

NER Annotation Tool Kullanım Kılavuzu

NER Nedir?

NER (Named Entity Recognition) veya Türkçe karşılığıyla **Adlandırılmış Varlık Tanıma**, doğal dil işleme (NLP) alanında kullanılan ve metin içindeki kişiler, kurumlar, yerler, tarihler, sayılar gibi belirli ve önemli varlıkları otomatik olarak tanıma ve sınıflandırma amacıyla geliştirilmiş bir makine öğrenimi modelidir.

NER modelleri, genellikle derin öğrenme teknikleri kullanılarak eğitilmektedir. Bu modeller, büyük miktardaki etiketlenmiş metin verileri üzerinde çalışarak varlıkları tanıma ve sınıflandırma konusunda uzmanlaşırlar. Eğitim sürecinde model, metin içindeki kelimelerin sırası, gramer yapısı ve bağlam gibi bilgileri kullanarak varlıkları belirlemeyi öğrenir.

Örnek: "Elon Musk known as Tesla's. Company's Headquarters is in California."

Bu cümlede bir NER modeli, "Elon Musk" ve "California" gibi adlandırılmış varlıkları tanıyarak sırasıyla "kişi" ve "yer" olarak sınıflandıracaktır.

```
import spacy
from spacy import displacy

# Spacy'nin İngilizce modelini yükle
nlp = spacy.load("en_core_web_sm")

# Örnek metin
text = "Elon Musk Tesla'nın CEO'su olarak biliniyor. Şirketin m"
```

```
# Metni işleme
doc = nlp(text)

# NER görselleştirme
displacy.render(doc, style="ent", jupyter=True)
```

Elon Musk PERSON known as Tesla ORG 's. Company's Headquarters is in California GPE .

Sonuç olarak, NER modelleri, doğal dil işleme alanında büyük öneme sahip olup, metin verilerinden bilgi çıkarma ve analiz etme süreçlerini otomatikleştirerek birçok farklı alanda kullanılmaktadır.

NER Annotation Tool Neden Kullanılır?

NER modeller etiketli veriler ile eğitilir. NER model için etiketli veri hazırlanırken text verisi içinde etiketlenecek karakterlerin başlangıç ve bitiş karakterlerini tek tek belirtmek gerekir.

Örnek:

```
DATASET = [
    ("G4-3 Report the name of the organization.. 43", {"entities":
    ("G4-4 Report the primary brands, products, and services. 20
]
```

Bu işlem manual olarak yapıldığında oldukça uğraştırıcıdır ve saatler alabilmektedir. Bu nedenle NER Annotation isimli etiketleme araçları üretilmiştir.

NER Annotation Tool Nasıl Kullanılır?

Kullanılacak NER Annotation Tool: <https://tecoholic.github.io/ner-annotator/>

NER Annotation Tool Github Repo: <https://github.com/tecoholic/ner-annotator>

Adım 1: Dosya Yükle

- Etiketlenecek txt dosyasını **Text file** yazan yere tıklayarak yükle
- Etiketlemeye başlamak için **BEGIN!** tuşuna bas.

NER Text Annotator

Annotate text for spaCy NER Model training

Select your file(s), then press 'Begin'!

BEGIN!

Hint: You can also drag and drop files into this window!



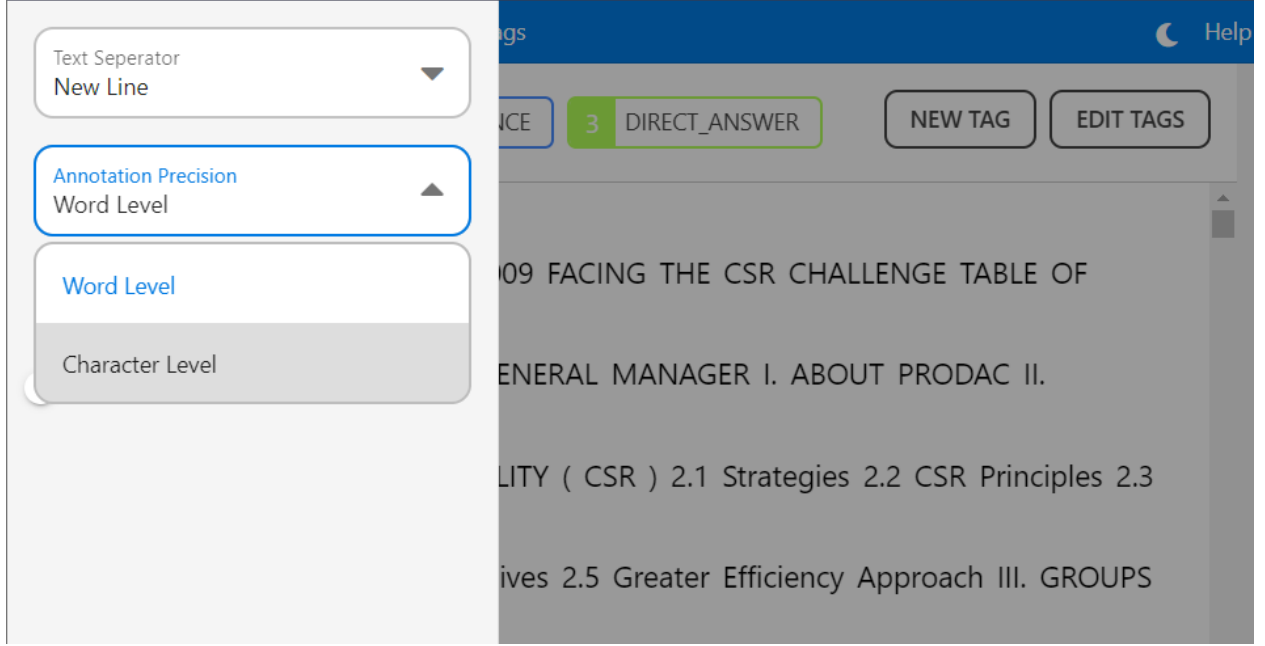
Text file



(Optional) Annotations

Adım 2: Etiketlemeye Başla

- Açılan ekranda **Character Level** seç, diğer ayarlar default haliyle kalabilir.

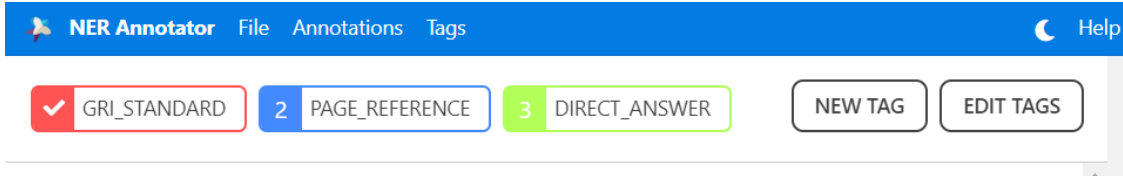


- NEW TAG butonuna tıkla. Aşağıdaki etiket isimlerini sırasıyla yaz. Her etiket ismini yazdıktan sonra + tuşuna basak etiketi ekle.

etiket isimleri: GRI_STANDARD , PAGE_REFERENCE , DIRECT_ANSWER



Tüm etiketleri oluşturduğunuzda aşağıdaki gibi gözükmeli.



- Artık etiketlemeye başlayabilirsiniz.
 - yukarıdaki kutucuklardan hangisi için etiketleme yapmak istediğini seç (mouse ile bir kez üzerine tıkla)



- text metin içinde etiketlemek istediğin yeri bul (mouse ile üzerine gel ve mouse a basılı tutarak etiketlenecek yerin başından sonuna mouse u kaydır. etiketlenecek yerin sonuna geldiğinde mouse tan elini kaldır.)

Creating alliances 3.7 Suppliers 60 61 STA

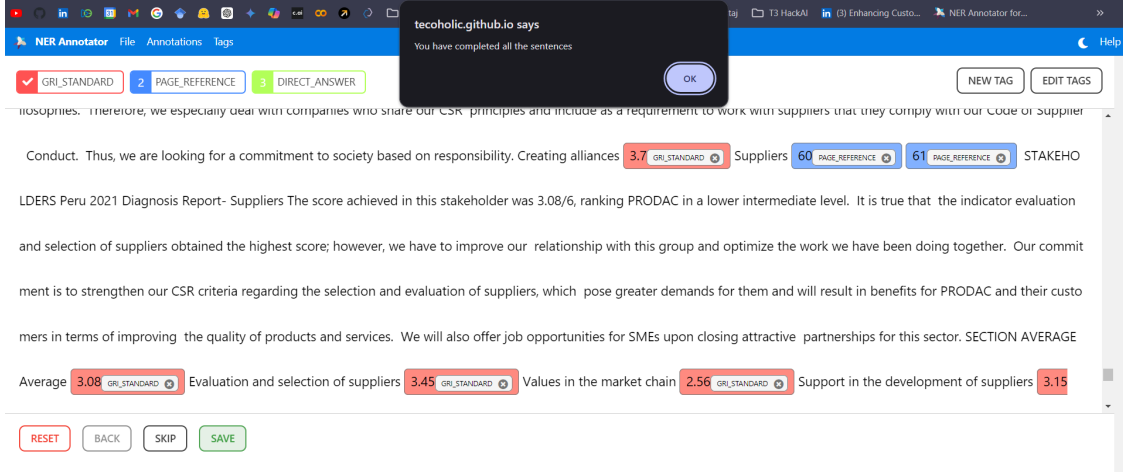
ty. Creating alliances 3.7 GRI_STANDARD Suppliers 60 61

!! Etiketlemeye başlamadan aşağıdaki yazıyı dikkatle okuyunuz:

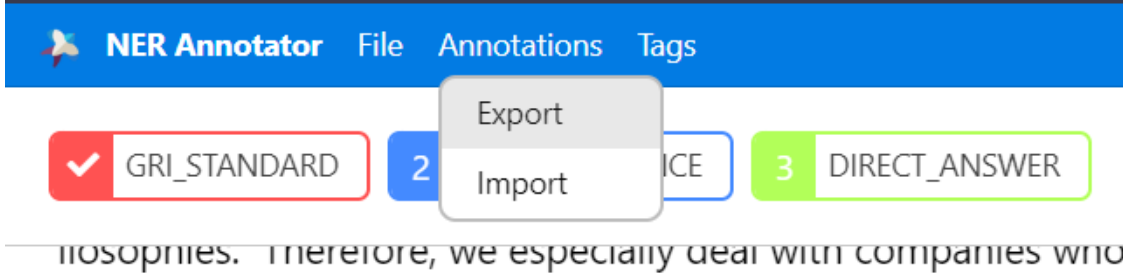
ETİKETLERKEN DİKKAT EDİLMESİ GEREKENLER

Adım 3: Etiketlenen Dosyayı Kaydet

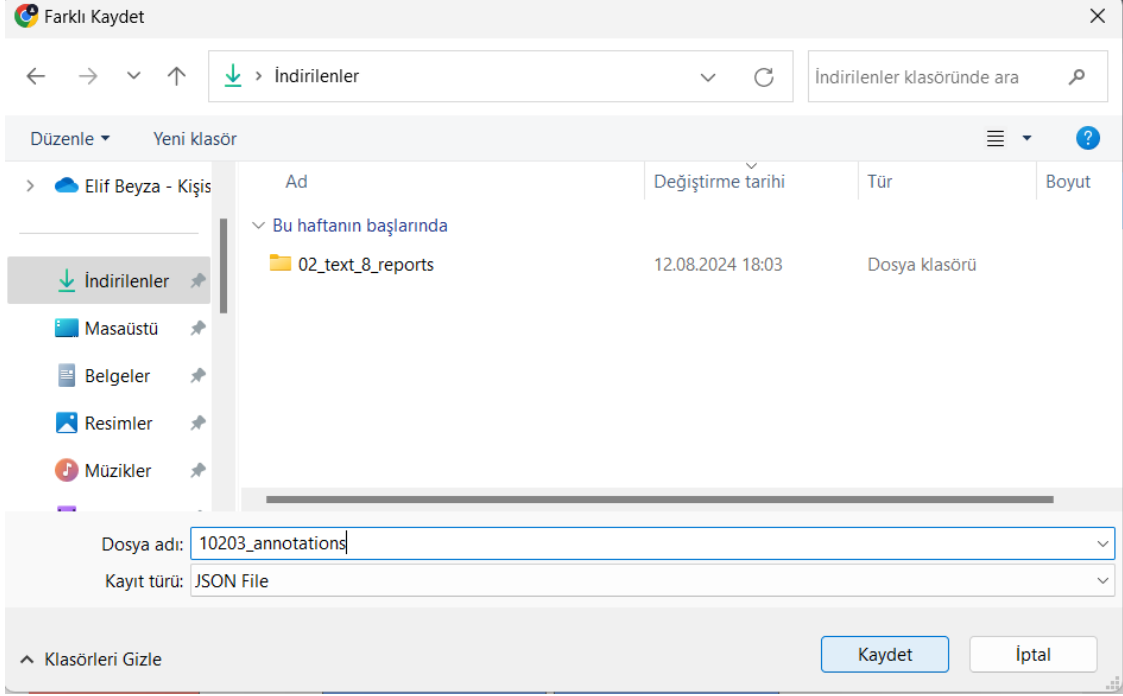
- Tüm etiketlemeleri tamamladıysan **SAVE** tuşuna basarak kaydet
- Çıkan popup OK basarak kapat



- sol üst köşede Annotations >> Export tuşlarına baş



- Etikelenen metni Report_ID_annotations.json adı ile kaydet. (Report_ID en başta tool içine yüklediğimiz txt dosyasının adıdır)



Tebrikler NER Annotation Tool ile etiketlemeyi tamamladınız! Şimdi sırada **etiketlemeleri Spacy Displacy ile kontrol etmeniz gerekiyor.**

Spacy Displacy Nedir?

Spacy Displacy, doğal dil işleme (NLP) alanında oldukça popüler olan bir Python kütüphanesi olan Spacy'nin, bir cümlenin veya metnin sintaktik yapısını ve adlandırılmış varlıklarını görselleştirmek için kullandığı bir araçtır. Basitçe söylemek gerekirse, Displacy, bir modelin bir metni nasıl anladığını görsel olarak sunar.

Spacy Displacy Nasıl Kullanılır?

- Boş bir python (.ipynb) dosyası oluşturun.
- Aşağıdaki kodları ekleyin.

```
import json
import spacy
```

```

from spacy import displacy

# Spacy modelini yükle
nlp = spacy.load("en_core_web_sm")

# JSON dosyasının yolunu belirtin
file_path = "/content/10203_annotations.json"

# JSON dosyasını açıp verileri yükleyin
with open(file_path, 'r', encoding='utf-8') as file:
    data = json.load(file)

    if 'annotations' in data:
        for annotation in data['annotations']:
            if annotation is not None:
                text = annotation[0] # Metin
                entities = [(ent[0], ent[1], ent[2]) for ent in

                # Displacy için gereken formatta veriyi hazırla
                spacy_displacy_data = {
                    "text": text,
                    "ents": [{"start": start, "end": end, "label": label} for start, end, label in entities],
                    "title": None
                }

                # Displacy ile görselleştirme yapın
                displacy.render(spacy_displacy_data, style="ent"

```

- file_path olarak Report_ID_annotations.json adlı etiketlenmiş dosyayı göster.
- Çıktıyı kontrol et. Karakterlerde kayma var mı? Etiketlemeler doğru gözüküyor mu?

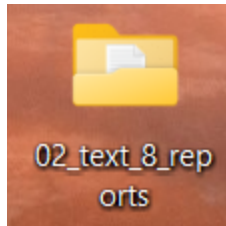
embedded in their philosophies. Therefore, we especially deal with companies who share our CSR principles and include as a requirement to work with suppliers that they comply with our Code of Supplier Conduct. Thus, we are looking for a commitment to society based on responsibility. Creating alliances 3.7 GRI_STANDARD Suppliers 60 PAGE_REFERENCE 61 PAGE_REFERENCE STAKEHOLDERS Peru 2021 Diagnosis Report- Suppliers The score achieved in this stakeholder was 3.08/6, ranking PRODAC in a lower intermediate level. It is true that the indicator evaluation and selection of suppliers obtained the highest score; however, we have to improve our relationship with this group and optimize the work we have been doing together. Our commitment is to strengthen our CSR criteria regarding the selection and evaluation of suppliers, which pose greater demands for them and will result in benefits for PRODAC and their customers in terms of improving the quality of products and services. We will also offer job opportunities for SMEs upon closing attractive partnerships for this sector. SECTION AVERAGE Average 3.08 GRI_STANDARD Evaluation and selection of suppliers 3.45 GRI_STANDARD Values in the market chain 2.56 GRI_STANDARD Support in the development of suppliers 3.15 GRI_STANDARD Our commitment to suppliers also involves supporting their development. Hence, in 2009, we set ourselves the objective to achieve 240 man/hours of training customers and suppliers. We not only achieved this goal, but also surpassed it, reaching only considering training to suppliers 198 man/hours. In addition, and considering that the first steps to set up a company can be complicated, this year we formalized our Inclusive Business program. This program seeks to promote the incorporation of small businesses providing seed capital and legal advice: examples of which are: Transportes Aldair, Las Orquideas, and Transporte Zaavedra. 2009 CSR training Las Orquideas Gardening company – Inclusive Business 1 2 62 63 Peru 2021 Diagnosis Report- State The outcome of the

Her şey yolundaysa ve etiketlemelerde sorun yoksa:

- Boş bir klasör oluşturun.
- Klasöre isim ver. Klasörün ismi etiketlenmek için gönderilen txt dosyalarının bulunduğu klasöre göre isimlendirilmeli.
- Report_ID_annotations.json adlı etiketlenmiş dosyayı ve bundan sonra etiketlediğin her dosyayı bu klasöre ekle
- Tebrikler! Etiketlenmek için bekleyen bir sonraki dosyaya geçebilirsiniz.

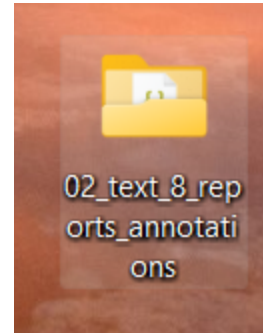
Örnek:

Size gönderilen klasör (etiketlenmek için gönderilen txt dosyalarının bulunduğu klasör) ⇒
xx_text_x_reports



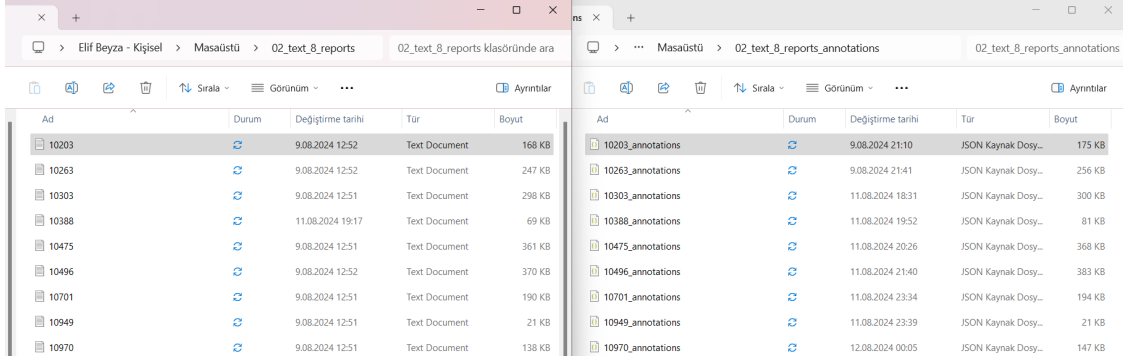
Yeni oluşturulan boş klasör (Report_ID_annotations.json adlı etiketlenmiş dosyaları bunun içine ekle)

⇒ **xx_text_x_reports_annotations**



Solda etiketlenmek için bekleyen
txt dosyaları

Sağda etiketlenmiş json dosyaları



Ad	Durum	Değiştirme tarihi	Tür	Boyut
10203	✓	9.08.2024 12:52	Text Document	168 KB
10263	✓	9.08.2024 12:52	Text Document	247 KB
10303	✓	9.08.2024 12:51	Text Document	298 KB
10388	✓	11.08.2024 19:17	Text Document	69 KB
10475	✓	9.08.2024 12:51	Text Document	361 KB
10496	✓	9.08.2024 12:52	Text Document	370 KB
10701	✓	9.08.2024 12:51	Text Document	190 KB
10949	✓	9.08.2024 12:51	Text Document	21 KB
10970	✓	9.08.2024 12:51	Text Document	138 KB

Ad	Durum	Değiştirme tarihi	Tür	Boyut
10203_annotations	✓	9.08.2024 21:10	JSON Kaynak Dosy...	175 KB
10263_annotations	✓	9.08.2024 21:41	JSON Kaynak Dosy...	256 KB
10303_annotations	✓	11.08.2024 18:31	JSON Kaynak Dosy...	300 KB
10388_annotations	✓	11.08.2024 19:52	JSON Kaynak Dosy...	81 KB
10475_annotations	✓	11.08.2024 20:26	JSON Kaynak Dosy...	368 KB
10496_annotations	✓	11.08.2024 21:40	JSON Kaynak Dosy...	383 KB
10701_annotations	✓	11.08.2024 23:34	JSON Kaynak Dosy...	194 KB
10949_annotations	✓	11.08.2024 23:39	JSON Kaynak Dosy...	21 KB
10970_annotations	✓	12.08.2024 00:05	JSON Kaynak Dosy...	147 KB

İsmlendirmeler verileri sonrasında takip edebilmek açısından önemlidir.
İsmlendirmeleri dikkatle yapmanız rica olunur.