

Processamento de Linguagem Natural

Engenharia Biomédica

Trabalho Prático 2

2024-2025

O TP2 de Processamento de Linguagem Natural em Engenharia Biomédica consiste em desenvolver um sistema que permita a utilização dos dados em formato JSON que foram extraídos dos PDFs médicos durante o trabalho prático anterior. Em primeiro lugar, pretende-se que enriqueça este conjunto de dados adicionando novas informações provenientes de fontes externas, tais como websites (Web scraping) ou outros dicionários médicos que encontre. Recomendamos ainda que seja feita uma análise dos conceitos anteriormente extraídos e das suas possíveis relações entre si, de modo a agrupá-los por domínios ou outro tipo de categorias.

De seguida, pretende-se que crie uma ferramenta que permita manipular e representar os dados contidos no seu dataset. A ferramenta deve representar os dados de forma adequada, sendo capaz de explorar as relações presentes entre conceitos. Aconselhamos que crie um caso de estudo para demonstrar a aplicação durante a sua apresentação.

Tarefas

- **Enriquecer o dataset:** Utilize fontes externas, tais como websites, dicionários médicos entre outras fontes relevantes, para adicionar mais informações ao seu dataset. Pode adicionar novos conceitos ao dataset, bem como enriquecer os conceitos já existentes adicionando novos atributos, tais como sinónimos, definições, termos relacionados, domínios ou categorias. Por exemplo, "Fémur" pode ser associado à categoria "ossos".
- **Plataforma Web:** Crie uma ferramenta de visualização dos dados que permita manipular e representar o dataset enriquecido. Esta plataforma deve permitir uma navegação adequada aos seus dados, utilizando as relações existentes entre conceitos. Por exemplo, deve ser possível organizar e ordenar os dados por diferentes atributos ou categorias. Esta ferramenta deve também permitir atualizar ou acrescentar novas informações ao conjunto de dados. Qualquer alteração feita ao sistema deve ser preservada após reinicializações do mesmo.

1 Restrições

- O projeto deve ser desenvolvido na linguagem de programação Python;
- O relatório técnico deve ser produzido em LaTeX;
- O projeto deve ser desenvolvido por grupos de alunos compostos pelos mesmos membros do trabalho prático anterior.

- Este trabalho prático deve ser entregue até ao dia anterior à data da apresentação do mesmo.

2 Deliverables

Na entrega deste projeto deve entregar os seguintes elementos:

- Código desenvolvido no âmbito deste projeto;
- Ficheiro JSON com a informação enriquecida;
- Relatório técnico do sistema desenvolvido.
- Conjunto de slides para apresentação oral do projeto (entre 10 a 15 minutos).