

Natural Language Processing

Biomedical Engineering

Assignment 2

2022-2023

In this assignment, you will continue working on the JSON dataset extracted from medical PDFs that you retrieved in the previous assignment. The goal is to enrich the dataset by adding more information from external sources such as online websites (Web scraping) and other medical dictionaries. We also recommend that you perform an analysis of the terms and their possible relations trying to group them into domains or categories.

Once you have a satisfactory dataset, you will create a tool to manipulate and represent your data. The tool should be able to exploit the relations in the data and represent its information adequately. We advise you to create a case study to demonstrate the application.

Tasks

- **Enrich the dataset:** Use external sources such as online websites, medical dictionaries, or other relevant sources to add more information to your dataset. You should aim to add new attributes to the existing JSON file, such as synonyms, definitions, related terms, domains or categories. For example, "Femur" can be associated with the category "bones".
- **Tool Development:** Create a tool that can manipulate and represent the enriched dataset. The tool should be able to exploit the relations in the data and represent its information adequately. For example, organise and sort your data by different attributes. This tool should also allow to update or add new information to the dataset.