Fine Models to Extract Knowledge

Project Proposal

Supervisor: João Paulo Cordeiro

2023/2024

1 Objective

Deep learning models (DLMs), like BERT, have come to stay and solve many of the problems related to text manipulation. Nowadays, various LLMs are open access, such as *Mistral 7B*, and can be used on different tasks. More fundamentally, well-consolidated DLMs can easily be adjusted (fine-tuned) and used in various problems, like extracting knowledge from textual data. For example, from a collection of online news, one can build a network of interconnected entities and concepts that allow for a good text preview and a conceptual representation of knowledge to be computationally processed in a symbolic and human-meaning way.

Thus, this project aims to implement a system that takes advantage of DLMs to generate knowledge networks that can be visualized and stored in a standardized format.

2 Work Plan

T1: The main study of the problem. (⏱ 2 weeks)

T2: Datasets gathering and pre-processing. (⏱ 3 weeks)

T3: Knowledge engineering and fine-tuning. (⏱ 3 weeks)

T4: Knowledge extraction, storage, and visualization. (⏱ 4 weeks)

T5: Writing of the project report. (⏱ 3 weeks)

3 Academic Prerequisites

Must have good programming skills, especially in Python, and general knowledge of Machine Learning.

4 Expected Results

A system capable of extracting knowledge from a collection of documents and storing it in a conventional format. The system must also allow the graphical visualization of the extracted items and their relations.

5 Contacts

João Paulo da Costa Cordeiro (jpaulo at di . ubi . pt)
UBI | Informatics Department, Office 4.3
https://www.di.ubi.pt/~jpaulo/