Mestrado em Engenharia Biomédica - Processamento de Linguagem Natural em Engenharia Biomédica

TP2 - GLOSSÁRIO MÉDICO

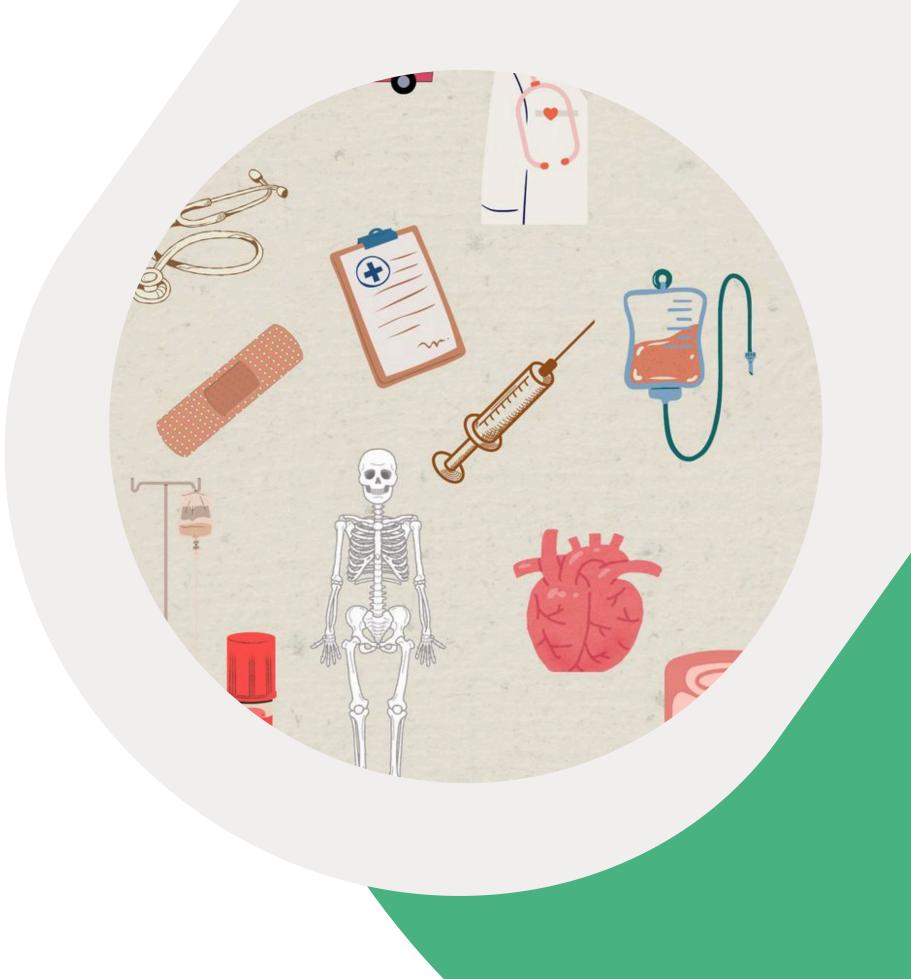
Processamento de Linguagem Natural em Engenharia Biomédica

Realizado por:

Ana Beatriz Macedo , pg55820 André Sousa, pg52564 Maria Vale , pg56144

Docentes:

Luís Filipe Costa Cunha José João Antunes Guimarães Dias Almeida



Melhoria da extração de conteúdo do dic Multilingue

Correção conceitos com definição "veg"

Extração das notas

ACA

veg. assaig aleatoritzat...

Ver detalhes →

ADG20

Fàrmac en investigació, amb efecte neutralitzant i antivíric, que s'uneix

Ver detalhes →

<u>ADN</u>

veg. àcid desoxiribonucleic...

Ver detalhes →

```
"cas": "2516243-50-0",

"area_tematica": "Principis actius",

"definicao": "Fàrmac en investigació, amb efecte neutralitzant i antivíric, que s'uneix a la superfície de la proteïna S del SARS-CoV-2 per a evitar que el virus in "nota": [

"L'ADG20 és un anticòs monoclonal que està en procés d'avaluació com a possible tractament per a la COVID-19. És d'origen biotecnològic.",

"1. La sigla ARN té un ús divulgatiu, mentre que en àmbits especialitzats se sol utilitzar la sigla anglesa RNA Aquesta recomanació és aplicable també a la parell

"2. La sigla RNA correspon a l'anglès ribonucleic acid ( 'àcid ribonucleic ')."
```

Construção de um Ficheiro json geral

O ficheiro JSON final, foi enriquecido com dados do TP1 e dados externos, estruturado em 5 grandes blocos: Abreviaturas, Anexos, Categorias, Conceitos e Siglas.

Os conceitos agregam traduções, definições, categorias, sinónimos e informação adicional (notas)

O ficheiro json final contêm informação dos seguintes documentos:

- Dicionário Multilingue COVID-19
- GlossárioNeo
- Glossário do Ministério da Saúde
- Harvard Health Publishing
- HGIS

Priorizámos as traduções portuguesas. Sempre que existia uma tradução em português para um termo estrangeiro, esse termo foi integrado como tradução do conceito em português correspondente, em vez de ser tratado como um novo conceito independente.

No final, temos um JSON unificado, que foi usado pela nossa aplicação web para visualização, edição e navegação de dados médicos.

Construção de um Ficheiro json geral

```
'ABREVS": {
   "ABEn": [
      "Associação Brasileira de Enfermagem"
   "BD-SIA/SUS": [
      "Banco de Dados Nacional do Sistema de Info
   "Bilat": [
      "Bilateral"
   "CAS": [
      "número CAS"
   "CN-DST/AIDS": [
      "Coordenação Nacional de Doenças Sexualment
   ],
   "CPqGM": [
      "Centro de Pesquisa Goncalo Moniz"
   "CPqHEC": [
      "Centro de Pesquisa Hospital Evandro Chagas
   "CPqRR":
      "Centro de Pesquisa René Rachou"
   "CRN-Aids": [
      "Centros de Referência Nacional de Aids"
```

```
"ANEXOS": [
       "Ano da Edição": "2001",
       "Mês da Edição": "setembro",
       "Número": "01",
       "Número da Edição": "68",
       "Título do Artigo": "A base da diferenca"
       "Ano da Edição": "2000",
       "Mês da Edição": "abril",
       "Número": "02",
       "Número da Edição": "52",
       "Título do Artigo": "A corrida pelo mapeamento"
       "Ano da Edição": "2004",
       "Mês da Edição": "julho",
       "Número": "03",
       "Número da Edição": "101",
       "Título do Artigo": "A favor da mulher"
       "Ano da Edição": "2003",
       "Mês da Edição": "junho",
       "Número": "04",
       "Número da Edição": "88",
       "Título do Artigo": "A flor da pele"
```

```
"CATEGORIAS": {
    "Acidentes e Violência": {
       "definicao": "Refere-se ao conjunto
       "subcategorias": [
            "Abuso financeiro aos idosos",
            "Abuso incestuoso",
            "Abuso sexual na infância",
            "Acidentes ampliadoscidentes de
            "Acidentes de trabalho fatais",
            "Acidentes de trabalho graves",
            "Acidentes de trânsito",
            "Acidentes de transporte",
            "Autonegligência",
            "Casas de apoio para mulheres em
            "situação de risco",
            "Comunicação de acidente do trab
            "Conselho tutelar",
            "Conselhos de defesa dos direito
            "criança e do adolescente",
            "Conselhos dos direitos da mulher
            "Delegacia",
            "Estupro",
            "Informação tóxico-farmacológica
```

```
"CONCEITOS": {
   "ACA": {
       "CAS": null,
       "artigos": [],
       "categoria area": [],
       "categoria lexica": [],
       "definicoes": [
                "veg. assaig aleatoritzat",
               "Dicionario Multilingue"
       "info enc": null,
       "sigla": null,
       "sinonimos": {}.
       "traducoes": {
           "ca": [
               "ACA"
    "ADG20": {
       "CAS": "2516243-50-0",
       "artigos": [],
       "categoria area": [
           "Principis actius"
       "categoria_lexica": [],
       "definicoes": [
               "Fàrmac en investigació, amb efecte neutral:
```

```
"SIGLAS": {
   "A": "Aborto",
   "A/C": "Assisto Controlado",
   "A/E": "A esclarecer",
   "AA": "Pode significar abdômen ag
   "AB": "Atenção Básica",
   "ABD": "Abdômen",
   "AC": "Ausculta Cardíaca",
   "ACM": "A Critério Médico",
   "ACV": "Ausculta Cardiovascular"
   "AD": "Água Destilada",
   "ADA": "Acrômio Direita Anterior
   "ADM": "Administração",
   "ADP": "Acrômio Direito Posterior
   "ADT": "Assistência Domiciliar Te
   "AE": "A esclarecer",
   "AEA": "Acrômio Direita Anterior 4
   "AEP": "Acrômio Direito Posterior
   "AF": "Água Filtrada",
```

Fonte 1: HGIS - Abreviações e Siglas

https://hgis.org.br/abreviacoes-e-siglas-padronizadas-para-registros-do-prontuario/



Figura 1 - Página de abreviações e siglas padronizadas do HGIS.

Fonte 1: HGIS - Abreviações e Siglas

https://hgis.org.br/abreviacoes-e-siglas-padronizadas-para-registros-do-prontuario/

- Grupo I.I - Diagnósticos

Abreviações / Siglas	Significado
A	Aborto
ADA ou AEA	Acrômio Direita Anterior ou Esquerda Anterior
ADP ou AEP	Acrômio Direito Posterior ou Esquerdo Posterior
AIG	Adequado para a idade gestacional
AIT ou TIA	Ataque isquêmico transitório (cerebral)
ANN	Anóxia Neonatal

Figura 1 - Página de abreviações e siglas padronizadas do HGIS.

Fonte 1: HGIS - Abreviações e Siglas

https://hgis.org.br/abreviacoes-e-siglas-padronizadas-para-registros-do-prontuario/

```
▼<div class="elementor-element elementor-element-5cd32ab elementor-widget elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elementor-elemento
    ntor-widget-accordion" data-id="5cd32ab" data-element type="widget" data-
    widget type="accordion.default">
      ▼<div class="elementor-widget-container">
            ▼<div class="elementor-accordion">

♦ <div class="elementor-accordion-item"> ( </div > )
                   / <div class="elementor-accordion-item"> ... </div>
                   ▶ <div class="elementor-accordion-item"> ... </div>
                   / <div class="elementor-accordion-item">()</div>
                   / <div class="elementor-accordion-item"> ... </div>
                   /div class="elementor-accordion-item">()</div>
                   ▶ <div class="elementor-accordion-item"> ··· </div>
                   ▶ <div class="elementor-accordion-item"> ··· </div>
                   ▶ <div class="elementor-accordion-item"> ··· </div>
                 </div>
          </div>
    </div>
```

```
<div class="elementor-accordion-item"> == $0
\div id="elementor-tab-title-9731" class="elementor-tab-title" data-
 tab="1" role="button" aria-controls="elementor-tab-content-9731" aria-
 expanded="false" tabindex="-1" aria-selected="false"> --- </div>
▼<div id="elementor-tab-content-9731" class="elementor-tab-content eleme
 ntor-clearfix" data-tab="1" role="region" aria-labelledby="elementor-ta
 b-title-9731" style="display: none;" hidden="hidden">
 *
    *
     ▼
       <strong>Abreviações / Siglas</strong>
     ▼
       <strong>Significado</strong>
      ▼
     ▼
       <strong>A</strong>
      Aborto
    \tr class="FundoTabelaCinza">\cdots \( \lambda / \tr> \)
    \tr class="FundoTabelaCinza"> \cdots \langle /tr>

... 

    \...
    \tr class="FundoTabelaCinza"> \cdots \langle /tr>
    ▶  ··· 
    \tr class="FundoTabelaCinza">\cdot \langle /tr>
    ▶  ···
```

Fonte 1: HGIS - Abreviações e Siglas

1. LEITURA DO FICHEIRO 'GLOSSARIO_FINAL_COMPLETO.JSON'

2. EXTRAÇÃO DO CONTEÚDO

```
url == "https://hgis.org.br/abreviacoes-e-siglas-padronizadas-para-registros-do-prontuario/"
response == requests.get(url)
soup == BeautifulSoup(response.content, 'html.parser')

# Encontrar todos os grupos (accordions)
accordion_items == soup.select('.elementor-accordion-item')

* Adiv class="elementor-accordion-item"> (div) (di
```

Fonte 1: HGIS - Abreviações e Siglas

3. FILTRAGEM E NORMALIZAÇÃO

```
for item in accordion items:
*** # Verificar se é o grupo de símbolos
title = item.select one('.elementor-accordion-title')
                                                            <a class="elementor-accordion-title" tabindex="0">Grupo I.X -
if title and "Simbolos" in title.get_text(strip=True):
                                                            Símbolos</a> == $0
print("Saltar grupo de símbolos")
----continue
# Encontrar a tabela dentro do item
table = item.select one('table')
                                                         ▼
if not table:
                                                           \ \cdots </ra>
 continue
                                                           \\cdots 
*** # Processar cada linha da tabela (ignorando cabeçalhos)
 rows = table.select('tr:not(.FundoTabelaAzul)')
for row in rows:
cells = row.select('td')
if len(cells) < 2:
                                                         ▼
                                                          ▼
··· continue
                                                            <strong>A</strong>
abbr = cells[0].get text(strip=True)
meaning = cells[1].get text(strip=True)
                                                           Aborto
····# Pular entradas vazias ou sem significado
if not abbr or not meaning:
 continue
----#-Remover-asteriscos-e-espaços-extras
abbr = abbr.replace('*', '').strip()
# Processar múltiplas siglas (ex: "ADA ou AEA")
for single abbr in [a.strip() for a in abbr.split(' ou ') if a.strip()]:
```

Fonte 1: HGIS - Abreviações e Siglas

4. CLASSIFICAÇÃO E INSERÇÃO

```
#*Processar*múltiplas*siglas*(ex: "ADA ou AEA")

for single_abbr in [a.strip() for a in abbr.split('ou') if a.strip()]:

#*Classificar como SIGLA (todas maiúsculas) ou ABREV

if single_abbr.isupper() or not any(c.islower() for c in single_abbr):

target_dict = data["SIGLAS"]

new_value = meaning

else:

target_dict = data["ABREVS"]

new_value = [meaning]

#*Adicionar se não existir

if single_abbr not in target_dict:

target_dict[single_abbr] = new_value

##Print(f"adicionei {single_abbr} a {meaning}")
```

5. ATUALIZAÇÃO DO *DATASET*

Fonte 2: Harvard Health Publishing

https://hgis.org.br/abreviacoes-e-siglas-padronizadas-para-registros-do-prontuario/

Medical Dictionary of Health Terms: A-C

Medical Dictionary of Health Terms: A-C

Browse dictionary by letter:

A|B|C|D|E|E|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z

A

5-alpha reductase: A chemical that changes the sex hormone testosterone into a substance called dihydrotestosterone. This hormone can cause the prostate gland to grow abnormally.

abdominal muscles: A flat sheet of muscles on the front of the abdomen, between the ribcage and the pelvis.

abdominoplasty: A procedure to remove excess abdominal skin and tighten the underlying stomach muscles. Also known as a tummy tuck.

abduction: Movement of a body part, such as an arm or leg, away from the center of the body.

Figura 4 - Página com conceitos médicos e a sua respetiva descrição da Harvard Medical School.

Fonte 2: Harvard Health Publishing

https://hgis.org.br/abreviacoes-e-siglas-padronizadas-para-registros-do-prontuario/

```
<div class="content-repository-content prose max-w-md-lg mx-auto flow-root">
  <h1>Medical Dictionary of Health Terms: A-C</h1>
  chr>
 ▶  ← 
 ▶  (p) (-) 
  <h2>A</h2>
 ▼ >
   ▼<strong> == $0
     <a name="_GoBack"></a>
     "5-alpha reductase: "
    </strong>
    "A chemical that changes the sex hormone testosterone into a substance called
    dihydrotestosterone. This hormone can cause the prostate gland to grow abnormally."
  ▼ >
    <strong>abdominal muscles: </strong>
    "A flat sheet of muscles on the front of the abdomen, between the ribcage and the
    pelvis."
  ▶  ← 
 ▶  ← 
 ▶  @ 
 ▶  ← 
 ▶  ← 
 ▶  ·  · 
 ▶  ← 
 Ezns mizzes
```

Figura 5 - Estrutura da página HTML.

Fonte 2: Harvard Health Publishing

- 1. LEITURA DO FICHEIRO 'GLOSSARIO_SIGLAS_ABREVS.JSON'
- 2. NAVEGAÇÃO PELAS PÁGINAS DO WEBSITE

Fonte 2: Harvard Health Publishing

3. EXTRAÇÃO DOS TERMOS E DEFINIÇÕES

```
soup = BeautifulSoup(response.content, 'html.parser')
container = soup.select one('div.content-repository-content')
if not container:
print("Container principal não encontrado")
continue
Ctrl+L to chat, Ctrl+K to generate
# Encontrar todos os elementos de parágrafo que contêm termos
for p in container.find all('p'):
•••• if p.strong:
term_text = p.strong.get_text(strip=True)
# Extrair o termo (remover os dois pontos finais)
term = term text.rstrip(':').strip()
normalized term = term.lower()
Extrair definição (texto após o termo)
definition = p.get text().replace(term text, '', 1).strip()
if not definition:
····|····|····|····continue
```

```
<a name="_GoBack"></a>
    "5-alpha reductase: "
    </strong>
    "A chemical that changes the sex hormone testosterone into a substance called dihydrotestosterone. This hormone can cause the prostate gland to grow abnormally."

<a name="content-repository-content prose max-w-md-lg mx-auto flow-roo to a cause the prose max-w-md-lg mx-auto flow-roo to a cause flow-roo to a cause flow-roo flow-r
```

Fonte 2: Harvard Health Publishing

4. ASSOCIAÇÃO COM CONCEITOS EXISTENTES

```
# Procurar termo nas traduções do JSON
for concept, details in data["CONCEITOS"].items():
en translations = details["traducoes"].get("en", [])
     #print(f"en_translations:{en_translations}")
     # Transformar a lista de traduções em texto JSON simulado
  translations text = json.dumps(en translations, indent=2)
  #print(f"translations text:{translations text}")
   *** # Aplicar a regex no texto
     matches = re.findall(r'((?:(?:(N{2,}-?(?:('w*)?|(d-?(?:('w*)?))))), translations text)
  #print(f"matches:{matches}")
     ···· # Extrair os termos encontrados pela regex
     normalized translations = [m.strip().lower() for m in matches]
     #print(f"normalized translations:{normalized translations}")
  for termo in normalized translations:
  #print(f"normalized term: {normalized term}")
#print(f"termo: {termo}")
-----if-termo-==-normalized term:
#print(f"normalized term: {normalized term}")
     #print(f"A adicionar definição para: {concept}")
-----details["definicoes"].append([definition, "Harvard"])
```

- Backend

O código pode ser dividido nas seguintes secções:

- Funções de Apoio
- Funções de Pesquisa
- Rotas Principais
- Rotas de detalhe
- Funções de Operações CRUD

Funções de Apoio

- carregar_dados(): carrega o ficheiro JSON com todos os dados para a memória, ficando acessível para as restantes funções viade uma variável global.
- salvar_dados(): guarda as alterações feitas aos dados no JSON.

processar_links():

- enriquece dinamicamente qualquer bloco de texto presente no frontend com hiperligações para os termos definidos no glossário.
- Exemplo, deteção de conceitos: Varre cada conceito presente na base de dados carregada, construindo para cada ocorrência uma tag <a> que aponta para a rota de detalhe desse conceito;

Backend

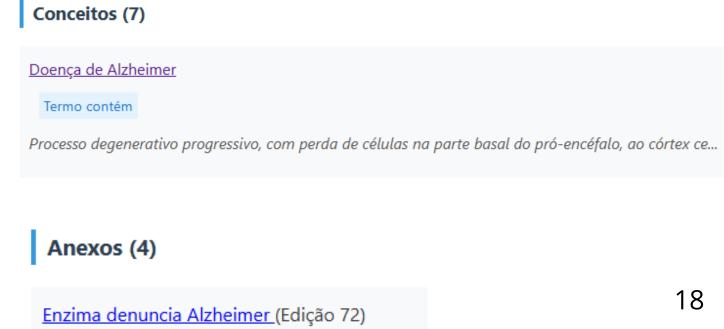
Informações Enciclopédicas

 <u>L'ambroxol</u> s'empra en el tractam metabòlit actiu de la bromhexina. S corresponent a la <u>DCI</u>.

Funções de pesquisa

- Existe uma função para cada objeto da base de dados: conceitos, siglas, abreviaturas, anexos e categorias.
- Permitem encontrar uma sequência de caracteres em qualquer campo de qualquer objeto da base de dados.
- Permitem pesquisa exata ou parcial
- Funcionamento pesquisar_conceitos():
 - Compara a query com o nome do conceito, com as suas definições, com as traduções em várias línguas, entre outros;
 - Se encontrar correspondência, regista o local da ocorrência e adiciona o conceito aos resultados com essa informação contextual.

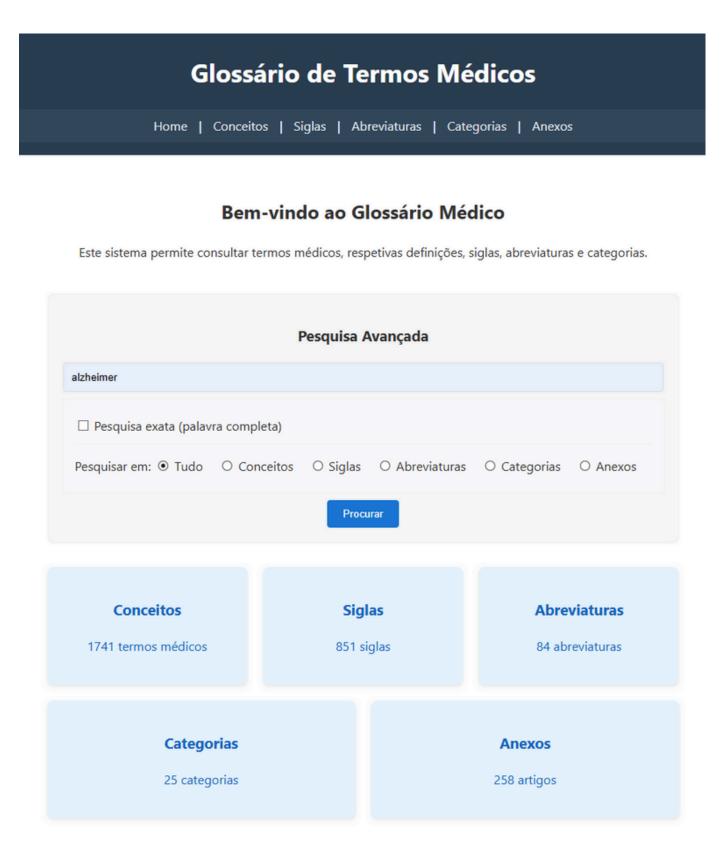




Backend

Rotas Principais

• index(): Apresenta a página inicial com estatísticas globais da aplicação.

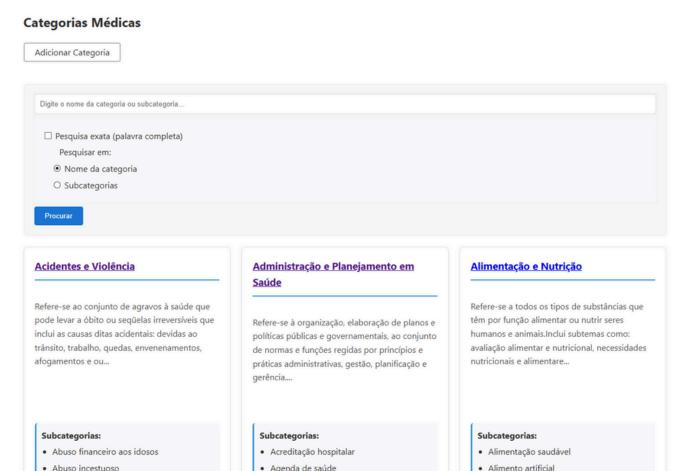


Backend

Rotas Principais

- pesquisar():
 - Centraliza a pesquisa por qualquer termo nas várias coleções da aplicação.
 - Encaminha o utilizador para a página com o resultado da pesquisa.
- Rotas de listagem (listar_conceitos(), listar_siglas(), listar_abrevs(), ...)
 - Existe uma rota para cada tipo de objeto da base de dados.
 - Encaminham o utilizador para listas pesquisáveis e navegáveis para cada tipo de recurso.
 - Permite filtrar informação em templo real da pesquisa.





Backend

Rotas de detalhe

- Cada rota apresenta informação detalhada de um único recurso (conceito, sigla, abreviatura ou categoria).
- Funcionamento detalhe_conceito(nome):
 - carrega o conceito indicado e aplica a função processar_links() para criar hiperligações entre termos relacionados.
 - As rotas detalhe_sigla(), detalhe_abreviatura() e detalhe_categoria() seguem a mesma lógica adaptada ao respetivo tipo de conteúdo.

Operações CRUD (Create, Read, Update Delete)

- As funcionalidades de criar, editar e eliminar estão disponíveis para todas as entidades da aplicação: Conceitos, Siglas, Abreviaturas, Categorias e Anexos.
- Cada operação segue um padrão comum: renderizar formulário (GET), processar input (POST), atualizar o JSON, e redirecionar com feedback.

 Adicionar Conceito

abeta
Editar Eliminar
Categoria Léxica: <u>s.f.</u>
Definições
Proteina que pode ser encontrada em todos os tipos de celulas do organismo humano. Ao acumular-se excessivamente no cortex cerebral do ser humano
pode contribuir para o aceleramento do <u>mal de alzheimer</u> . <i>(Fonte: GlossarioNeo)</i>



Demonstração —

Mestrado em Engenharia Biomédica - Processamento de Linguagem Natural em Engenharia Biomédica

TP2 - GLOSSÁRIO MÉDICO

Processamento de Linguagem Natural em Engenharia Biomédica

Realizado por:

Ana Beatriz Macedo , pg55820 André Sousa, pg52564 Maria Vale , pg56144

Docentes:

Luís Filipe Costa Cunha José João Antunes Guimarães Dias Almeida

