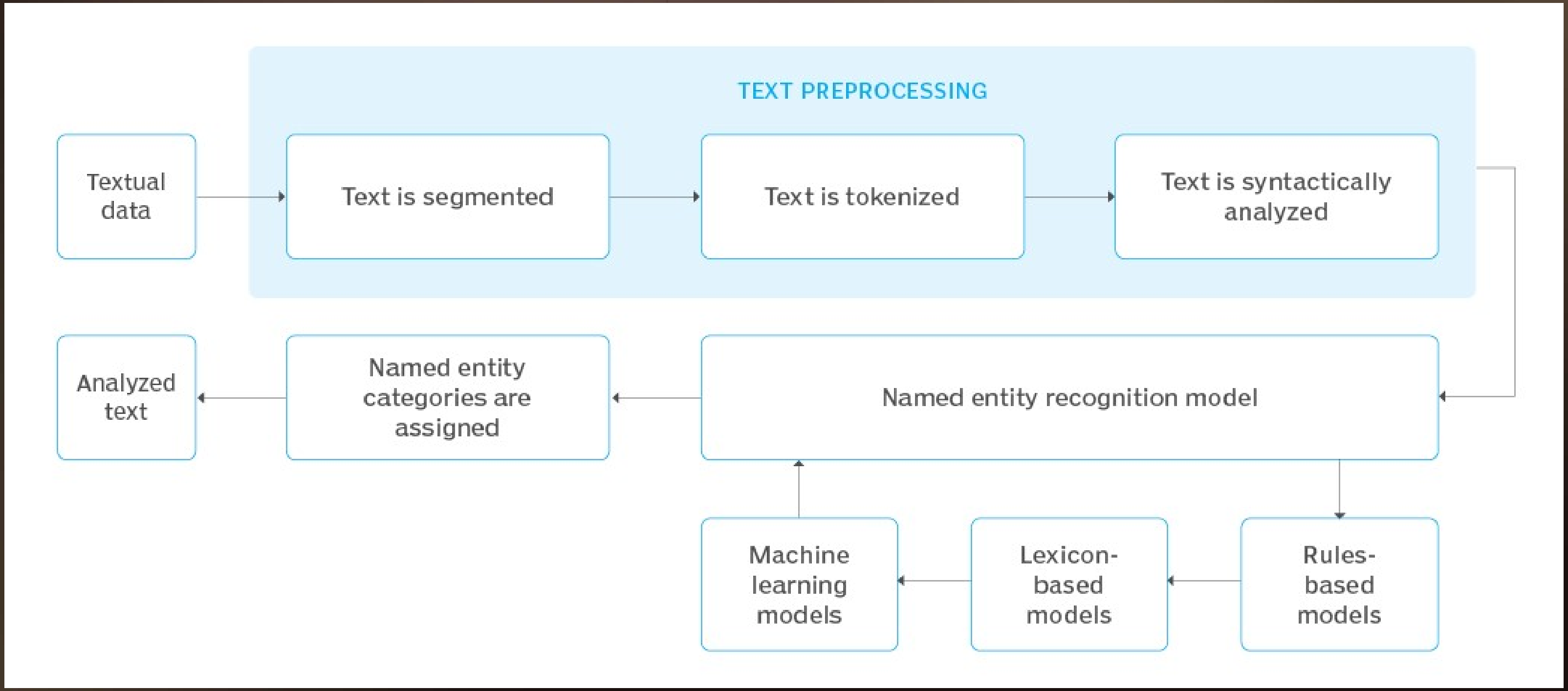


Ousted WeWork founder Adam Neumann lists his Manhattan penthouse for $37.5 million

[organization] [person] [location] [monetary value]

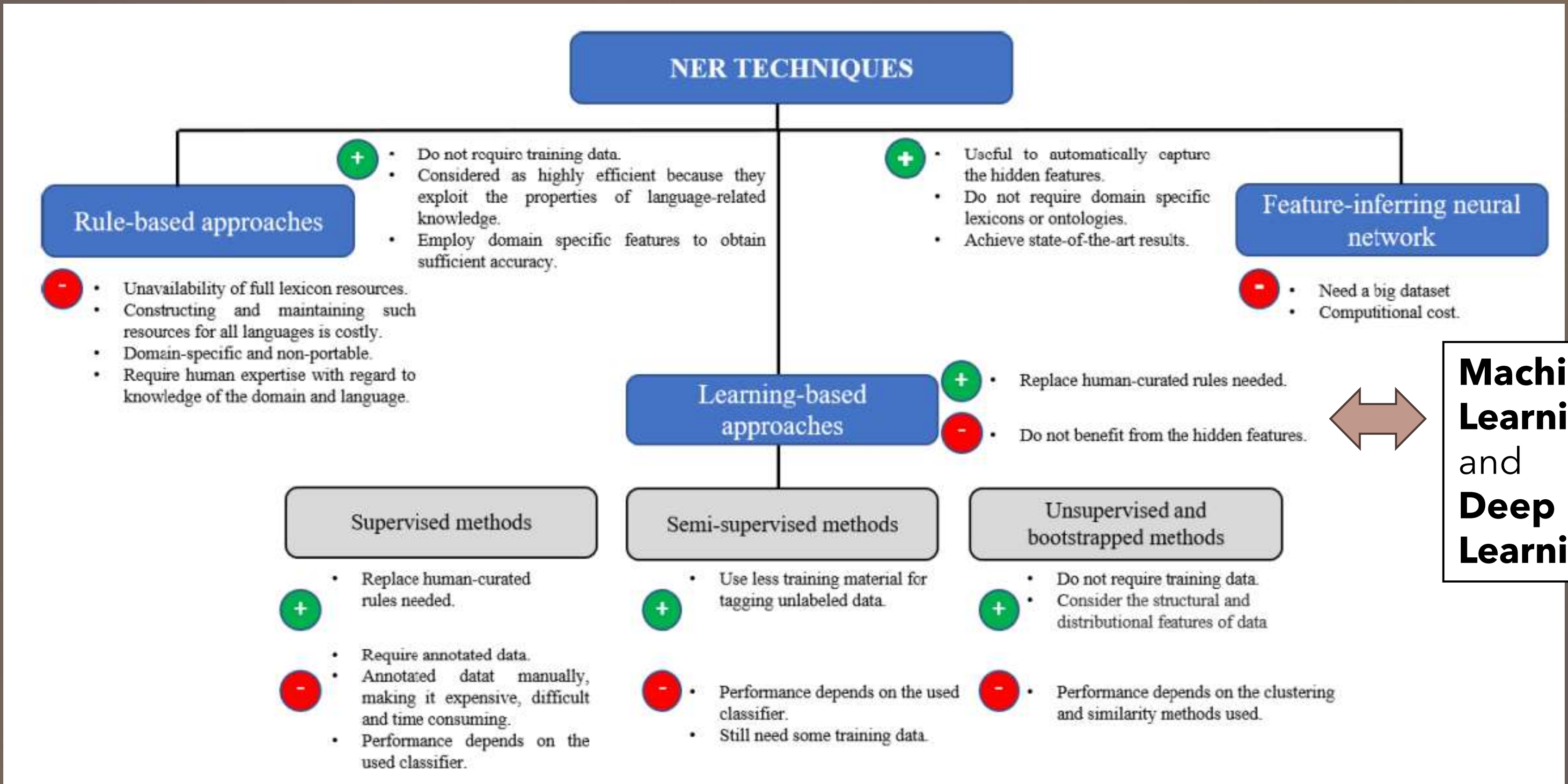
Name Entity Recognition (NER)

-- Process --



Source: https://www.techtarget.com/whatis/definition/named-entity-recognition-NER

Current Methods



NER TECHNIQUES

Rule-based approaches

+
- Do not require training data.
- Considered as highly efficient because they exploit the properties of language-related knowledge.
- Employ domain specific features to obtain sufficient accuracy.

−
- Unavailability of full lexicon resources.
- Constructing and maintaining such resources for all languages is costly.
- Domain-specific and non-portable.
- Require human expertise with regard to knowledge of the domain and language.

Feature-inferring neural network

+
- Useful to automatically capture the hidden features.
- Do not require domain specific lexicons or ontologies.
- Achieve state-of-the-art results.

−
- Need a big dataset
- Computitional cost.

Learning-based approaches

+
- Replace human-curated rules needed.

−
- Do not benefit from the hidden features.

Supervised methods

+
- Replace human-curated rules needed.

−
- Require annotated data.
- Annotated datat manually, making it expensive, difficult and time consuming.
- Performance depends on the used classifier.

Semi-supervised methods

+
- Use less training material for tagging unlabeled data.

−
- Performance depends on the used classifier.
- Still need some training data.

Unsupervised and bootstrapped methods

+
- Do not require training data.
- Consider the structural and distributional features of data

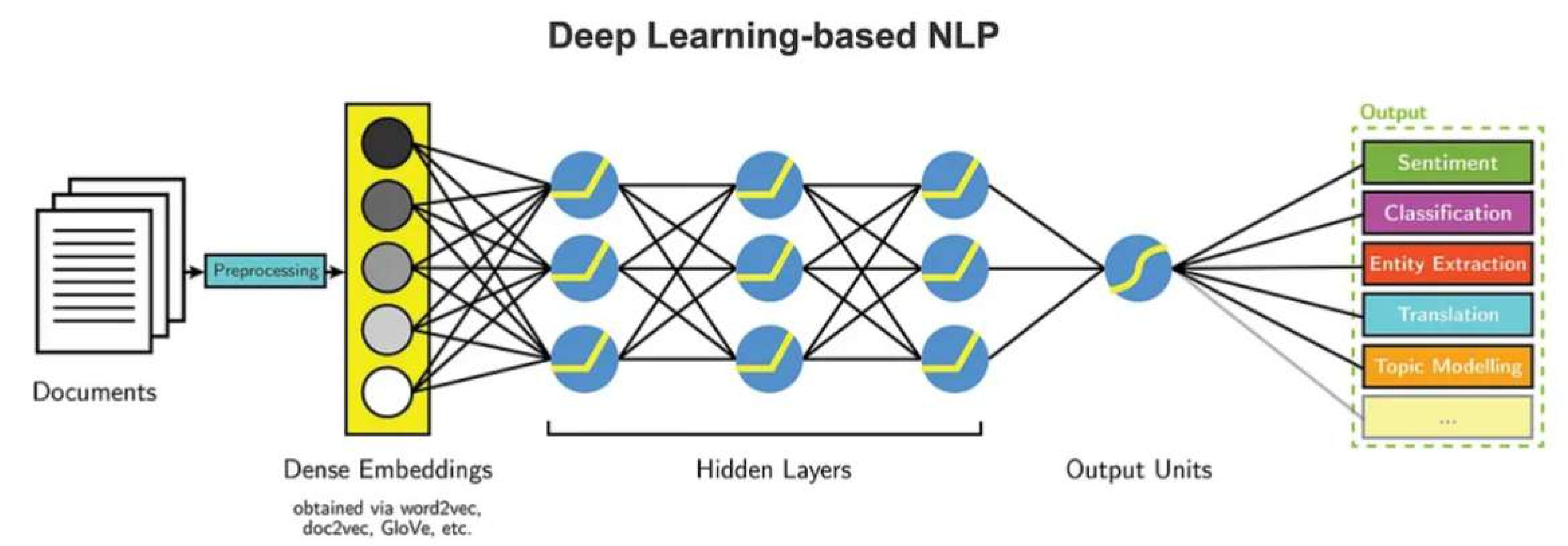
−
- Performance depends on the clustering and similarity methods used.

⇔

Machine Learning and **Deep Learning**

Source: https://ieeexplore.ieee.org/ielx7/6287639/8948470/08999622.pdf

Current Methods
-- Deep Learning --



Source: https://medium.com/intro-to-artificial-intelligence/entity-extraction-using-deep-learning-8014acac6bb8

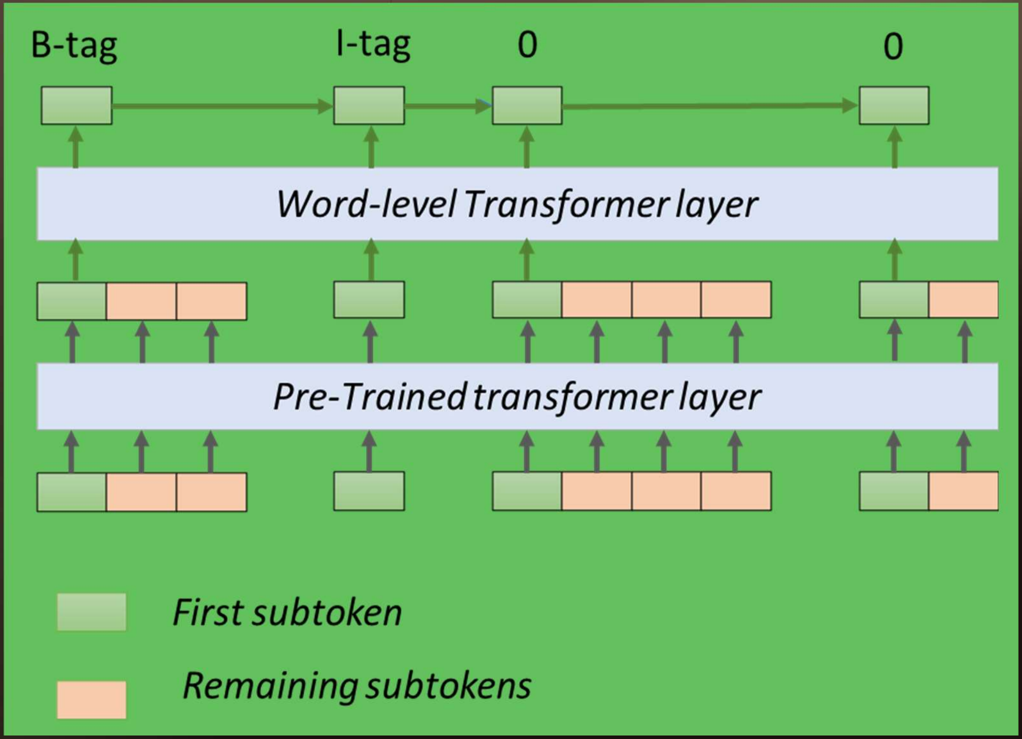
Name Entity Recognition (NER)

-- Deep Learning --

- There are many pre-trained models for this task. For example, you can see the previous paper here (https://aclanthology.org/2020.coling-main.334.pdf, https://pahulpreet86.github.io/name-entity-recognition-pre-trained-models-review/ and https://pahulpreet86.github.io/name-entity-recognition-pre-trained-models-review/).
- In this work, I used the Huggingface pre-trained model, namely Distillbert. This is because it is a small model with six deep learning layers and is suitable to run on my machine. You can try with other models such as Roberta and Bert-base. You can find it with the NER task here.

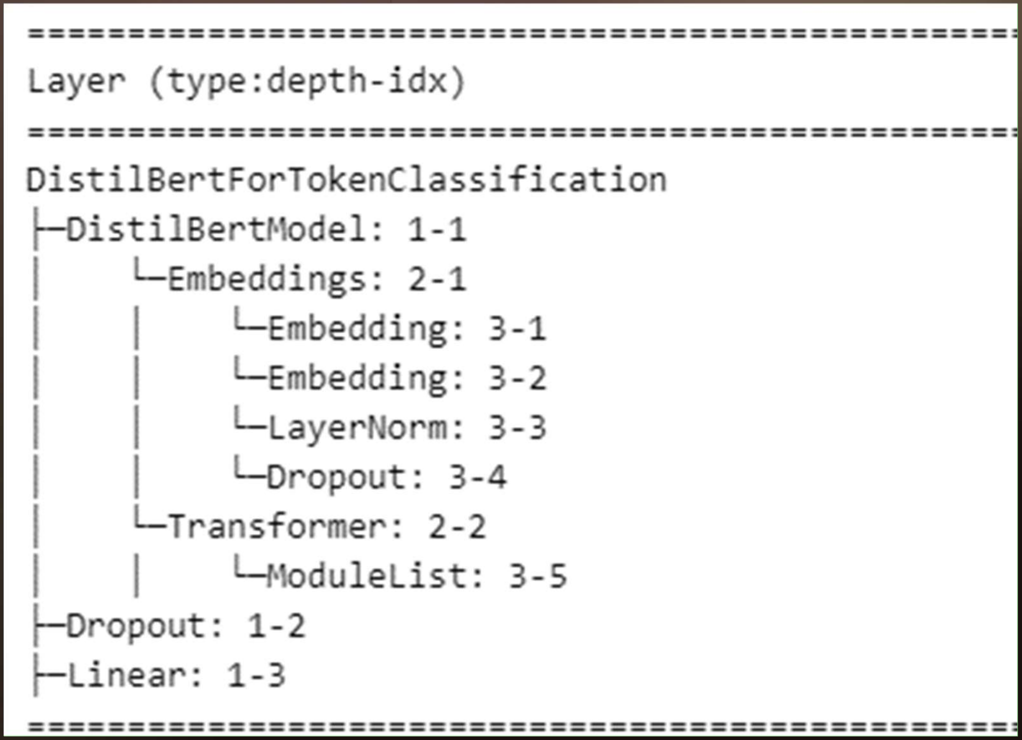
Name Entity Recognition (NER)

-- HuggingFace--



Name Entity Recognition (NER)

-- Distillbert Pre-trained Model --



```
==================================================
Layer (type:depth-idx)
==================================================
DistilBertForTokenClassification
├─DistilBertModel: 1-1
│    └─Embeddings: 2-1
│    │    └─Embedding: 3-1
│    │    └─Embedding: 3-2
│    │    └─LayerNorm: 3-3
│    │    └─Dropout: 3-4
│    └─Transformer: 2-2
│    │    └─ModuleList: 3-5
├─Dropout: 1-2
├─Linear: 1-3
==================================================
```

This Repository works

- Scenario A: Select a model from huggingface and run the MultiNERD data for English. Fine-tune the parameters with suitable metrics to get the best results.
- Scenario B: Similar to scenario A. However, the entity types belong to one of the following five entity types: PERSON(PER), ORGANIZATION(ORG), LOCATION(LOC), DISEASES(DIS), ANIMAL(ANIM).

This Repository works

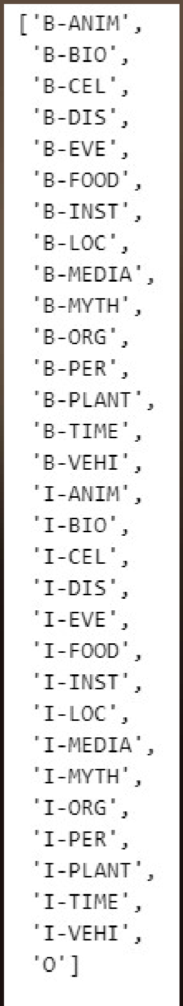
- You can find out the results of SystemA at SystemADistilbert.ipynb
- Testing the SystemA with an example sentence (SystemADistilbert_Testing.ipynb).
- You can find out the results of SystemB at SystemBDistilbert.ipynb
- Testing the SystemB with an example sentence (SystemBDistilbert_Testing.ipynb).

This Repository works

-- Pre-processing data --

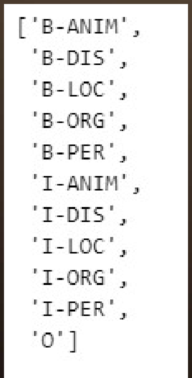
- I only loaded the data and followed the instructions from Huggingface.
- I selected small samples for all training, validation and test sets in two systems because of the limitation of my Machine. You can see it in the codes.
- For SystemB, I only kept the entity types which belong to one of the following five entity types: PERSON(PER), ORGANIZATION(ORG), LOCATION(LOC), DISEASES(DIS), ANIMAL(ANIM).

A



```
['B-ANIM',
 'B-BIO',
 'B-CEL',
 'B-DIS',
 'B-EVE',
 'B-FOOD',
 'B-INST',
 'B-LOC',
 'B-MEDIA',
 'B-MYTH',
 'B-ORG',
 'B-PER',
 'B-PLANT',
 'B-TIME',
 'B-VEHI',
 'I-ANIM',
 'I-BIO',
 'I-CEL',
 'I-DIS',
 'I-EVE',
 'I-FOOD',
 'I-INST',
 'I-LOC',
 'I-MEDIA',
 'I-MYTH',
 'I-ORG',
 'I-PER',
 'I-PLANT',
 'I-TIME',
 'I-VEHI',
 'O']
```

B

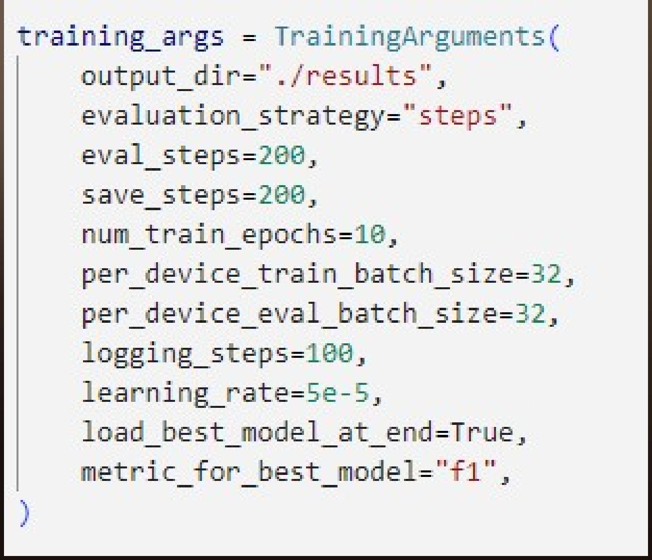


```
['B-ANIM',
 'B-DIS',
 'B-LOC',
 'B-ORG',
 'B-PER',
 'I-ANIM',
 'I-DIS',
 'I-LOC',
 'I-ORG',
 'I-PER',
 'O']
```

This Repository works

-- Pre-processing data --

- For two scenarios, I used the same configuration to compare the outcomes of the two systems.
    - Metrics: F1 score
    - Learning rate: `5e-5 = 0.00005`
    - Epochs: 10

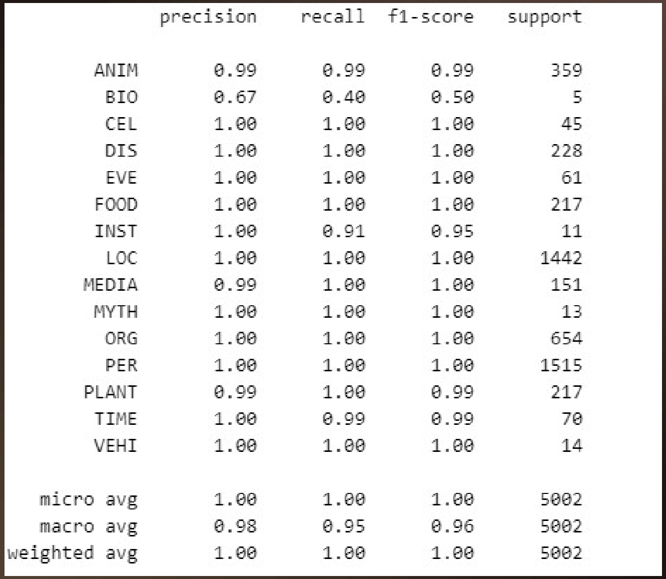


```
training_args = TrainingArguments(
    output_dir="./results",
    evaluation_strategy="steps",
    eval_steps=200,
    save_steps=200,
    num_train_epochs=10,
    per_device_train_batch_size=32,
    per_device_eval_batch_size=32,
    logging_steps=100,
    learning_rate=5e-5,
    load_best_model_at_end=True,
    metric_for_best_model="f1",
)
```

This Repository works

-- Results -- SystemA

Training set

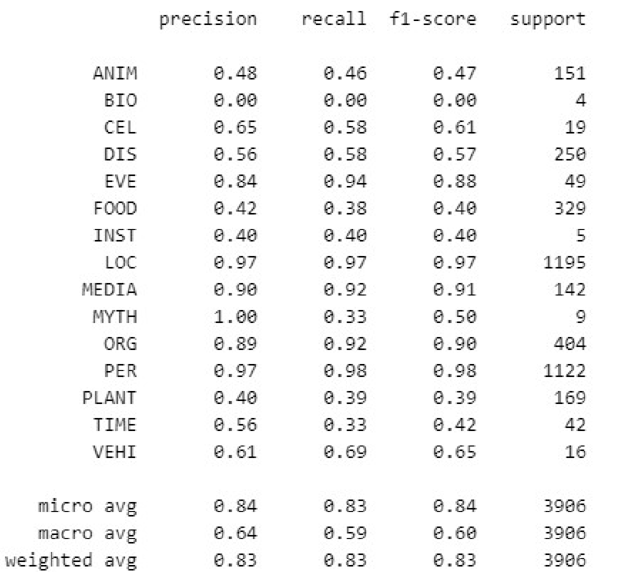


```
              precision    recall  f1-score   support

        ANIM       0.99      0.99      0.99       359
         BIO       0.67      0.40      0.50         5
         CEL       1.00      1.00      1.00        45
         DIS       1.00      1.00      1.00       228
         EVE       1.00      1.00      1.00        61
        FOOD       1.00      1.00      1.00       217
        INST       1.00      0.91      0.95        11
         LOC       1.00      1.00      1.00      1442
       MEDIA       0.99      1.00      1.00       151
        MYTH       1.00      1.00      1.00        13
         ORG       1.00      1.00      1.00       654
         PER       1.00      1.00      1.00      1515
       PLANT       0.99      1.00      0.99       217
        TIME       1.00      0.99      0.99        70
        VEHI       1.00      1.00      1.00        14

   micro avg       1.00      1.00      1.00      5002
   macro avg       0.98      0.95      0.96      5002
weighted avg       1.00      1.00      1.00      5002
```

Validation set

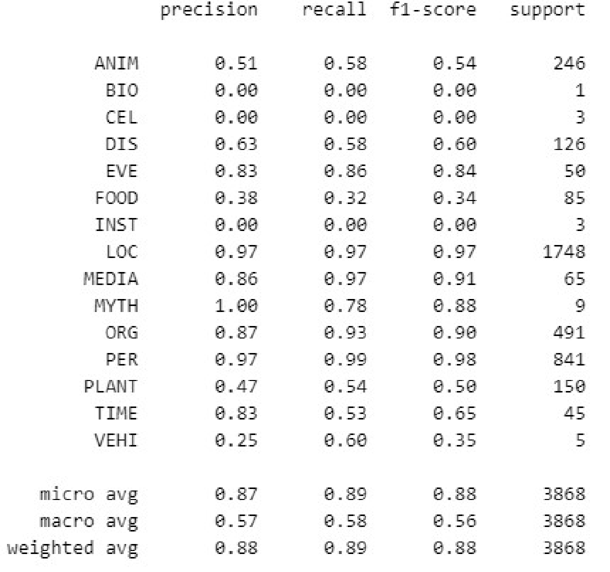


```
              precision    recall  f1-score   support

        ANIM       0.48      0.46      0.47       151
         BIO       0.00      0.00      0.00         4
         CEL       0.65      0.58      0.61        19
         DIS       0.56      0.58      0.57       250
         EVE       0.84      0.94      0.88        49
        FOOD       0.42      0.38      0.40       329
        INST       0.40      0.40      0.40         5
         LOC       0.97      0.97      0.97      1195
       MEDIA       0.90      0.92      0.91       142
        MYTH       1.00      0.33      0.50         9
         ORG       0.89      0.92      0.90       404
         PER       0.97      0.98      0.98      1122
       PLANT       0.40      0.39      0.39       169
        TIME       0.56      0.33      0.42        42
        VEHI       0.61      0.69      0.65        16

   micro avg       0.84      0.83      0.84      3906
   macro avg       0.64      0.59      0.60      3906
weighted avg       0.83      0.83      0.83      3906
```

Test set



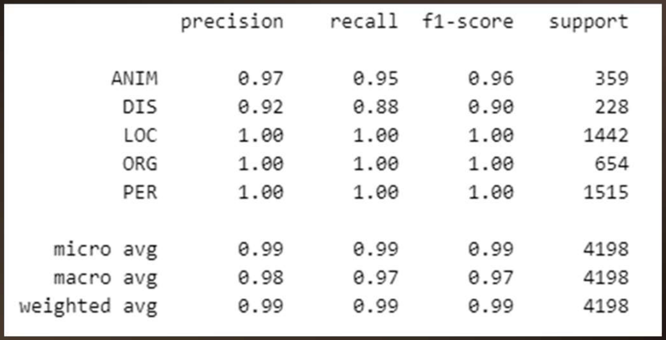
```
              precision    recall  f1-score   support

        ANIM       0.51      0.58      0.54       246
         BIO       0.00      0.00      0.00         1
         CEL       0.00      0.00      0.00         3
         DIS       0.63      0.58      0.60       126
         EVE       0.83      0.86      0.84        50
        FOOD       0.38      0.32      0.34        85
        INST       0.00      0.00      0.00         3
         LOC       0.97      0.97      0.97      1748
       MEDIA       0.86      0.97      0.91        65
        MYTH       1.00      0.78      0.88         9
         ORG       0.87      0.93      0.90       491
         PER       0.97      0.99      0.98       841
       PLANT       0.47      0.54      0.50       150
        TIME       0.83      0.53      0.65        45
        VEHI       0.25      0.60      0.35         5

   micro avg       0.87      0.89      0.88      3868
   macro avg       0.57      0.58      0.56      3868
weighted avg       0.88      0.89      0.88      3868
```

This Repository works
-- Results -- SystemB

Training set

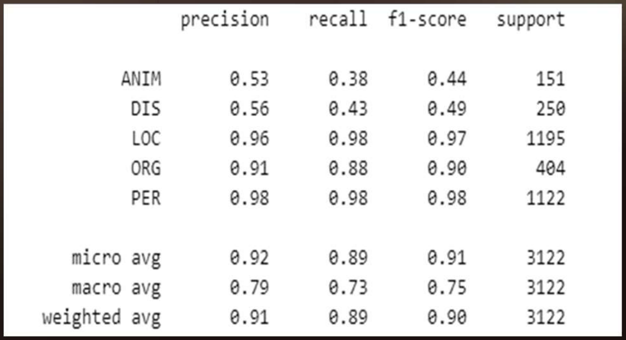


```
              precision    recall  f1-score   support

        ANIM       0.97      0.95      0.96       359
         DIS       0.92      0.88      0.90       228
         LOC       1.00      1.00      1.00      1442
         ORG       1.00      1.00      1.00       654
         PER       1.00      1.00      1.00      1515

   micro avg       0.99      0.99      0.99      4198
   macro avg       0.98      0.97      0.97      4198
weighted avg       0.99      0.99      0.99      4198
```

Validation set

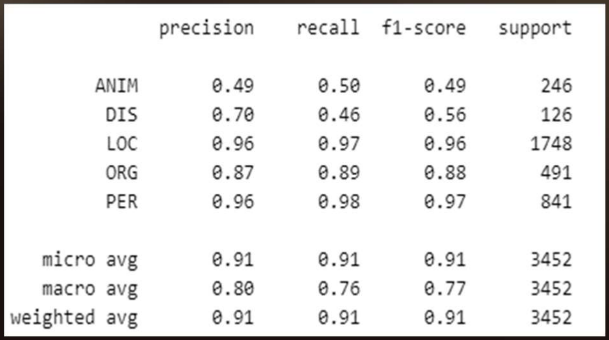


```
              precision    recall  f1-score   support

        ANIM       0.53      0.38      0.44       151
         DIS       0.56      0.43      0.49       250
         LOC       0.96      0.98      0.97      1195
         ORG       0.91      0.88      0.90       404
         PER       0.98      0.98      0.98      1122

   micro avg       0.92      0.89      0.91      3122
   macro avg       0.79      0.73      0.75      3122
weighted avg       0.91      0.89      0.90      3122
```

Test set



```
              precision    recall  f1-score   support

        ANIM       0.49      0.50      0.49       246
         DIS       0.70      0.46      0.56       126
         LOC       0.96      0.97      0.96      1748
         ORG       0.87      0.89      0.88       491
         PER       0.96      0.98      0.97       841

   micro avg       0.91      0.91      0.91      3452
   macro avg       0.80      0.76      0.77      3452
weighted avg       0.91      0.91      0.91      3452
```

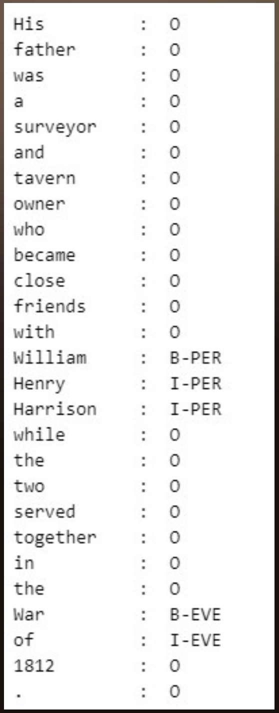
This Repository works

-- Load saved models and run sample sentence --

- Testing the SystemA with an example sentence (SystemADistilbert_Testing.ipynb).

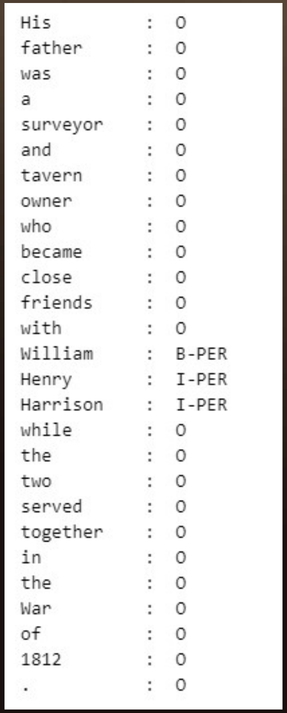
- Testing the SystemB with an example sentence (SystemBDistilbert_Testing.ipynb).

A



```
His         :  O
father      :  O
was         :  O
a           :  O
surveyor    :  O
and         :  O
tavern      :  O
owner       :  O
who         :  O
became      :  O
close       :  O
friends     :  O
with        :  O
William     :  B-PER
Henry       :  I-PER
Harrison    :  I-PER
while       :  O
the         :  O
two         :  O
served      :  O
together    :  O
in          :  O
the         :  O
War         :  B-EVE
of          :  I-EVE
1812        :  O
.           :  O
```

B



```
His         :  O
father      :  O
was         :  O
a           :  O
surveyor    :  O
and         :  O
tavern      :  O
owner       :  O
who         :  O
became      :  O
close       :  O
friends     :  O
with        :  O
William     :  B-PER
Henry       :  I-PER
Harrison    :  I-PER
while       :  O
the         :  O
two         :  O
served      :  O
together    :  O
in          :  O
the         :  O
War         :  O
of          :  O
1812        :  O
.           :  O
```

This Repository works

-- Findings -

- The testing set is missed a label ('I-BIO' - there is only 30 labels) compared to training set (31) and validation set (31).
- Time for running:
  - SystemA: 5619.6813 s ~ 1.56 hour. (31 Labels)
  - SystemB: 5557,2894 ~ 1.5436915.

  → The times for training of 2 systems are not far different. This is because there are changes in the last layer for the output layer with Transformers. Other layers are trained in similar ways. However, it can be different with larger datasets because I only test these systems on small random data from the original data because of time limitations and my machine.

This Repository works
-- Findings -

- Accuaracy (Precision | recall | f1 score)
  - System A: 1.0 | 0.83 | 0.88
  - System B: 0.99 | 0.90 | 0.91

  → System B outperformed System A on average score. This is because of fewer labels compared to system A.
- Limitations:
  - Randomly small parts of the data can lead to incorrect evaluation of the models. Therefore, It needs to run with a powerful computer for all datasets.
  - Because of time limitations, I only tested on the default learning rate (example) from HuggingFace. It can change the learning rate with different optimizers like Adam. (See more here)
  - Epoch = 10. It needs to be tested with more epochs. Note that I run 1 epoch with a full training set. It took more than 8 hours. According to the system, you can consider how many epochs you need for training. GPU is another solution to reduce the time for training.

Thanks