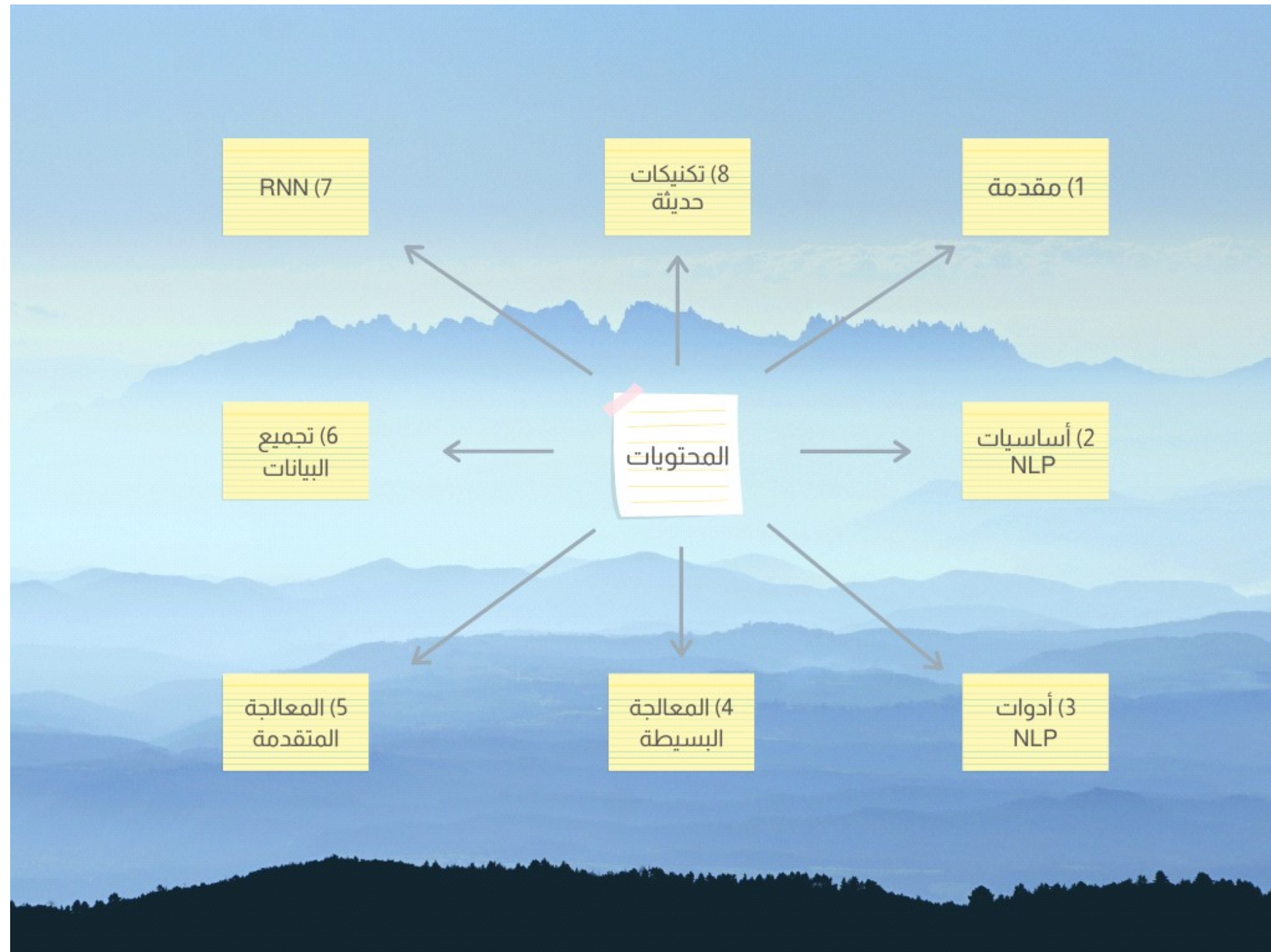


NATURAL LANGUAGE PROCESSING

المعالجة اللغوية الطبيعية



المحتويات

				التطبيقات	العقبات و التحديات	تاريخ NLP	ما هو NLP	المحتويات	1) مقدمة
					البحث في النصوص	ملفات pdf	الملفات النصية	المكتبات	2) أساسيات NLP
T.Visualization	Syntactic Struc.	Matchers	Stopwords	NER	Stem & Lemm	POS	Sent. Segm.	Tokenization	3) أدوات NLP
	Dist. Similarity	Text Similarity	TF-IDF	BOW	Word2Vec	T. Vectors	Word embed	Word Meaning	4) المعالجة البسيطة
T. Generation	NGrams	Lexicons	GloVe	L. Modeling	NMF	LDA	T. Clustering	T. Classification	5) المعالجة المتقدمة
	Summarization & Snippets		Ans. Questions		Auto Correct	Vader	Naïve Bayes	Sent. Analysis	
Search Engine	Relative Extraction		Information Retrieval		Information Extraction		Data Scraping	Tweet Collecting	6) تجميع البيانات
					Rec NN\TNN	GRU	LSTM	Seq to Seq	7) RNN
Chat Bot	Gensim	FastText	Bert	Transformer	Attention Model	T. Forcing	CNN	Word Cloud	8) تكنيكات حديثة

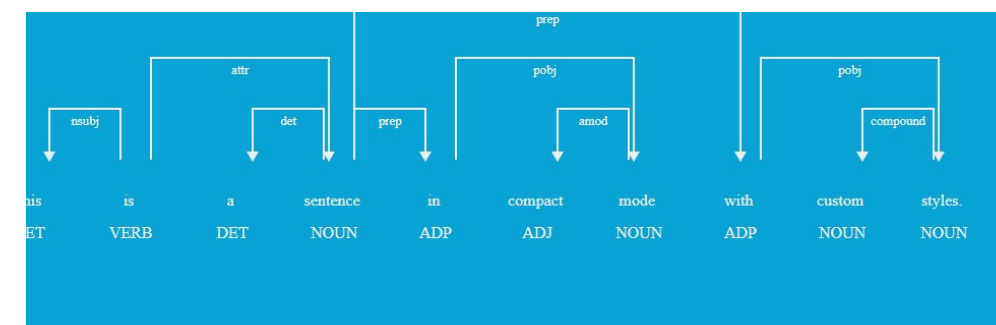
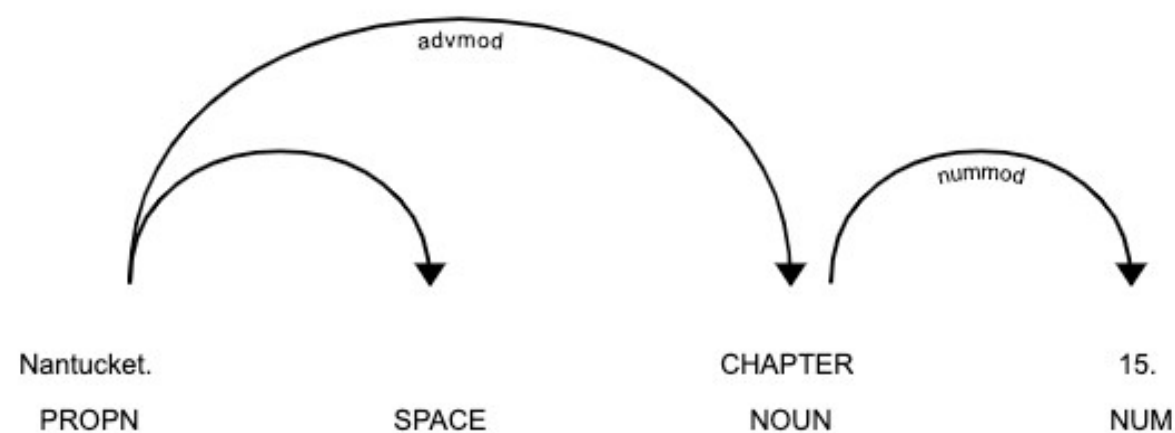
القسم الثالث : أدوات NLP

الجزء التاسع : Text Visualization

=====

و هي الأداة التي تقوم بعرض العلاقات بين الكلمات بشكل مرأي واضح , سواء في لظهار , dep , NER
و تأتي من اداة displacy من نفس مكتبة سباسي , بعدها نقوم باعطائه جملة معينة , و نطلب عمل توضيع مرأي لها
و تأتي في شكلين أساسيين , اظهار العلاقات بين الكلمات dependencies او الكلمات الهامة entities
و يتم هذا عبر استدعاء كلاس displacy من مكتبة spacy , وتحديد نوع ال style , هل هو ent , dep
و في نوع العلاقات dep يمكن ان تكون السهم دائرية او خطوط مستقيمة
كما يمكن عرض الرسم علي جوبيتر فقط بأمر render او إلحاقه بصفحة علي موقع بأمر serve

F.B.I. Agent **Peter Strzok PERSON**, **Who Criticized Trump PERSON** in Texts, Is **Fired GPE** - **The New York Times ORG** SectionsSEARCHSkip to contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's **PaperAdvertisementSupported ORG** byF.B.I. Agent **Peter Strzok PERSON**, **Who Criticized Trump PERSON** in Texts, Is FiredImagePeter Strzok, a top **F.B.I. GPE** counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President **Trump PERSON** were uncovered, was fired. **CreditT.J. Kirkpatrick PERSON** for **The New York TimesBy Adam Goldman ORG** and **Michael S. SchmidtAug PERSON**. **13 CARDINAL**, **2018WASHINGTON CARDINAL** — **Peter Strzok PERSON**, the **F.B.I. GPE** senior counterintelligence agent who disparaged President **Trump PERSON** in inflammatory text messages and helped oversee the **Hillary Clinton PERSON** email and **Russia GPE** investigations, has been fired for violating bureau policies, Mr. **Strzok PERSON**'s lawyer said **Monday DATE**.Mr. Trump and his allies seized on the texts — exchanged during the **2016 DATE** campaign with a former **F.B.I. GPE** lawyer, **Lisa Page — in PERSON** assailing the **Russia GPE** investigation as an illegitimate “witch hunt.” Mr. **Strzok PERSON**, who rose over **20 years DATE** at the **F.B.I. GPE** to become one of its most experienced counterintelligence agents, was a key figure in **the early months DATE** of the inquiry.Along with writing the texts, Mr. **Strzok PERSON** was accused of sending a highly sensitive search warrant to his personal email account.The **F.B.I. GPE** had been under immense political pressure by Mr. **Trump PERSON** to dismiss Mr. **Strzok PERSON**, who was removed **last summer DATE** from the staff of the special counsel, **Robert S. Mueller III PERSON**. The president has repeatedly denounced Mr. **Strzok PERSON** in posts on **Twitter EVENT**, and on **Monday DATE** expressed satisfaction that he had been sacked.Mr. **Trump's ORG** victory traces back to **June DATE**, when Mr. **Strzok PERSON**'s conduct was laid out in a wide-ranging inspector general's report on how the **F.B.I. GPE** handled the investigation of **Hillary Clinton's PERSON** emails in the run-up to the **2016 DATE** election. The report was critical of Mr. **Strzok PERSON**'s conduct in sending the

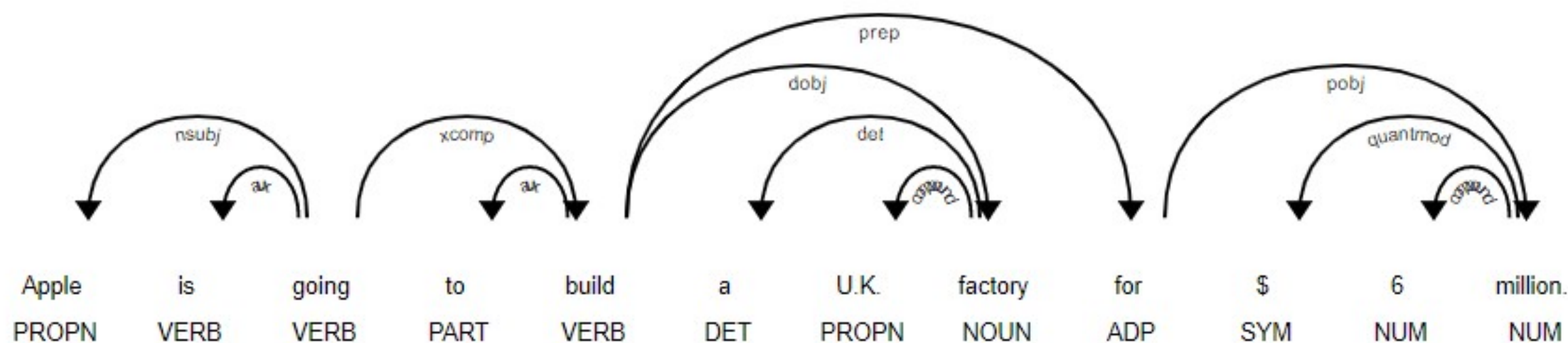


الكود المستخدم :

```
import spacy
from spacy import displacy
nlp = spacy.load('en_core_web_sm')
```

و هنا مثال علي نموذج dep

```
doc = nlp(u'Apple is going to build a U.K. factory for $6 million.')
displacy.render(doc, style='dep', jupyter=True, options={'distance': 80})
```



فهنا تم رسم علاقات بين الكلمات و بعضها , علي كل اسهم و توضيح نوع كل كلمة و علاقاتها بالباقي و هذا عبر ستايل dep

و من الممكن عرض الـ dependencies لرؤيتها

```
for token in doc:  
    print(f'{token.text:{10}} {token.pos_:{7}} {token.dep_:{7}} {spacy.explain(token.dep_)})'
```

و لعمل هذا علي موقع معين , يتم استخدام أمر `serve`

```
doc = nlp(u'This is a sentence.')  
displacy.serve(doc, style='dep')
```

ولو كان لدينا جمل كثيرة , فيفضل عمل تقطيع للجمل كاجزاء (spans) قبل رسمها

```
text = """In ancient Rome, some neighbors live in three adjacent houses. In the center is  
the house of Senex, who lives there with wife Domina, son Hero, and several slaves, the  
lovely Philia, one of the courtesans in the House of Lycus (albeit still a virgin)."""  
doc = nlp(text)  
sentence_spans = list(doc.sents)
```

```
displacy.render(sentence_spans, style="dep",options={'distance': 50})
```

او هكذا

```
doc1 = nlp("This is a sentence.")  
doc2 = nlp("This is another sentence.")  
html = displacy.render([doc1, doc2], style="dep", page=True)
```

مثال اخر

```
doc = nlp("Rats are various medium-sized, long-tailed rodents.")  
displacy.render(doc, style="dep")
```

استخدام اداة HTML لعرضها

```
from IPython.core.display import display, HTML
```

```
html = displacy.render(doc, style="dep")  
display(HTML(html))
```

و يمكن حفظ الجراف كصورة svg هكذا


```
from pathlib import Path

nlp = spacy.load("en_core_web_sm")
sentences = ["This is an example.", "This is another one."]
for sent in sentences:
    doc = nlp(sent)
    svg = displacy.render(doc, style="dep", jupyter=False)
    file_name = '-'.join([w.text for w in doc if not w.is_punct]) + ".svg"
    output_path = Path( file_name)
    output_path.open("w", encoding="utf-8").write(svg)
```

*_**

و يمكن عمل الاسهم مستطيلة , عبر تغيير خاصية **"compact": True** كما يمكن تغيير الالوان و باقي الخواص

```
doc = nlp(u'Over the last quarter Apple sold nearly 20 thousand iPods for a profit of $6 million.')
```

[illegible]

* * * * *

و هنا يتم استخدام ستايل ent لاظهار شكل توضيحي مختلف , وهذه الاداة تقوم باظهار قيم الـ NER بشكل مرئي

```
doc = nlp(u'Over the last quarter Apple sold nearly 20 thousand iPods for a profit of $6 million.')
```

```
displacy.render(doc, style='ent', jupyter=True)
```

و فيه تم عمل هايلايت علي الكلمات الهامة , وتحديد نوعها و معناها , و هي تقوم بانتقاء الكلمات الهامة , وعمل هايلايت لها , وكتابة نوعها , مع معلومية ان الهايلايت يكون بلون و ستايل مود لكل نوع معين , فالشركات ازرق و الارقام رمادي و هكذا

و ايضا يمكن عرضها علي السيرفر

```
displacy.serve(doc, style='ent')
```

و هنا امثلة اخري و تغيير الخصائص فيها

```
doc = nlp(u'Over the last quarter Apple sold nearly 20 thousand iPods for a profit of $6 million. '
```

```
u'By contrast, Sony sold only 7 thousand Walkman music players.')
```

```
displacy.render(doc, style='ent', jupyter=True)
```

ويمكن عمل دالة بحيث تقوم بعمل هايلايت للسطور التي فيها كلمات هامة , وطابعة باقي الجمل التي ليس فيها هكذا :

```
doc2 = nlp(u'Over the last quarter Apple sold nearly 20 thousand iPods for a profit of $6 million. '
```

```
u'By contrast, my kids sold a lot of lemonade.')
```

```
for sent in doc2.sents:
```

```
    docx = nlp(sent.text)
```

```
    if docx.ents:
```

```
        displacy.render(docx, style='ent', jupyter=True)
```

```
    else:
```

```
        print(docx.text)
```

لكن بعمل قاموس فيه انواع معينة سيظهرها و يتجاهل الباقي

```
options = {'ents': ['ORG', 'PRODUCT']}
```

```
displacy.render(doc, style='ent', jupyter=True, options=options)
```

و يمكن التحكم في الالوان عبر عمل قاموس الالوان الذي يدخل في قاموس اوبشنز

```
colors = {'ORG': 'linear-gradient(90deg, #aa9cfc, #fc9ce7)', 'PRODUCT': 'radial-gradient(yellow, green)'}
```

```
options = {'ents': ['ORG', 'PRODUCT'], 'colors': colors}
```

```
displacy.render(doc, style='ent', jupyter=True, options=options)
```

```
text = "When Sebastian Thrun started working on self-driving cars at Google in 2007,  
few people outside of the company took him seriously."
```

```
doc = nlp(text)
```

```
displacy.render(doc, style="ent")
```

```
colors = {"ORG": "linear-gradient(90deg, #aa9cfc, #fc9ce7)"}
```

```
options = {"ents": ["ORG"], "colors": colors}
```

```
displacy.render(doc, style="ent", options=options)
```

```
doc = nlp("This is a sentence about Google.")
```



```
doc.user_data["title"] = "This is a title"
displacy.render(doc, style="ent")
```

```
ex = [{"text": "But Google is starting from behind.",
      "ents": [{"start": 4, "end": 10, "label": "ORG"}],
      "title": None}]
html = displacy.render(ex, style="ent", manual=True)
```

* * * * *

و في اللغة العربية , تقوم المكتبة بعرض المهام بكفاءة متوسطة