

Web Semântica e Processamento de Linguagem Natural na prática

Newton Calegari



Processamento de Linguagem Natural

Python NLTK

Exercícios

Processamento de Linguagem Natural

Natural Language Processing

Python NLTK

```
pip install nltk
```

```
import nltk
```

```
nltk.download('popular')
```

- ▶ `from nltk.corpus import brown`
- ▶ `FreqDist`
- ▶ `text.concordance('word')`
- ▶ `text.similar('word')`
- ▶ `text.count('word')`
- ▶ `text.collocations('word')`
- ▶ `POS_TAG (from nltk.tag import pos_tag)`


```
▶ from nltk.corpus import machado, stopwords
▶ machado.fileids()
▶ dom_casmurro = machado.words('romance/marm08.txt')
▶ nltk.Text(dom_casmurro)
▶ FreqDist(dom_casmurro)
▶ freq.most_common(15)
▶ stopwords.words('portuguese')
```

```
▶ from url lib import request
  ▶ request.urlopen(url)
  ▶ response.read().decode('utf8')
▶ word_tokenize(raw)
▶ len(tokens), len(set(tokens))
▶ nltk.Text(tokens)
▶ text.collocations()
...

```

```
▶ from bs4 import BeautifulSoup
  ▶ request.urlopen(url).read().decode('utf8')
  ▶ BeautifulSoup(html).get_text()
▶ word_tokenize(raw)
...
```

```
▶ from nltk.corpus import wordnet
```

Part of Speech (pos_tag)

Tag	Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun

Tag	Description
PRP\$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, non3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Whdeterminer
WP	Whpronoun
WP\$	Possessive whpronoun
WRB	Whadverb