

spaCy

<https://spacy.io/>

インストール

<https://spacy.io/usage#quickstart>

- ・ Windows のコンソールで

```
pip install spacy
```

- ・ Windows のシェルで管理者として実行（右ボタン）

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

- ・ Mac の場合

```
pip3 install spacy  
python3 -m spacy download en_core_web_sm
```

- ・ Python のシェルで

```
import spacy
```

- ・ それか、スクリプトの先頭に書いておく

Anaconda Navigator の Environments で追加できないときは、

<https://github.com/conda/conda/issues/9423>

- ・ PowerShell で、

```
conda install -c conda-forge spacy  
conda install -c conda-forge spacy-model-en_core_web_sm
```

<https://anaconda.org/conda-forge/spacy>

しくみ

<https://spacy.io/usage/spacy-101>

- ・ 言語処理に必要な「データベース」が作ってある。
- ・ その「データベース」に基づいて、言語処理（自然言語解析）を行う。

処理

- ・ tokenizer: 単語に分けて

- `tagger` : 品詞分析して
- `parser` : 構文解析して
- `ner` : エンティティ付与

手順

- 「データベース ("en_core_web_sm")」の読み込み

```
spacy.load("en_core_web_sm")
```

- 読み込んだものを例えば `nlp` という名前で保存して、それを使って処理できるようにする。

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

- 処理するデータを変数に入れて

```
sample = "Parents drive their children everywhere."
```

- `nlp` で処理した結果を変数 (`sample_doc`) に保存する

```
sample_doc = nlp(sample)
```

- あとは、この `sample_doc` にたいして、`spaCy` の関数を使って「処理」をする。
 - 表面上は、`sample_doc` を表示させると、普通に文が表示されるだけ。
 - その裏で、処理結果の情報も蓄えられている。

属性 Attributes

- 言語情報が各 token の属性として付与されている。
 - 属性情報そのものは、プログラミング処理の都合上、数値 (hash values) になっている
- 人の目にわかるようにするには、属性名の後ろにアンダーバーを付ける

チャンク 名詞句

- `nlp` での分析結果には、チャンクという単位もある
- チャンクの役割 (文の構成要素) と依存関係も分析できる

<code>chunk.text</code>	<code>chunk.root.text</code>	<code>chunk.root.dep_</code>	<code>chunk.root.head.text</code>
Parents	Parents	nsubj	drive
their children	children	dobj	drive

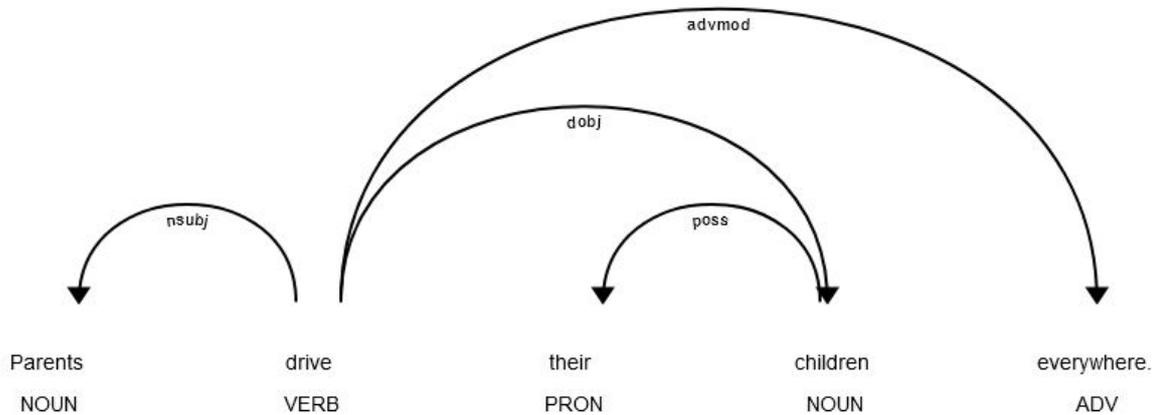
関数

依存関係の可視化

<https://spacy.io/usage/visualizers>

- Jupyter Notebook で実行する場合、最後の `displacy.serve` の代わりに、`displacy.render` を使

う。



Demo

<https://explosion.ai/demos/>

References

<https://github.com/explosion/spacy>

<https://ishitonton.hatenablog.com/entry/2018/11/24/004748>

<https://spacy.io/usage/spacy-101>

日本語処理

参考：<https://qiita.com/wf-yamaday/items/3ffdcc15a5878b279d61>

GiNZA 日本の自然言語処理ライブラリー

<https://megagonlabs.github.io/ginza/>

解説

<https://qiita.com/poyo46/items/7a4965455a8a2b2d2971>

インストール

- ・ [Anaconda Prompt](#) を管理者として実行

```
pip install ja-ginza
```

spaCy から使えるようにする