Automatic Medical Knowledge Graph Construction

Joel Pardo Ferrera,
Ontology Engineering Group, Departamento de Inteligencia Artificial, ETSI Informáticos
Universidad Politécnica de Madrid,
joel.pardof@upm.es

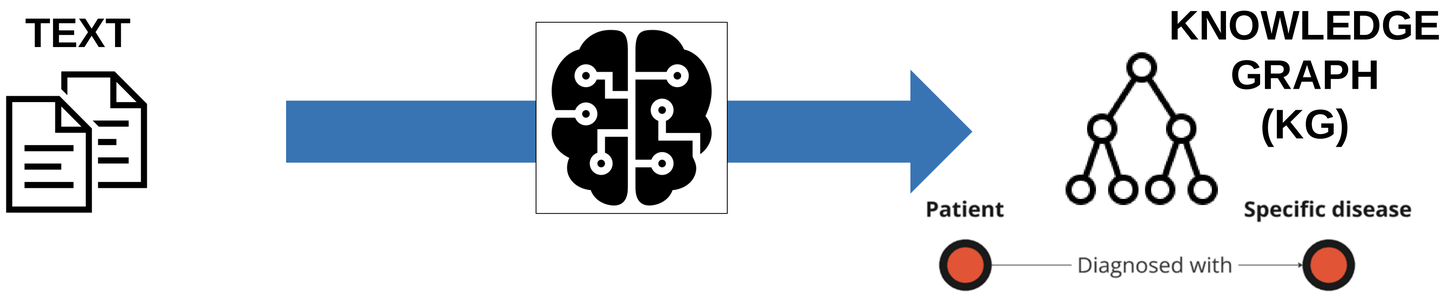
Motivation

- Increasing amount of data
- Unstructured format
- English-centric resources

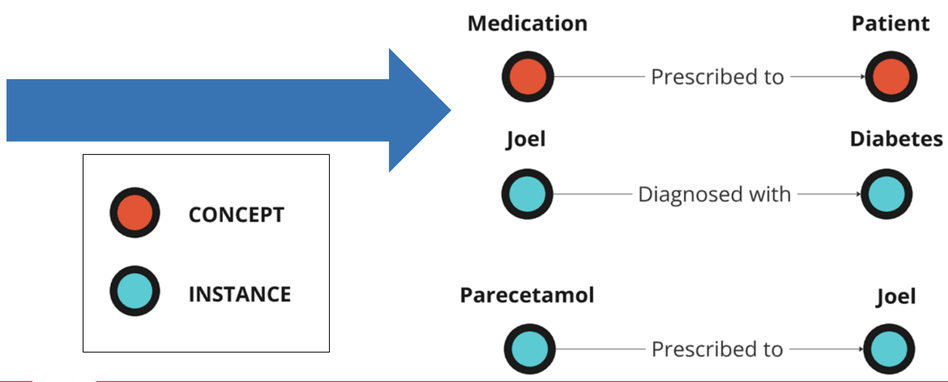
have a direct impact on

- Patient care
- Clinical decision-making
- Medical research

Goal

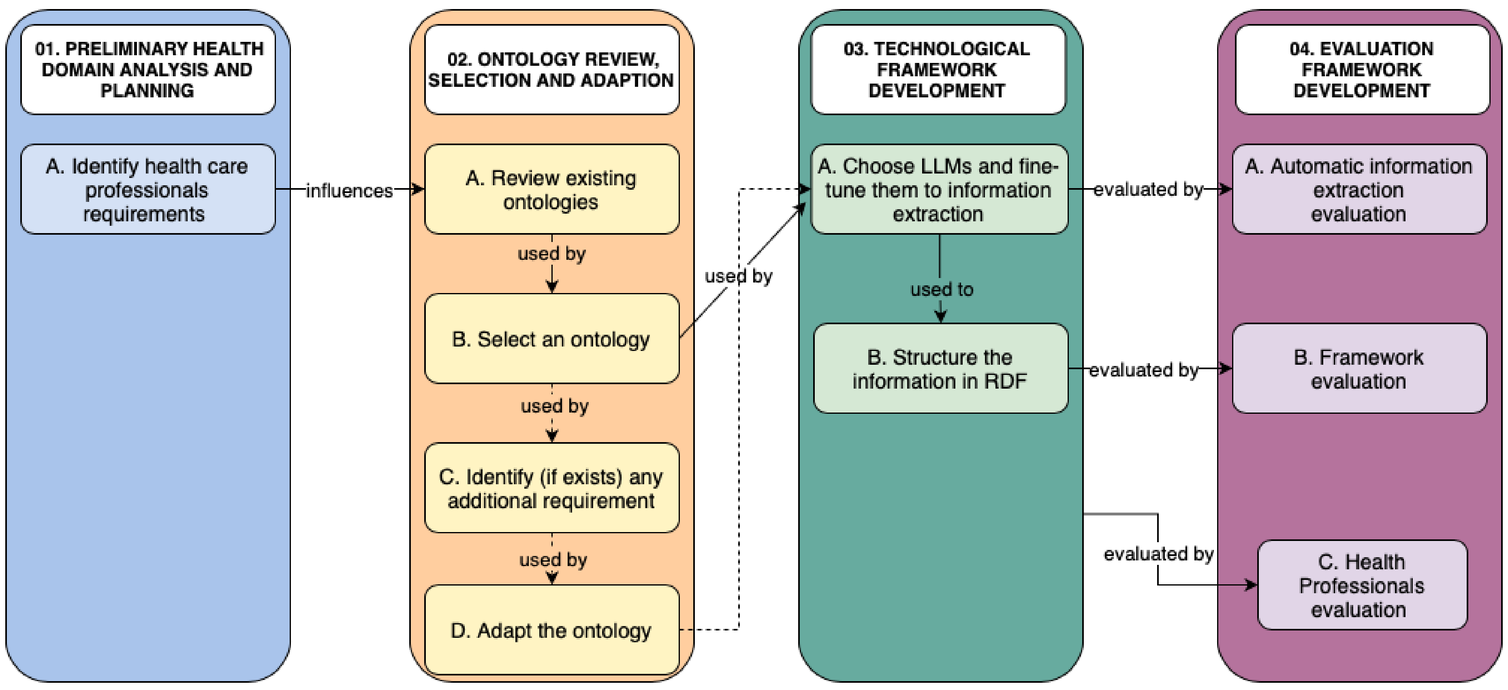


A patient can be *diagnosed with* a specific disease or *prescribed* a medication, such as Joel, a patient *diagnosed with* diabetes and *prescribed* Paracetamol.



Methodology

Developing an automated framework to generate **domain-specific KGs** for the **electronic health sector**, aimed at assisting **medical professionals** in the **decision-making process**.



- How to **automatically extract a KG** from **diverse unstructured text** in **different languages**?
- How to **effectively (re)use existent ontologies** from the **Health Sector** for KG construction?
- How to **evaluate the generated domain-specific KG** in a **real-world healthcare decision-making scenario**?

Challenges

- Reliance on traditional methods for Information Extraction tasks
- Lack of medical resources in non-English languages
- Designing ontologies by scratch is challenging
- Absence of a clear KG construction evaluation benchmark

Future Work

Collaboration with **healthcare professionals**

Setting a **unified** KG construction **evaluation benchmark**

Use of Large Language Models (**LLMs**) **to** refine **Information Extraction** tasks

Cooperate with us

- **Help developing or expanding multilingual medical corpora** like E3C to improve entity and relation extraction in non-English languages.
- **Contribute to aligning, reusing, or expanding existing medical ontologies** (e.g., SNOMED-CT) to support structured knowledge integration across different datasets.
- **Collaborate in improving NLP tasks** such as Named Entity Recognition (NER) and Relation Extraction using state-of-the-art models like LLMs for better accuracy in cross-lingual KG construction.
- **Participate in establishing comprehensive evaluation benchmarks** to measure the quality and performance of KGs in the medical domain.

Funding

This work has been funded by **INESData (Infraestructura para la INvestigación de ESpacios de DAtos distribuidos en UPM)** project, from **Ministerio para la Transformación Digital** y de la **Función Pública (PRTR, UNICO I+D CLOUD, EU NextGeneration)**.

Some icons are extracted from www.iconos8.es

@oeg_upm
www.linkedin.com/company/ontologyengineeringgroup
www.oeg.fi.upm.es

Joel Pardo Ferrera
PhD Student, Universidad Politécnica de Madrid (UPM)
AI Researcher, Ontology Engineering Group (OEG)