

Universidade do Minho Escola de Engenharia

Trabalho Prático 2

Processamento de Linguagem Natural em Engenharia Biomédica Mestrado em Informática Médica Ano Letivo 2023/2024

Alunos:

Madalena Passos, pg54023 Mariana Ribeiro, pg54061 Mariana Almeida, pg54062

Docentes: Luís Filipe da Costa Cunha e José João Almeida.

Conteúdo

1 Introdução								
2	2 Informação Recolhida							
3	Funcionalidades da plataforma							
	3.1	ário Ministério da Saúde	4					
		3.1.1	Estrutura dos Dados	4				
		3.1.2	Rotas do Flask e Templates Jinja2	6				
3.2 Anatomia na Prática								
		3.2.1	Estrutura dos dados	10				
		3.2.2	Rotas Flask e Templates Jinja2	11				
	3.3	Minid	icionário de Cardiologia	15				
		3.3.1	Estrutura dos Dados	15				
		3.3.2	Rotas Flask e Templates Jinja2	16				
4	4 Conclusão							
5	5 Bibliografia							

1 Introdução

No âmbito da unidade curricular de Processamento de Linguagem Natural em Engenharia Biomédica, foi proposto o desenvolvimento de um sistema capaz de utilizar dados em formato JSON extraídos de PDFs médicos.

Para a realização deste projeto, foram escolhidos três dos quatro dicionários gerados no 1º trabalho prático: Anatomia na Prática, Glossário Ministério da Saúde e Minidicionário de Cardiologia. Estes dicionários foram integrados num único ficheiro JSON, o qual foi posteriormente enriquecido através de técnicas de web scraping para inclusão de notícias relacionadas com o Ministério da Saúde e pelo uso do Google Translator, visando adicionar mais traduções ao dicionário de Cardiologia.

Após a integração e o enriquecimento dos dados, foi utilizado o framework Flask em conjunto com o template Jinja2 para criar as rotas necessárias e a interface da página web. Este sistema permitiu a manipulação e visualização dos dados, explorando as relações entre os conceitos e categorias presentes. Com efeito, a interface desenvolvida possibilitou a navegação pelas categorias e descritores do Ministério da Saúde, pelos conceitos básicos anatómicos do Sistema Musculoesquelético e pelos conceitos e respetivas traduções da área de Cardiologia. Além disso, na secção de Cardiologia, a plataforma permite ainda a adição de novos conceitos e a remoção de entradas existentes, garantindo que qualquer alteração seja preservada após reinicializações do sistema.

Em suma, o projeto incidiu na criação de uma plataforma robusta que não apenas organiza dados médicos, mas também facilita a exploração de relações entre os conceitos, proporcionando uma ferramenta útil para a manipulação e visualização de informações num contexto biomédico.

2 Informação Recolhida

No seguimento do primeiro trabalho prático, foram utilizados três documentos pdf cuja informação foi recolhida, organizada e armazenada num só ficheiro JSON compilado. Neste ficheiro estão presentes informações proveniente de documentos relativos a:

- Categorias do Ministério da Saúde, incluindo as suas definições e descritores;
- Legendas de estruturas musculares e ósseas do corpo humano;
- Traduções inglês-português e português-inglês de termos de cardiologia.

Porém, para tornar a plataforma mais abrangente e completa, foram adicionadas novas informações.

Primeiramente, foi realizado o scraping de notícias atualizadas relativas às categorias do Ministério da Saúde, utilizando como fonte o website Atlas da Saúde [1] e a biblioteca BeautifulSoup. Este processo envolveu a extração de secções específicas, relativas ao título, introdução e link de cada notícia, permitindo assim que os utilizadores da plataforma tenham acesso a informações recentes e contextualizadas sobre as políticas e iniciativas do Ministério da Saúde.

Adicionalmente, ampliou-se o leque de traduções de termos de cardiologia, incluindo agora traduções para o francês e espanhol, além das já existentes em inglês e português. Esta inclusão visa facilitar a comunicação e compreensão dos termos técnicos de cardiologia em contextos multilíngues.

3 Funcionalidades da plataforma

Nesta secção, serão explicadas as funcionalidades da plataforma, associadas às funções Python, com auxílio da biblioteca Flask, desenvolvidas para o efeito. É importante realçar que as funções em Python estão associadas a rotas e a templates HTML. Estas páginas HTML dinâmicas foram implementadas através das $templates\ Jinja2$.

Foi então desenvolvida a página *layout.html* com informações do *header* e *footer* e extendido o *layout* destes elementos nas restantes páginas *HTML*, adicionando os diferentes componentes de forma a apresentar as diferentes informações.

A título de exemplo, mostra-se na Figura 1, a página home.html onde são apresentados os temas da plataforma, nomeadamente, a visualização das terminologias das categorias do ministério da saúde, bem como pesquisa e notícias relacionadas, visualização e identificação de estruturas anatómicas e, por fim, a tradução de termos relacionados a cardiologia em várias

línguas.

É de notar que esta página apresenta o header, isto é, a navbar e o footer, provenientes do layout.html.



Figura 1: Página home.html

3.1 Glossário Ministério da Saúde

3.1.1 Estrutura dos Dados

A seccção do ficheiro JSON relativa ao Ministério da Saúde está organizada de forma hierárquica, contendo categorias, descritores, descrições e notícias associadas. A estrutura dos dados é a seguinte:

```
"Descritor2": {
                     "descricao": "Descrição do Descritor2",
                     "assuntos": [
                         ["Notícia3", "URL3", "Resumo3"]
                     ]
                }
            }
        },
        "Categoria2": {
            "descrição categoria": "Descrição da Categoria2",
            "descritores categoria": {
                "Descritor3": {
                     "descricao": "Descrição do Descritor3",
                     "assuntos": [
                         ["Notícia4", "URL4", "Resumo4"]
                     ]
                }
            }
        }
    },
}
```

Ministério da Saúde: No JSON gerado, a chave "Ministério da Sáude" contém todas as categorias a ele relacionadas. Estas categorias referem-se a áreas específicas de interesse ou a funções dentro do Ministério da Saúde.

Categoria: Cada categoria contém uma breve descrição, bem como um conjunto de subtemas designados por descritores.

Descrição da Categoria: Breve descrição que fornece uma visão geral da categoria.

Descritores da Categoria: Dicionário onde cada chave é o nome de um descritor e o valor é um dicionário com a descrição do descritor e uma lista de assuntos (notícias) relacionados.

Descrição do Descritor: Breve descrição que fornece uma descrição detalhada

do descritor.

Assuntos: Lista de listas, onde cada sublista contém uma notícia representada por um título, um link e um resumo da notícia.

3.1.2 Rotas do Flask e Templates Jinja2

Para o desenvolvimento do sistema web, foram implementadas várias rotas utilizando o Flask, de forma a ser possível a navegação e manipulação dos dados.

Listar Categorias

A rota /ministerio/categorias é responsável por recuperar todas as categorias relacionadas ao Ministério da Saúde a partir do ficheiro JSON. Essas categorias são então enviadas para um template HTML designado categorias.html. Neste template, as categorias são apresentadas em forma de lista, permitindo ao utilizador visualizar todas as opções disponíveis.

Para além disso, foi utilizada ainda a *framework Bootstrap* para estilizar a página e uma barra de pesquisa para facilitar a localização das categorias. A filtragem dinâmica de categorias foi possível graças ao uso de *Javascript*.

Através da Figura 2, é possível observar o resultado da página renderizada.

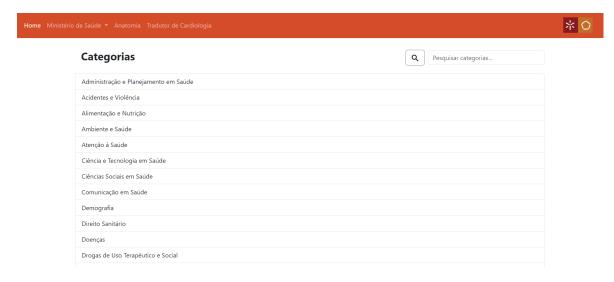


Figura 2: Página categorias.html

Consultar Categoria

Após a seleção de uma determinada categoria, a rota /ministerio/categorias/<cat> obtém a descrição da categoria e a lista de descritores a ela associados. Estas informações são enviadas para o template HTML descrição.html, que demonstra de forma estruturada a descrição da categoria e os descritores associados.

A Figura 3 apresenta a página renderizada resultante.



Figura 3: Página descricao.html

Consultar Descritor

Com a seleção de um determinado descritor na página anterior, a rota /ministerio/categorias /<cat> /<cat> /<descritor> irá recuperar a descrição do respetivo descritor e as notícias a ele relacionadas. Mais uma vez, estas informações são enviadas para um template HTML, descritor.html, que apresenta a descrição do descritor e a lista de notícias associadas.Caso o descritor ou a categoria não forem encontrados no dicionário JSON, é mostrada uma mensagem de erro no template erro.html.

Adicionalmente, o template descritor.html possui um botão para voltar à lista de categorias, um text-card que contém a descrição do descritor, e uma secção de notícias dividida em duas colunas, onde cada notícia é representada por um título, um breve resumo e um link "Ver mais..."que abre a notícia completa num novo separador.

A respetiva página renderizada está apresentada na Figura 4.



Figura 4: Página descritor.html.

Listar Notícias

Para além das funcionalidades descritas anteriormente, caso o utilizador pretenda apenas consultar notícias relacionadas ao Ministério da Saúde, pode optar por clicar na opção "Notícias" na NavBar dentro da secção Ministério da Saúde. Para tal, foi criada uma rota que lista todas as notícias disponíveis resultantes de web scraping na página web do Atlas da Saúde. Esta rota, acessível em /ministerio/noticias, recupera todas as notícias relacionadas ao Ministério da Saúde.

Além disso, para facilitar a navegação, são apresentadas funcionalidades de procura e paginação. O utilizador pode pesquisar por termos específicos nas notícias, e os resultados são retornados com os termos encontrados destacados a amarelo, tanto nos títulos quanto nos resumos das notícias. Por outro lado, a paginação divide os resultados em páginas, com 10 notícias por página, permitindo uma navegação mais eficiente através dos resultados obtidos. Os resultados são renderizados no template *noticias.html*, cuja página renderizada pode ser observada na Figura 5.

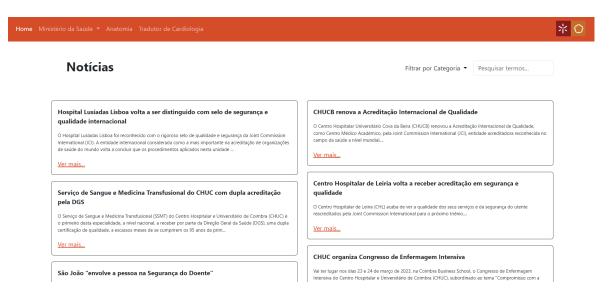


Figura 5: Página noticias.html.

Listar Notícias por Categoria

A rota /ministerio/noticias/<categoria> permite listar todas as notícias relacionadas a uma determinada categoria do Ministério da Saúde. Quando uma categoria é selecionada, são apresentadas as notícias associadas aos descritores dessa categoria no template HTML noticias cat.html.

A interface apresenta um menu dropdown para filtrar notícias por categoria, um botão para visualizar todas as notícias e um campo de pesquisa que destaca os termos encontrados.

Adicionalmente, as notícias são divididas em duas colunas para uma melhor organização, sendo que cada uma é apresentada num text-card onde é incluído o título, o resumo e o link da notícia. A respetiva página renderizada está apresentada na Figura 6.

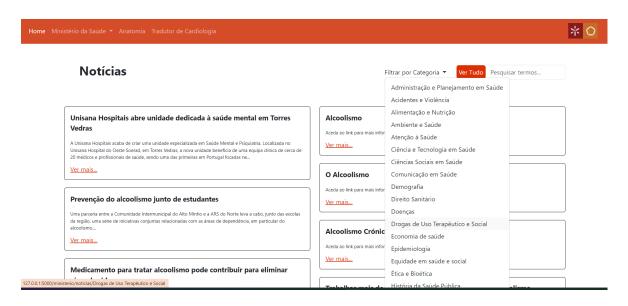


Figura 6: Página noticias_cat.html

3.2 Anatomia na Prática

3.2.1 Estrutura dos dados

A secção do ficheiro *JSON* gerado relativa à anatomia, apresenta uma estrutura idêntica à conseguida no trabalho prático anterior. Foi apenas necessário criar uma chave "Anatomia" à qual está associado o *JSON* com as informações do ficheiro Anatomia na Prática, como se apresenta no esquema a seguir.

Como se pode observar existem 2 temas principais no ficheiro relativos ao Sistema Es-

quelético e Articular e o Sistema Muscular. Posteriormente, cada um destes temas apresenta subtemas e sub-subtemas. Desta forma, o dicionário está organizado hierarquicamente desde os temas até às legendas das estrutras anatómicas respetivas.

3.2.2 Rotas Flask e Templates Jinja2

Para o utilizador poder visualizar os dados do ficheiro Anatomia na Prática de forma organizada, desenvolveu-se um conjunto de funções em *Python e Flask* e templates *HTML* que permite ao utilizador navegar nos temas e subtemas, visualizando as estrutras anatómicas pretendidas.

Consultar temas

Foi definida a rota "/anatomia" que quando acedida executa a função consultar_temas_ossos. Esta função obtém os temas, isto é, as chaves, da secção anatomia do dicionário json_all. Os temas são adicionados a uma lista que é passada na renderização da página anatomia.html. A página anatomia.html apresenta os temas em forma de botão, de forma a permitir ao utilizar escolher o tema que pretende visualizar. É de notar que foram também adicionadas figuras que, assim como os botões, têm redireccionamento para URLs específicas de cada sistema anatómico.



Figura 7: Página anatomia.html

Consultar subtemas

A rota "/anatomia/<tema>" foi definida para poder ser executada a função consultar_subtemas_ossos. A função recebe um tema como parâmetro do link URL, verificando se esse tema se encontra no dicionário relativo à anatomia, json all["Anatomia"]. Em caso posi-

tivo, obtém os subtemas associados a esse tema e renderiza a template anatomia_tema.html, passando o tema e subtemas respetivos. A template desenvolvida, apresenta os subtemas em forma de botão para o utilizador poder escolher qual subtema pretende visualizar. Na figura 8, apresenta-se a página "anatomia/SISTEMA ESQUELÉTICO E ARTICULAR" que mostra os 4 subtemas deste tema. A página relativa aos subtemas do sistema Muscular é idêntica, mostrando, no entanto, os subtemas respetivos.

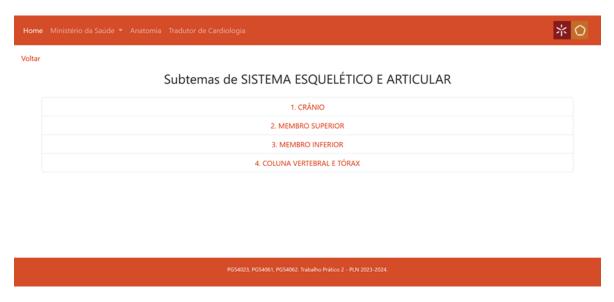


Figura 8: Página anatomia_tema.html

É de notar que foi adicionado um botão "Voltar" que permite ao utilizador retroceder na sua pesquisa.

Consultar sub-subtemas

A rota "/anatomia/<tema>/<subtema>" define um endpoint que executa a função consultar_subsubtemas_ossos, que recebe como parâmetros da URL o tema e subtema. Após a verificação da existência do tema no dicionário json_all["Anatomia"], é verificado se o subtema passado como parâmetro está presente nas chaves no dicionário relativo ao tema. Em caso positivo, são obtidos os subsubtemas e a introdução do subtema, e renderiza o template anatomia_subtema.html, passando os subsubtemas, subtema, tema e introdução. Como se pode observar na Figura 9, são apresentadas a introdução e também os subtemas de "2. MEMBRO SUPERIOR".



Figura 9: Página anatomia subtema.html

Consultar Sub-sub-subtemas

Por fim, foi definida rota "/anatomia/<tema>/<subtema> /<subsub>" define um endpoint que executa a função consultar_sub_sub_sub_temas_ossos. De forma a apresentar as figuras e legendas de cada sub-sub-subtema, a função recebe três parâmetros tema, subtema e sub-subtema, verificando a existência dos mesmos nos dicionários respetivos. Como mostrado na estrutura dos dados na secção Estrutura dos Dados, cada sub-subtema apresenta as legendas de uma imagem. Para as poder recuperar e enviar para a template posteriormente, estas são acedidas através das chaves dos dicionários, json_all["Anatomia"][tema][subtema][subsub], e guardadas numa variável. Para poder ser apresentada a opção de navegar para o sub-subtema anterior e seguinte ao sub-subtema apresentado é necessário obter uma lista com todos os sub-subtemas. Posteriormente, identificar o índice do sub-subtema que se pretende visualizar e, consequentemente, determinar o sub-subtema anterior e seguinte, se estes existirem. Por fim, para garantir que a imagem associada ao sub-subtema seja apresentada corretamente, o título do sub-subtema passa por um processo de limpeza, ajustando-o para corresponder ao nome do arquivo da respetiva imagem. É então renderizada a template e passadas as variáveis relativas às legendas, sub-subtemas anterior e seguinte e, por fim, o nome da imagem.



Figura 10: Página anatomia legendas.html

Como se pode verificar pela figura 10, a template anatomia_legendas.html mostra a figura relativa à Vista Inferior da Clavícula Direita e a legenda das suas diferentes partes. Apresentam-se também os botões "Anterior" e "Seguinte" que permitem navegar intuitivamente entre os sub-subtemas. Por outro lado, o botão "Voltar" permite retroceder à página anatomia_subtema.html.

Pesquisar Termo

De forma a poder ser pesquisado um termo anatómico em todo o conjunto de dados do ficheiro Anatomia na Prática, foi desenvolvida a rota "/pesquisa_anatomia" que executa a função pesquisar_termo_anatomia que recebe como parâmetros um termo para pesquisa. Esta função chama uma outra função auxiliar que procura recursivamente nos vários dicionários o termo pesquisado, adicionando os resultados relevantes a uma lista.

A lista total de resultados encontrados é passada à template pesquisa_ossos.html renderizada.

Uma outra função auxiliar utilizada é a função construir_link, que constrói um link direto para a secção onde o termo foi encontrado. Esta função é utilizada diretamente na template pesquisa_ossos.html.



Figura 11: Procura de termos anatómicos.

3.3 Minidicionário de Cardiologia

3.3.1 Estrutura dos Dados

Relativamente à secção do ficheiro JSON responsável pelos dados dos termos de cardiologia e suas traduções, este encontra-se organizado do seguinte modo:

Pode-se verificar que existem quatro dicionários distintos, referentes aos pares de idiomas português-inglês, inglês-português, português-espanhol e português-francês, incluídos na chave Cardiologia do ficheiro JSON.

Em cada um destes dicionários estão presentes pares chave, valor, onde o primeiro é o termo de cardiologia a traduzir e o segundo a tradução do mesmo.

3.3.2 Rotas Flask e Templates Jinja2

Á semelhança dos documentos referidos anteriormente, para os conceitos de cardiologia, foram definidas rotas permitindo o acesso a esta informação, a inserção e a deleção de conceitos.

Listar Conceitos

Na rota /tradutor_cardiologia é apresentada uma visão geral dos termos de cardiologia e das suas traduções compilados numa tabela de leitura fácil, podendo, também, ser seleci-

onados os idiomas envolvidos na tradução entre os pares português-inglês, inglês-português, português-francês e português-espanhol 12.

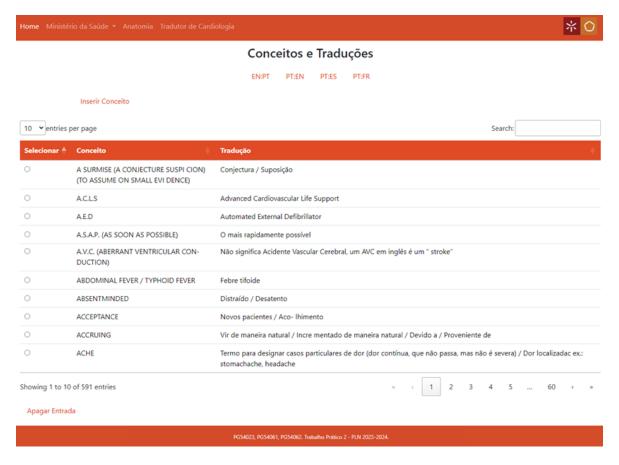


Figura 12: Página cardiologia.html

De modo a obter este resultado, aquando do acesso a esta secção, os dados relativos aos termos de cardiologia são extraídos do arquivo JSON compilado, determinando-se dinamicamente o idioma selecionado pelo utilizador. Utilizando este idioma como referência, a função renderiza uma página HTML, cardiologia.html, que exibe os conceitos juntamente com a sua respetiva tradução de forma clara e organizada.

Adicionar Conceitos

Relativamente à consulta de termos de cardiologia, era fundamental oferecer aos utilizadores a capacidade de expandir a base de dados da aplicação. Tal foi implementado recorrendo à rota /adicionar_conceitos.

Quando os utilizadores desejam incluir um novo termo de cardiologia e a sua tradução, eles preenchem um formulário fornecendo as informações necessárias 13. Após o envio do formulário, o sistema valida os dados recebidos, garantindo que todos os campos obrigatórios

estejam preenchidos.

	Conceitos e Traduções								
	EN:PT	PT:EN	PT:ES	PT:FR					
Fechar									
Adicionar Conceito	5								
Conceito									
Tradução									
Idioma					_				
EN:PT									
EN:PT									
PT:EN									
PT:ES PT:FR									

Figura 13: Formulário de adição de conceitos de cardiologia.

Por fim, o novo termo é adicionado ao dicionário correspondente ao par de idiomas selecionado no arquivo JSON compilado. Esta função desempenha um papel essencial na manutenção e atualização contínua da base de dados, garantindo que ela permaneça relevante e abrangente.

Apagar Conceitos

Na seção de remoção de conceitos, presente na rota /delete_entry, quando um termo é selecionado para exclusão, é realizada uma série de operações para garantir que a remoção seja feita de forma precisa e segura.

Primeiro, o termo selecionado e seu idioma de tradução são identificados a partir dos dados do formulário POST. Em seguida, é criada uma cópia do arquivo JSON compilado de modo a preservar os dados existentes 14.

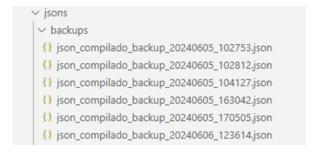


Figura 14: Cópias de segurança das edições efetuadas à base de dados.

Posteriormente, o termo selecionado é removido do dicionário correspondente no arquivo

JSON, assegurando que a base de dados permaneça consistente e atualizada. Para que se possa escolher ao certo os termos a eliminar, foi adicionada na tabela uma coluna contendo botões de seleção, que conectados à funcionalidade de deleção permitem a implementação do processo referido anteriormente 15.



Figura 15: Remoção de termos de cardiologia: antes (A) e depois (B).

4 Conclusão

O projeto implementado utilizando *Flask* e *Jinja2* proporcionou a criação de uma plataforma com diferentes funcionalidades e interativa para os utilizadores.

Ao explorar os dados do Ministério da Saúde, Anatomia na Prática e Dicionário de Cardiologia contidos no JSON compilado, foi permitido aos utilizadores não só visualizar informações relevantes, mas também interagir de forma fácil com a plataforma.

A capacidade de visualizar conceitos, aceder a notícias, imagens e legendas anatómicas, além de adicionar, eliminar e traduzir conceitos em múltiplos idiomas, mostrou significativamente a utilidade e a acessibilidade da ferramenta desenvolvida. Por outro lado, a utilização de *Flask* e *Jinja2* permitiu que todo este processo fosse facilitado e mais interativo.

5 Bibliografia

[1] Atlas da Saúde. (s.d.). Atlas da Saúde. Acedido em: https://www.atlasdasaude.pt/