

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 الثاني : أساسيات NLP

الجزء الأول : المكتبات المستخدمة في NLP

● المكتبات التي سيتم استخدامها في الكورس هي

- nltk
- spacy
- re
- genism
- fasttext
-

● و المكتبات التقليدية السابقة مثل

- **numpy**
- **pandas**
- **sklearn**
- **Matplotlib**

* * * * *

- أولًا مكتبة nltk



- تعد مكتبة NLTK من اقوي مكتبات NLP , وهي اختصار Natural Language Toolkit و التي نشأت عام 2001 علي يد Steven Bird, Edward Loper, Ewan Klein , و اخر اصدار لها حتي الان هو 3.5 في ابريل 2020

- و يتم تحميلها بأمر

`pip install nltk`

● ثم التأكد من الإصدار

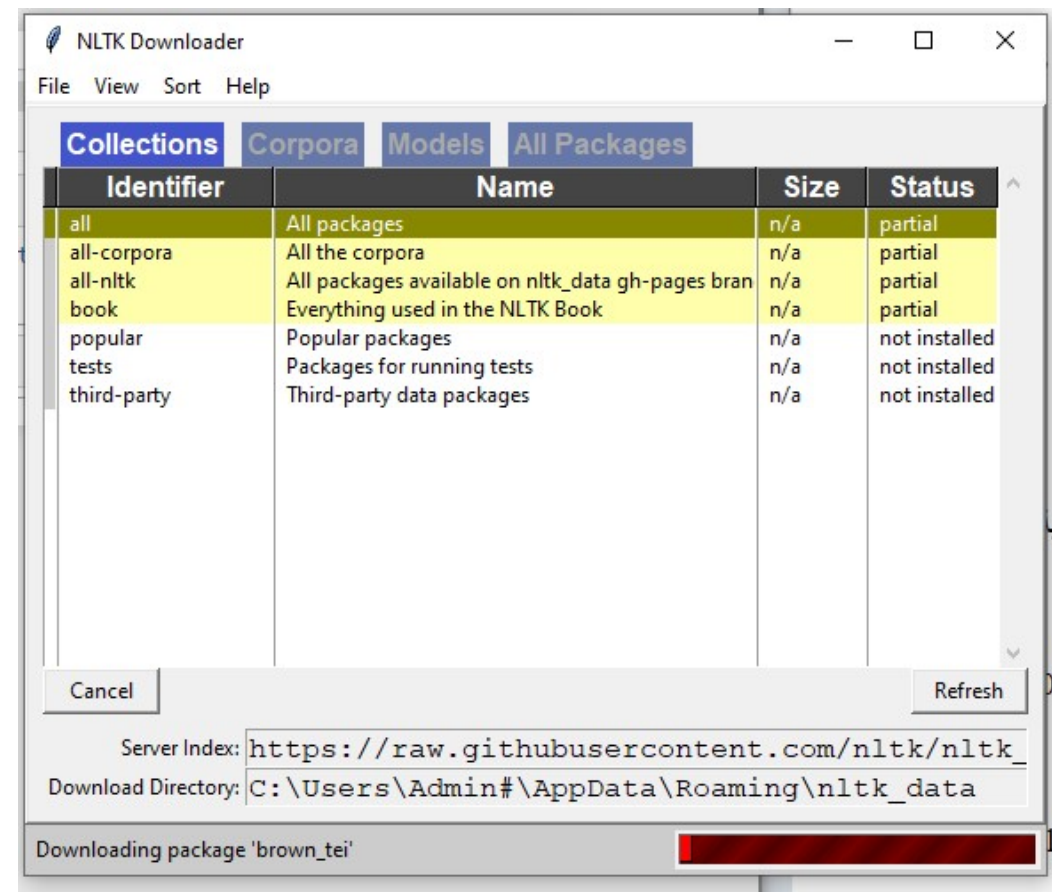
```
import nltk
```

```
nltk.__version__
```

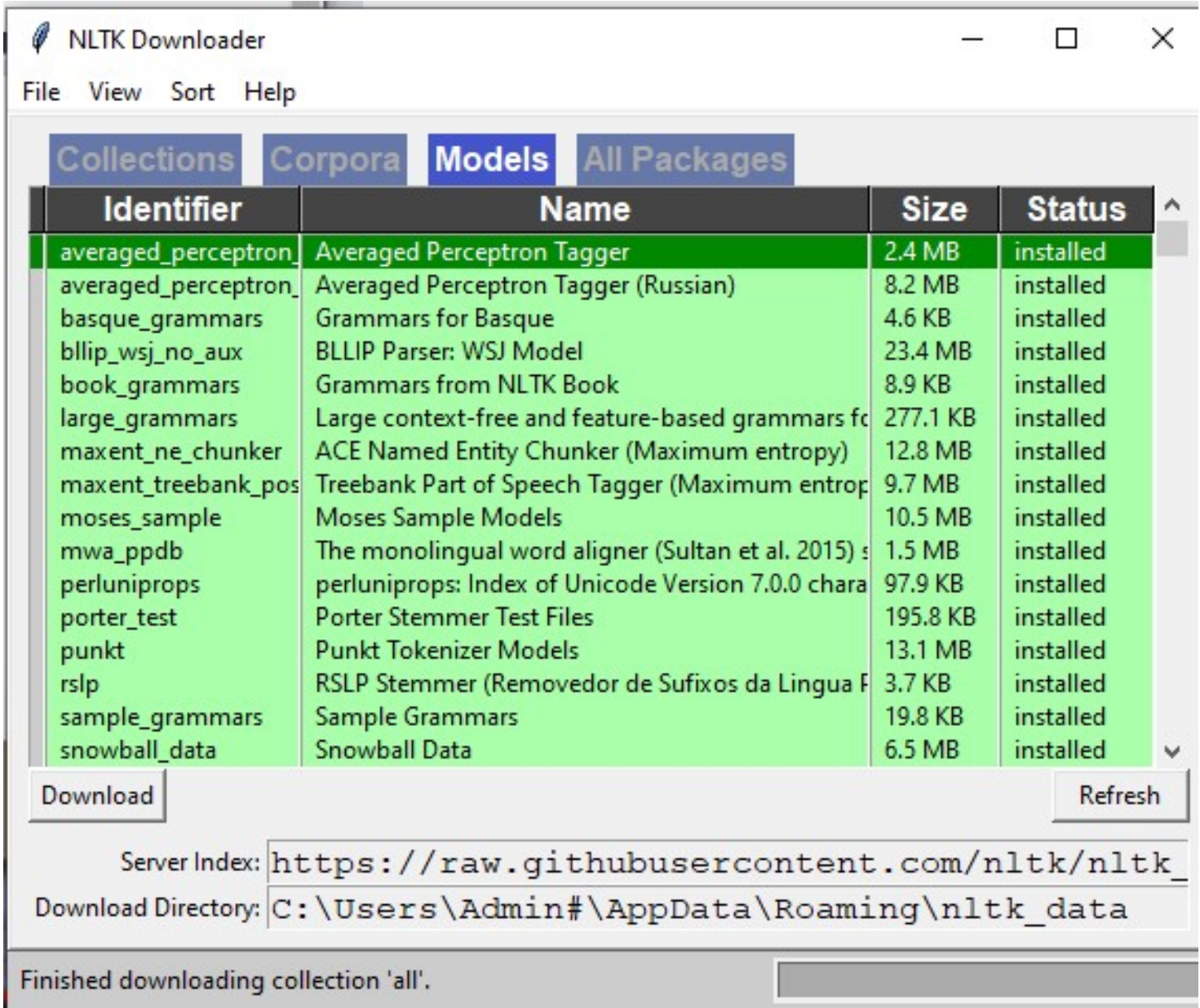
```
nltk.download()
```

● ولو قمت بكتابة أمر

● ستظهر لك هذه الشاشة , التي تقوم بتحميل كل ملحقات المكتبة بسهولة



● حتي تتحول الي هذا الشكل



* _ * _ * _ * _ * _ * _ * _ * _ * _ * _ * _ *

- مكتبة spacy هي مكتبة مفتوحة المصدر و خاصة بالتعامل مع العمليات , NLP و تم اطلاقها في 2015
- و يمكن تحميل المكتبة عبر الامر :

`pip install spacy`

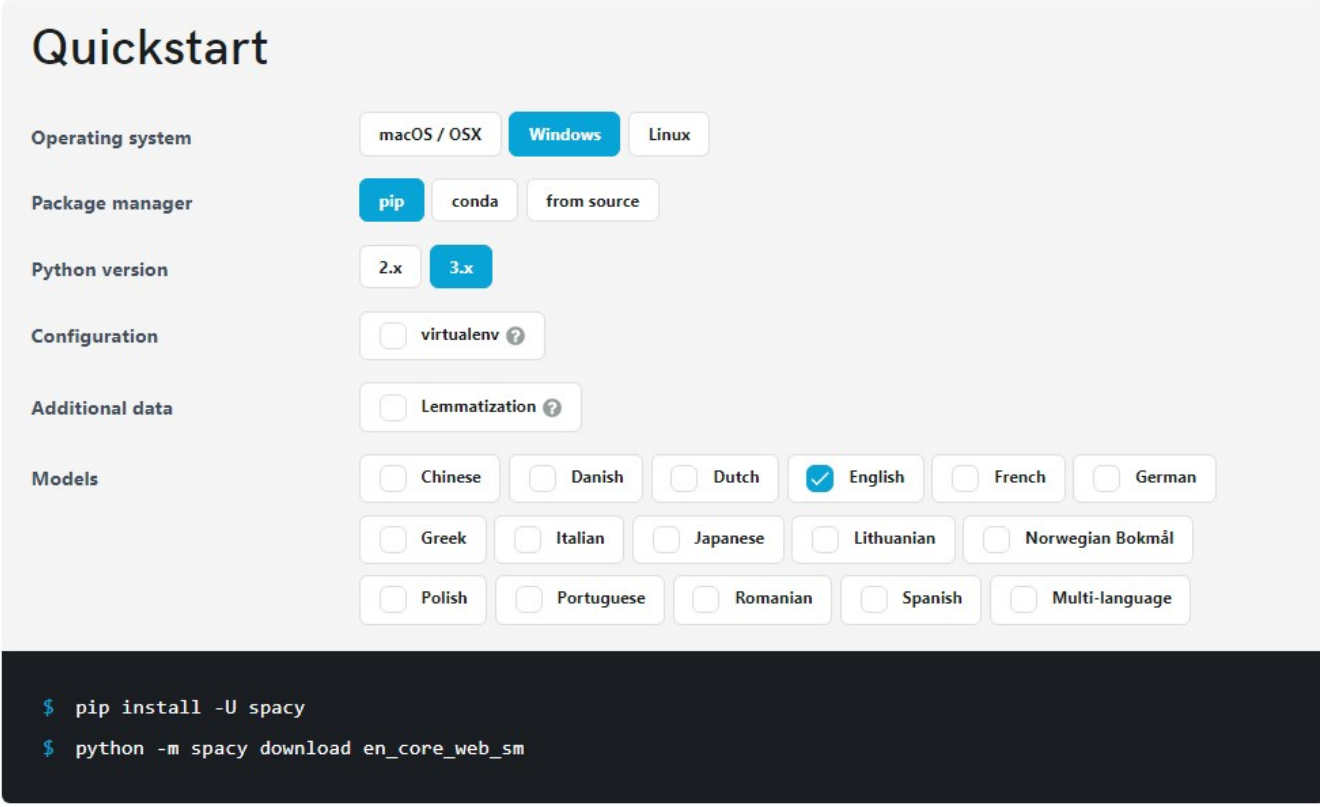
`import spacy`

`spacy.__version__`

- و التأكد من الإصدار هكذا :

- و هنا رابط ايجاد كود التحميل

<https://spacy.io/usage>



The image shows the 'Quickstart' form on the Spacy website. It is a configuration tool for installing Spacy. The form has several sections: 'Operating system' with buttons for 'macOS / OSX', 'Windows' (selected), and 'Linux'; 'Package manager' with buttons for 'pip' (selected), 'conda', and 'from source'; 'Python version' with buttons for '2.x' and '3.x' (selected); 'Configuration' with a checkbox for 'virtualenv' (unchecked); 'Additional data' with a checkbox for 'Lemmatization' (unchecked); and 'Models' with a grid of checkboxes for various languages. 'English' is selected. At the bottom, there is a dark box with terminal commands: '\$ pip install -U spacy' and '\$ python -m spacy download en_core_web_sm'.

Quickstart

Operating system: macOS / OSX, **Windows**, Linux

Package manager: **pip**, conda, from source

Python version: 2.x, **3.x**

Configuration: ☐ virtualenv ?

Additional data: ☐ Lemmatization ?

Models:

<input type="checkbox"/> Chinese	<input type="checkbox"/> Danish	<input type="checkbox"/> Dutch	<input checked="" type="checkbox"/> English	<input type="checkbox"/> French	<input type="checkbox"/> German
<input type="checkbox"/> Greek	<input type="checkbox"/> Italian	<input type="checkbox"/> Japanese	<input type="checkbox"/> Lithuanian	<input type="checkbox"/> Norwegian Bokmål	
<input type="checkbox"/> Polish	<input type="checkbox"/> Portuguese	<input type="checkbox"/> Romanian	<input type="checkbox"/> Spanish	<input type="checkbox"/> Multi-language	

```
$ pip install -U spacy
$ python -m spacy download en_core_web_sm
```

و الأهم , هو تحميل الباقيات في المكتبة

و يتم هذا عبر فتح نافذة anaconda prompt يتم كتابة الامر لتحميل اللغة الانجليزية

```
python -m spacy download en
```

و هناك العديد من اللغات , مثل الفرنسية و الالمانية و الاسبانية , لكن لا توجد العربية بعد

ثم تحميل الملف الصغير من هنا

```
python -m spacy download en_core_web_sm
```

أو الكبير من هنا

```
python -m spacy download en_core_web_lg
```

و يتم التأكد من تحميلها , عبر عمل ريستارت للكرنل , وتشغيل الكود

```
import spacy
```

```
nlp = spacy.load('en_core_web_sm')
```

علي ان يكون الجهاز اونلاين

و هناك عدد من المراحل الهامة في تعامل spacy مع التعامل اللغوي :

○ تحميل و استدعاء المكتبة

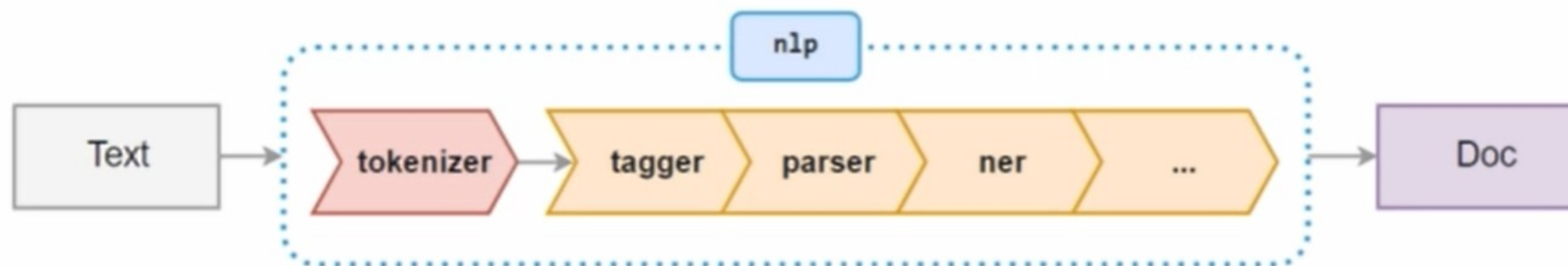
○بناء المسار pipeline

○ استخدام الرموز tokens

○ اختيار كلمات محددة Tagging

Token Attributes

وهذا الرسم يوضح المسار الذي يسلكه `spacy` , وهذا هو المسار المذكور `pipeline`



ففي البداية يدخل النص , ثم يتم عمل كل عمليات الـ NLP مثل الترميز tokenizer ثم الوسم tagger ثم مصفوفة البارسر , يليها عملية التعرف علي الكائنات المسماة Named Entity Recognition و اختصارها NER كذلك عمليات Stemming و POS و Lemmatization و غيرها

* * * * *

● ثالثا مكتبة re

و هي اختصار كلمة regular expressions اي التعبيرات المعتادة

و غالبا ما تكون محملة بالفعل مع باقة اناكوندا , ويمكن التأكد عبر تجربتها هنا :

```
import re
```

و قد تم وضع أسسها في الخمسينات , لكن نشأت كمكتبة في التسعينات علي يد المبرمج فيليب هازل

* * * * *

أما مكتبتَي genism & FastText فسيتم تناولها في القسم الأخير