SPACY CHEATSHEET

WRITTEN BY

LESLEY CORDERO

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.

install

```
pip3 install spacy
```

import

```
import spacy
```

data

python3 -m spacy.en.download all

```
x = spacy.load('en')
```

languages

```
parser = spacy.en.English()
```

```
spacy.de.German()
```

```
<spacy.de.German at 0x157666ba8>
```

```
spacy.fr.French()
```

```
<spacy.fr.French at 0x1eb8590b8>
```

```
spacy.es.Spanish()
```

```
<spacy.es.Spanish at 0x14b461a20>
```

```
spacy.it.Italian()
```

```
<spacy.it.Italian at 0x14b461f98>
```

```
spacy.pt.Portuguese()
```

```
<spacy.pt.Portuguese at 0x14b461b38>
```

```
spacy.nl.Dutch()
```

```
<spacy.nl.Dutch at 0x14b461ef0>
```

```
spacy.sv.Swedish()
```

```
<spacy.sv.Swedish at 0x13b89e978>
```

```
spacy.fi.Finnish()
```

```
<spacy.fi.Finnish at 0x13b89e828>
```

```
spacy.hu.Hungarian()
```

```
<spacy.hu.Hungarian at 0x13b89e6a0>
```

```
spacy.bn.Bengali()
```

```
<spacy.bn.Bengali at 0x13b89e7f0>
```

```
spacy.he.Hebrew()
```

```
<spacy.he.Hebrew at 0x13b89e550>
```

```
spacy.zh.Chinese()
```

```
<spacy.zh.Chinese at 0x13b89ee10>
```

specific data

```
parser.vocab['NASA']
parser.vocab['apple']
parser.vocab['UNK']
```

```
<spacy.lexeme.Lexeme at 0x195764360>
```

loading parent doc

```
x = x("Hello, I like to program. My favorite language is Python.")
```

parent doc type

```
x[0].lang_
```

```
'en'
```

sentences

```
for i in x.sents:
    print(i)
```

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

lower

```
x[0].orth_
```

```
'Hello'
```

```
x[0].lower_
```

```
'hello'
```

prefix

```
x[0].prefix_
```

```
'H'
```

suffix

```
x[0].suffix_
```

```
'llo'
```

shape

```
x[0].shape_
```

```
'Xxxxx'
```

log probability

```
x[0].prob
```

```
-11.369197845458984
```

sentiment

```
x.sentiment
```

```
0.0
```

brown cluster ID

```
x[0].cluster
```

```
1726
```

vectors

```
king = x.vocab['king'].vector
```

lemmatizing

```
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
. : .
```

parts of speech

```
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
```

entity types

```
x[0].ent_type
```

```
0
```

```
x[0].ent_iob_
```

```
'O'
```

0 = no tag is assigned.
1 = `I` = inside an entity.
2 = `O` = no tag is assigned.
3 = `B` = begins an entity.

PoS string

```
x[0].dep_
```

```
'intj'
```

entities

```
for i in x.ents:
    print(i,i.label_)
```

```
Python PERSON
```

nounphrases

```
for i in x.noun_chunks:
    print(i)
```

similarity

```
print(x[5],x[9])
x[5].similarity(x[9])
```

dependency trees

```
for i in x.sents:
    print(i.root)
    print(list(i.root.children))
```

matchers

```
from spacy.matcher import Matcher
```

```
x = spacy.load('en')
matcher = Matcher(x.vocab)
```

entities

```
matcher.add_entity(
    "GoogleNow",
)
```

patterns

```
from spacy.attrs import ORTH
```

```
matcher.has_entity(LOWER)
```

```
matcher.add_pattern(
    "GoogleNow",
    [{ORTH: "Google"},
    {ORTH: "Now"}],
    label=None
)
```

third party modules

sense2vec
displaCy
textacy
spacyr

Like the Cheatsheet?
Have Questions for Lesley?

Join our Slack Channel at
www.byteacademy2.slack.com
and get in touch with her at
@lesleycordero

WRITTEN BY

LESLEY CORDERO