

Trabalho Prático 1

Projecto integrado, Humanidades Digitais

Contents

1 Enunciados

1

Ligado à biblioteca de processamento de lingua natural **SPAcy**, há várias funções que poderão ser úteis para estes trabalhos.

```
import spacy
nlp=spacy.load('pt_core_news_md')
texto = ....
...
doc = nlp(texto)

for token in doc: .... ## token.pos_ lemma_
for frase in doc.sent: ....
for entidade in doc.ents: .... ## entidade.
```

Seguidamente propomos alguns enunciados de complexidade bastante diferentes. De entre eles, escolha alguns.

1 Enunciados

1. Lista das obras de camilo. Listar as obras de Camilo existentes na pasta `Obra/` do nosso repositório. Ou seja apresentar a lista dos ficheiros de texto existentes nessa pasta.
2. Título e autor de um livro Dado um livro (ficheiro de texto, ex: `Amor_de_predição.txt`) mostrar as duas primeiras frases.
3. Contar o número de verbos, e mostrar o respectivo lemma (realizado nas aulas).
4. Calcular a tabela de ocorrências dos verbos (lemma, número de ocorrências).
5. Quantos tokens tem um livro?, Quantas frases?
6. Calcular o comprimento médio das frase de um livro.
7. Calcular o comprimento médio das frase de todos os livros.
8. Tabela das entidades (Nomes de pessoas, lugares, ...) e respectivo número de ocorrências.
9. Calcular as 5 personagens / lugares que mais aparecem num livro.
10. Estender o anterior para calcular personagens em todos os livros.
11. Tomando um livro que tenha CAPÍTULOS, calcular lugares mais falados em cada capítulo.
12. Para algum dos exercícios anteriores, produzir saídas em formado html ou markdown(.md)