



Trabalho Prático

Extração e Estruturação de Termos Médicos

Beatriz Amorim, pg56112

Carolina Santos, pg56116

Catarina Nunes, pg56117



Objetivos

- Enriquecer o *dataset* obtido no TP1
- Criar uma plataforma web que permita visualizar, manipular e representar o *dataset*.

Fontes utilizadas:

Glossário

Termos técnicos usados no website

Pesquisar termos do glossário (expressão regular permitida)

pesquisar...

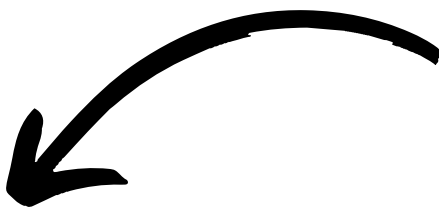
Q

Limpar

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

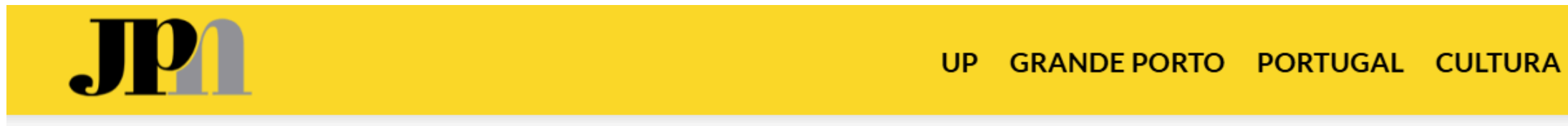
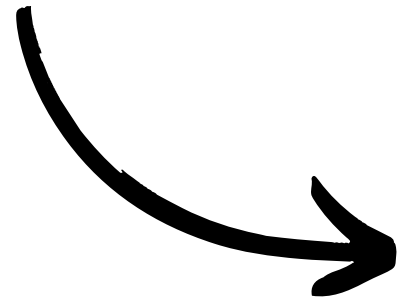
Termo	Definição Principal
<u>Abcesso</u>	Coleção de pus bem delimitada, desenvolvida numa cavidade neoformada.
<u>Ablação</u>	Extirpação cirúrgica de uma parte do corpo.
<u>Acalásia</u>	Impossibilidade de relaxar uma abertura ou esfíncter.
<u>Adenoidectomia</u>	Operação para ablação das adenoides.
<u>Admissão</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.
<u>Aférese</u>	Corte ou remoção de uma parte de um órgão; ablação. Processo de separação de elementos do sangue, permitindo a recolha seletiva de um ou mais dos seus componentes.

Centro Hospitalar de Lisboa Ocidental (CHLO):
Glossário



Fontes utilizadas:

JPN - covid-19: As palavras que entraram no nosso vocabulário.



- **Cadeia de contágio** – conjunto ordenado, série, com capacidade de transmissão.

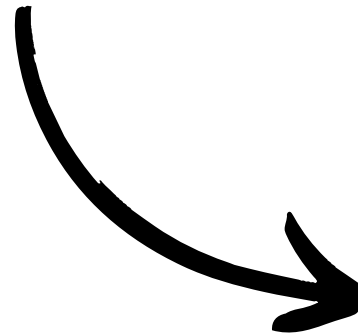
Uma pessoa contagiada pelo novo coronavírus pode originar uma cadeia de transmissão. O contágio ocorre em sucessão, de forma sequencial e interligada, ou seja, cria-se uma rede de infetados que vão contagiar as pessoas com quem estiverem em contacto e assim sucessivamente.

- **Calamidade** – Grande mal que atinge muita gente; infortúnio público.

Ovar entrou em Estado de Calamidade no passado dia 18 de março devido à propagação da COVID-19. Como consequência, ninguém sai e ninguém entra no município e os habitantes têm que permanecer nas suas casas e só podem sair à rua em situações extremamente necessárias.

Fontes utilizadas:

Hospital da Luz: Glossário para Covid-19



HOSPITAL DA LUZ Hospital da Luz Serviços e Médicos Acordos Guia de Saúde Hospitais e Clínicas Marcações

A

- **Anticorpos** - Proteínas específicas (imunoglobulinas) produzidas pelo organismo para combater infeções. A produção de anticorpos faz parte dos mecanismos de defesa do nosso sistema imunitário contra agressores (antígenos), por exemplo, vírus, bactérias, fungos, parasitas, entre outros.
- **Automonitorização de sintomas** - Verificação, pelo próprio, se tem sintomas de doença. No caso da covid-19 refere-se, mais frequentemente, a febre, tosse e dificuldade respiratória.

C

- **Comorbilidade** - Ocorrência simultânea, ou coexistência, de mais do que uma doença na mesma pessoa. Normalmente o termo comorbilidade é aplicado a doenças crónicas.
- **Cordão sanitário** - Medida de impedimento de deslocações para fora ou para dentro de uma determinada área geográfica para conter a transmissão e disseminação de uma infeção.
- **Coronavírus** - designação comum dos vírus da família *Coronaviridae*. Diversos coronavírus são causa de doença, tanto no homem como nos animais. O coronavírus que causa a doença covid-19 tem o nome de SARS-CoV-2 e foi identificado no homem pela primeira vez em 2019, em Wuhan, na China.
- **Covid-19** - Nome abreviado (do inglês *coronavirus disease 2019*) dado à doença provocada pelo coronavírus SARS-CoV-2. O nome não abreviado em português é doença por coronavírus 2019.

IR PARA TOPO

Integração dos novos dados

Os dados extraídos por *web scraping* foram integrados ao dicionário, evitando duplicações ao comparar termos sem distinguir maiúsculas e minúsculas, e juntando conteúdos sem perder informação.

Transformação e Normalização dos dados

- Termos duplicados que diferiam apenas na capitalização foram unificados, e o dicionário foi normalizado para que todos os termos comecem por letra maiúscula.
- Vários campos com a mesma informação, mas escritos de maneiras distintas foram normalizados para uma única chave:
 1. Sinónimo, sinonimo ou Sinónimos => **Sinónimos**
 2. Traduções e traducoes => **Traduções**
 3. sigla e Sigla => **Sigla**

Dicionário multilingue da COVID-19



Termo principal encontra-se em catalão



Substituiu-se estas chaves, sempre que possível, pelas respectivas traduções em português, mantendo o termo original como tradução em catalão.

As descrições em português obtidas por *scraping* foram adicionadas no campo **Descricao_pt**, mantendo a versão original em catalão no campo **Descrição**.

Transformação e Normalização dos dados

Após o *scraping*, obteve-se um novo campo - **Definição**



Essencialmente, a informação é equivalente à do campo **Descrição** que já se encontrava no dicionário.

Juntou-se numa lista os conteúdos de ambos, sob a chave **Descrição**.



Remissivas e Sinónimos complementares representam o mesmo tipo de informação



Unificados sob a chave **Remissivas**

Categorização e Similaridade

BERTimbau Base (aka "bert-base-portuguese-cased")



Categorias:

- Princípios activos;
- Etiopatogénese;
- Tratamento;
- Clínica;
- Epidemiologia;
- Prevenção;
- Contexto social;
- Conceitos gerais;
- Diagnóstico;
- Governança e administração pública;
- Políticas e intervenções na saúde pública;
- Monitoramento e avaliação;
- Gestão de dados e informação;
- Planeamento e estratégia;
- Metodologias e ferramentas;
- Componentes e conceitos farmacológicos;
- Anatomia.

Demonstração Plataforma Web