

TRABALHO PRÁTICO 2

PROCESSAMENTO DE LINGUAGEM NATURAL EM
ENGENHARIA BIOMÉDICA



Beatriz Rodrigues, PG53696

|

Bruno Machado, PG53709

|

Rúben Ganança, PG54203

ÍNDICE

01 Introdução

02 Enriquecimento do *Dataset*

- Correções nos ficheiros anteriores;
- Glossário obtido num *site*;
- Concatenação dos *datasets*;

03 Plataforma *Web*

- Funcionalidades da plataforma;
- Jinja

04 Conclusão

INTRODUÇÃO

Objetivo:

- Obter um glossário global a partir de todos os documentos *JSON*;
- A partir deste glossário desenvolver uma aplicação *Web* que permitisse navegar pelos dados, atualizar ou remover conceitos, ordenar os dados e filtrá-los;

ENRIQUECIMENTO DO *DATASET*

Correção dos *JSON* obtidos no Trabalho Prático 1:

- Transformação das designações dos conceitos para minúscula;
- Remoção de pontuação nas chaves (designações) dos conceitos;
- Remoção no *wipo.json* de sinónimos nas traduções dos conceitos.

```
"Traducoes": {  
  "AR": "ةمزلاتمةقئاضلاةيسفنتلاة د احلا",  
  "DE": "akutes Atemnotsyndrom des Erwachsenen, (syn.) ARDS",  
  "ES": "síndrome de dificultad respiratoria aguda, (syn.) SDRA",  
  "FR": "syndrome de détresse respiratoire aiguë, (syn.) SDRA",  
  "JA": "急性呼吸窮迫症候群, (syn.) 急性呼吸促迫症候群, ARDS",  
  "KO": "급성 호흡곤란 증후군, (syn.) ARDS",  
  "PT": "síndrome do desconforto respiratório agudo, (syn.) SDRA",  
  "RU": "острый респираторный дистресс-синдром, (syn.) ОРДС",  
  "ZH": "急性呼吸窘迫综合征, (syn.) ARDS"  
}
```

```
"Traducoes": {  
  "AR": "ةمزلاتمةقئاضلاةيسفنتلاة د احلا",  
  "DE": "akutes Atemnotsyndrom des Erwachsenen",  
  "ES": "síndrome de dificultad respiratoria aguda",  
  "FR": "syndrome de détresse respiratoire aiguë",  
  "JA": "急性呼吸窮迫症候群",  
  "KO": "급성 호흡곤란 증후군",  
  "PT": "síndrome do desconforto respiratório agudo",  
  "RU": "острый респираторный дистресс-синдром",  
  "ZH": "急性呼吸窘迫综合征"  
}
```

GLOSSÁRIO OBITIDO DE UM SITE

Glossário de termos técnicos provenientes do Centro Hospitalar de Lisboa Ocidental:

- Contêm informação relativa a termos médicos, nomeadamente, doenças, material médico ou conceitos médicos;
- Para retirar informação do site foi necessário recorrer à biblioteca *Beautiful Soup*.

Todas

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Termo	Definição Principal
<u>Abcesso</u>	Coleção de pus bem delimitada, desenvolvida numa cavidade neoformada.
<u>Ablação</u>	Extirpação cirúrgica de uma parte do corpo.
<u>Acalásia</u>	Impossibilidade de relaxar uma abertura ou esfíncter.
<u>Adenoidectomia</u>	Operação para ablação das adenoides.
<u>Admissão</u>	Ato da responsabilidade de um profissional de saúde credenciado que estabelece o ingresso de um indivíduo num estabelecimento de saúde ou programa de saúde, com o propósito de prestação de cuidados de saúde ou de cuidados continuados integrados.

</> Elementos

><div id="seog-glossary-description">...</div>

><div id="glossarysearch">...</div>

><div class="glossaryalphabet seopagination">...</div>

><table class="glossaryclear table " id="glossarylist">

><thead>...</thead>

><tbody>

><tr class="row0">

><td>

> Abcesso

></td>

><td>

><div>

><p>Coleção de pus bem delimitada, desenvolvida numa cavidade neoformada.</p>

></div>

><div class="tags"> </div>

></td>

></tr>

><tr class="row1">

><td>

> Ablação

></td>

><td>

><div> == \$0

><p>Extirpação cirúrgica de uma parte do corpo.</p>

></div>

><div class="tags"> </div>

></td>

></tr>

><tr class="row0">...</tr>

><tr class="row1">...</tr>

GLOSSÁRIO OBITIDO DE UM SITE

"abcesso": "coleção de pus bem delimitada, desenvolvida numa cavidade neoformada.",
"ablação": "extirpação cirúrgica de uma parte do corpo.",
"acalásia": "impossibilidade de relaxar uma abertura ou esfíncter.",
"adenoidectomia": "operação para ablação das adenoides.",
"admissão": "ato da responsabilidade de um profissional de saúde credenciado que estabelece o ingresso de um indivíduo num estabelecimento de saúde ou programa de saúde, com o propósito de prestação de cuidados de saúde ou de cuidados continuados integrados.",

CONCATENAÇÃO DOS *DATASETS*

Datasets concatenados:

- *conceitos.json*: Proveniente do documento “glossario_ministerio_saude.pdf” ;
- *termos_glossario.json*: Proveniente do documento “Glossario de Termos Medicos Tecnicos e Populares.pdf” ;
- *termos_anatomia.json*: Proveniente do documento “anatomia geral.pdf” ;
- *wipo_modificado.json*: Proveniente do documento “WIPOPearl_COVID-19_Glossary.pdf” ;
- *doencas_site.json*: Proveniente do documento *site* do Centro Hospitalar de Lisboa Ocidental.

De modo a não haverem chaves repetidas, sempre que um conceito existisse em mais do que um *dataset* a sua informação era acrescentada à chave já existente com o mesmo nome. Caso contrário era adicionada uma nova chave com a informação.

CONCATENAÇÃO DOS *DATASETS*

```
"eczema": {  
  "Categoria": "doenças",  
  "Descricao": "é uma doença inflamatória da pele caracterizada por eritema, edema, vesículas, escomas, crostas e liquenificação. a",  
  "Descricao 2": "dermatite; inflamação da pele",  
  "Descricao SNS": "doença cutânea manifestada por um processo inflamatório, superficial, com características e causas diversas, mu
```

```
"farmacoepidemiologia": {  
  "Categoria": "medicamentos, vacinas e insumos",  
  "Descricao": "aplicação do método e raciocínio epidemiológico no estudo dos efeitos, benéficos e adversos, e do uso de medicamentos",  
  "Traducoes": {  
    "AR": "ةلديصة يئابو",  
    "DE": "Pharmakoepidemiologie",  
    "ES": "farmacoepidemiología",  
    "FR": "pharmaco-épidémiologie",  
    "JA": "薬剤疫学",  
    "KO": "약물역학",  
    "PT": "farmacoepidemiologia",  
    "RU": "фармакоэпидемиология",  
    "ZH": "药物流行病学"  
  }  
}
```


CONCATENAÇÃO DOS *DATASETS*

Contudo, de forma a haver traduções para conceitos, mesmo para aqueles cujo nome não é exatamente o mesmo que o da tradução, foi usado um modelo do *Hugging Face* (BioBERT) próprio para termos biomédicos que verifique a similaridade entre palavras e acrescente uma tradução à palavra se ela for mais do que 95% similar à original.

```
"retenção": {  
  "Categoria": "Sem Categoria",  
  "Descricao": "conservação, dentro do corpo, de materiais ou líquidos que se excretam normalmente",  
  "Traducoes Relacionadas": {  
    "AR": "ءاوتجا",  
    "DE": "Eindämmung",  
    "ES": "contención",  
    "FR": "endiguement",  
    "JA": "封じ込め",  
    "KO": "전파차단",  
    "PT": "retenção",  
    "RU": "сдерживание",  
    "ZH": "围堵"  
  }  
}
```

APLICAÇÃO WEB - JINJA

- Criado um documento chamado *parent.html* que atua como *template* base para toda a aplicação. Define uma estrutura *HTML* que as outras páginas herdarão, facilitando a manutenção e a consistência da aplicação;
- Permite a existência de ciclos *for* e a utilização de lógica condicional diretamente nos *templates*, para poder iterar sobre conjuntos de dados e ocultar ou mostrar informação dependendo da condição lógica;
- Foram também usados *forms* para poder alterar ou atualizar informação dentro dos *datasets* (usando o método “*POST*”) e foi ainda usado *CSS* e a framework *Bootstrap* para dar estilo à página.

APLICAÇÃO *WEB*

Demonstração das Funcionalidades da
Aplicação

CONCLUSÃO

- Com este trabalho foi possível verificar a flexibilidade e utilidade tanto do *Flask* como do *Jinja* e aplicá-los a ambos para fazer uma aplicação *web* que fosse dinâmica e permitisse manipular os dados de um glossário de termos médicos/biomédicos;
- **Trabalho Futuro:** utilizar uma API da wikipédia que ao encontrar uma página com o nome do conceito iria adicionar o seu link ao mesmo, permitindo a sua consulta.