

# Processamento de Lingua Natural para a Eng. Biomédica

Teste, 2023-06-01

## 1 Python

Escreva uma função python que receba um dicionário EN-PT (palavra EN → lista de palavras PT) e calcule o dicionário inverso (palavra PT: palavra EN). Exemplo:

```
inverte( { dog : [ cão, cachorro ] , table : [mesa, tabela] ... })
```

deverá retornar o dicionário:

```
{ cão : dog, cachorro : dog, mesa : table, tabela : table ....}
```

## 2 Expressões regulares

De que é que as pessoas gostam?

Escrever uma função python que dado um texto devolva a lista das palavras que se seguem a **gostar de**, (mais precisamente “gost... de/do/da/dos/das X”).

Exemplo: *O Manel gosta de passear e sempre gostou da praia.* → [passear, praia]

## 3 Spacy

Pretende-se determinar os principais lugares dum texto. Para tal, crie um programa python que usando Spacy, procure as entidade (função ‘ents()’) geográficas (com e.label\_ GPE ou LOC)

- GPE - Countries, cities, states. ou
- LOC - Other locations, mountain ranges, bodies of water.

e imprima as 5 que aparecem mais vezes.

## 4 Jinja2

A seguinte estrutura de dados representa uma lista de doenças. Use a biblioteca Jinja2 para gerar dinamicamente uma página páginas HTML que representem todos os campos desta estrutura de forma adequada. Use um elemento HTML <table>.

Sugestão: P{{ list|join(' ') }}

```
[
  {
    "designation": "Fibrilhação auricular",
    "category": ["Diseases", "Cardiology", "Arrhythmias"],
    "desc": "Atrial fibrillation is a disease that ... " # This field is optional
    "en", "Atrial Fibrillation",
    "es": "Fibrilación auricular",
    "de": "Vorhofflimmern",

  },
  ...
]
```

## 5 Web Scrapping

Implemente um programa em Python que dado o URL de um plano de estudos de um curso, extraia o nome, descrição e créditos de todas as Unidades curriculares desse curso.

Tome como exemplo as páginas HTML apresentadas abaixo.

Exemplo de página do plano de estudos: (<https://www.uminho.pt/cursos/engbiomedica>)

```
<div class="disciplina-table">
  <table>
    <tr>
      <th>Semestre</th>
      <th>Disciplina</th>
      <th>ECTS</th>
    </tr>
    <tr>
      <td>1º Semestre</td>
      <td> <a href="https://www.uminho.pt/cursos/engbiomedica/biomecanica"> Biomecânica </a> </td>
      <td>10</td>
    </tr>
    <tr>
      <td>2º Semestre</td>
      <td> <a href="https://www.uminho.pt/cursos/engbiomedica/eletronicabiomedica"> Eletrônica Biomédica </a> </td>
      <td>8</td>
    </tr>
    (...)
  </table>
</div>
```

Exemplo de página de uma disciplina (<https://www.uminho.pt/cursos/engbiomedica/biomecanica>)

```
<div class ="container">
  Ano letivo 2022-2023
</div>
<div class="container">
  <div class="disciplina-title">
    <h1> Engenharia Biomédica </h1>
    <h2> <b> Biomecânica </b> </h2>
  </div>
  <div class="disciplina-desc">
    <p>
      A disciplina de Biomecânica estuda as forças e os movimentos do corpo humano (...)
    </p>
  </div>
</div>
```

O resultado deve ser gravado num ficheiro JSON.

Exemplo de saída esperada:

```
{
  "Biomecânica": { "ECTS": "10", "desc":"A disciplina de Biomecânica estuda ..." } ,
  "Eletrônica Biomédica": { "ECTS": "8", "desc":"A disciplina de ..." },
  (...)
}
```

## 6 Flask

Crie uma rota Flask que calcula o Índice de Massa Corporal (IMC). Esta rota deve receber como parâmetros o peso e altura de um indivíduo e devolver o IMC resultante. Deve ainda guardar o valor calculado num ficheiro, registando um histórico de IMCs calculados (uma entrada por linha).

$IMC = weight / (height * height)$ .