

Processamento de Linguagem Natural

Flávio Codeço Coelho
FGV-EMAp
(Dated: September 6, 2019)

CONTENTS

I. Contando Palavras

1

I. CONTANDO PALAVRAS

Neste capítulo iremos começar a interagir com os textos no nível da linguagem, por meio das ferramentas do Processamento de Linguagem Natural (PLN). Vamos progredir gradualmente nossa representação da linguagem a partir da morfologia, passando pela sintaxe e chegando à semântica.

Nesta etapa, faremos uso de bibliotecas especializadas em PLN como o [NLTK](#) e a [Spacy](#).

```
In [1]: import nltk
        import spacy
```

Bibliotecas de PLN requerem o carregamento de modelos de linguagem para funcionar de maneira apropriada: para este capítulo iremos carregar os modelos específicos da língua portuguesa. Para isso precisamos executar comandos no terminal do sistema operacional:

```
In [2]: !python3 -m spacy download pt
```

```
Requirement already satisfied: pt_core_news_sm==2.1.0 from
https://github.com/explosion/spacy-models/releases/download/pt_core_news_sm-2.1.0/pt_core_news_sm-2.1.0.tar.gz#egg=pt_core_news_sm==2.1.0 in /usr/local/lib/python3.6/dist-packages (2.1.0)
```

```
Download and installation successful
```

```
You can now load the model via spacy.load('pt_core_news_sm')
```

```
Couldn't link model to 'pt'
```

```
Creating a symlink in spacy/data failed. Make sure you have the required
permissions and try re-running the command as admin, or use a virtualenv. You
can still import the model as a module and call its load() method, or create the
symlink manually.
```

```
/usr/local/lib/python3.6/dist-packages/pt_core_news_sm -->
```

```
/usr/local/lib/python3.6/dist-packages/spacy/data/pt
```

```
Download successful but linking failed
```

```
Creating a shortcut link for 'pt' didn't work (maybe you don't have admin
permissions?), but you can still load the model via its full package name: nlp =
spacy.load('pt_core_news_sm')
```

```
In [4]: nltk.download()
```

```
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml
```

```
[4]: True
```

```
In [ ]:
```