spacy

July 21, 2017

1 SpaCy

Spacy is an open source platform available in Python for Natural Language Processing. Its speed, comprehensiveness, and thorough documentation make it a strong choice for both industry and academia.

1.0.1 install

```
pip3 install spacy
```

1.0.2 import

```
In [171]: import spacy
```

1.0.3 data

python3 -m spacy.en.download all

```
In [222]: x = spacy.load('en')
```

1.0.4 languages

```
In [173]: parser = spacy.en.English()

In [174]: spacy.de.German()

Out[174]: <spacy.de.German at 0x157666ba8>

In [175]: spacy.fr.French()

Out[175]: <spacy.fr.French at 0x1eb8590b8>

In [176]: spacy.es.Spanish()

Out[176]: <spacy.es.Spanish at 0x14b461a20>

In [177]: spacy.it.Italian()

Out[177]: <spacy.it.Italian at 0x14b461f98>
```

```
In [178]: spacy.pt.Portuguese()

Out[178]: <spacy.pt.Portuguese at 0x14b461b38>

In [179]: spacy.nl.Dutch()

Out[179]: <spacy.nl.Dutch at 0x14b461ef0>

In [180]: spacy.sv.Swedish()

Out[180]: <spacy.sv.Swedish at 0x13b89e978>

In [181]: spacy.fi.Finnish()

Out[181]: <spacy.fi.Finnish at 0x13b89e828>

In [182]: spacy.hu.Hungarian()

Out[182]: <spacy.hu.Hungarian at 0x13b89e6a0>

In [183]: spacy.bn.Bengali()

Out[183]: <spacy.bn.Bengali at 0x13b89e7f0>

In [184]: spacy.he.Hebrew()

Out[184]: <spacy.he.Hebrew at 0x13b89e550>

In [185]: spacy.zh.Chinese()

Out[185]: <spacy.zh.Chinese at 0x13b89ee10>
```

1.0.5 specific data

```
In [245]: parser.vocab['NASA']
          parser.vocab['apple']
          parser.vocab['UNK']

Out[245]: <spacy.lexeme.Lexeme at 0x195764360>
```

1.0.6 loading parent doc

```
In [223]: x = x("Hello, I like to program. My favorite language is Python.")
```

1.0.7 parent doc type

```
In [214]: x[0].lang_

Out[214]: 'en'
```

1.0.8 sentences

```
In [188]: for i in x.sents:
              print(i)
```

```
Hello, I like to program.
My favorite language is Python.
```

1.0.9 lower

```
In [189]: x[0].orth_
```

```
Out[189]: 'Hello'
```

```
In [190]: x[0].lower_
```

```
Out[190]: 'hello'
```

1.0.10 prefix

```
In [191]: x[0].prefix_
```

```
Out[191]: 'H'
```

1.0.11 suffix

```
In [192]: x[0].suffix_
```

```
Out[192]: 'llo'
```

1.0.12 shape

```
In [193]: x[0].shape_
```

```
Out[193]: 'Xxxxx'
```

1.0.13 log probability

```
In [194]: x[0].prob
```

```
Out[194]: -11.369197845458984
```

1.0.14 sentiment

```
In [231]: x.sentiment
```

```
Out[231]: 0.0
```

1.0.15 brown cluster ID

```
In [195]: x[0].cluster
```

```
Out[195]: 1726
```

1.0.16 vectors

```
In [196]: king = x.vocab['king'].vector
```

1.0.17 lemmatizing

```
In [197]: for i in x:
              print(i,":",i.lemma_)
```

```
Hello : hello
, : ,
I : -PRON-
like : like
to : to
program : program
. : .
My : -PRON-
favorite : favorite
language : language
is : be
Python : python
. : .
```

1.0.18 parts of speech

```
In [198]: for i in x:
              print(i,":",i.pos_)
```

```
Hello : INTJ
, : PUNCT
I : PRON
like : VERB
to : PART
program : VERB
. : PUNCT
My : ADJ
favorite : ADJ
language : NOUN
is : VERB
Python : PROPN
. : PUNCT
```