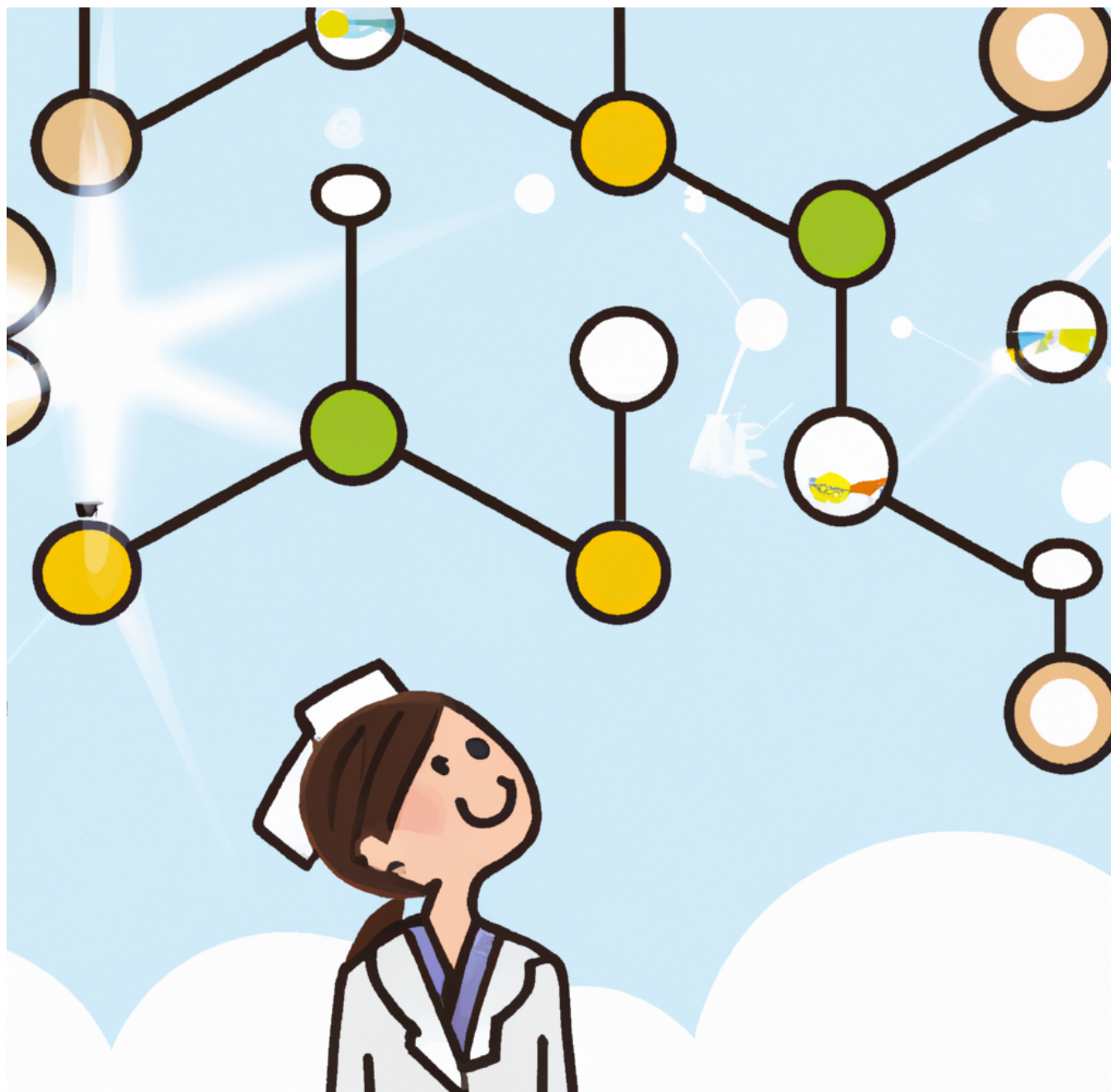# Drugs4Covid

**Knowledge graph creation based on clinical evidence automatically retrieved from scientific literature. Complex searches guided by drugs, diseases, genes and proteins. Natural language queries.**

## Contact information

**Address:** Laboratorio 3206
Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF)
Universidad Politécnica de Madrid
Avda. Montepríncipe, s/n, Boadilla del Monte, 28660, Spain
**Phone number:** 910673520
**Website:** drugs4covid.oeg.fi.upm.es
**Email:** carlos.badenes@upm.es

**Technological Offers type**

Technological solutions

**Research and innovation areas**

- Digital Technologies, Artificial Intelligence, Cybersecurity, 5G, Robotics
- Health and Wellbeing

**ODS**



**Available from:** 2022

**Where?**

Ontology Engineering Group

**Files**

Download additional documentation (pdf)

Keywords: | Biomedical Entities | Biomedical Ontology | Drug Catalogue | Evidence | Knowledge Graphs | QA

---

**Brief description of the solution and the added value it delivers**

The main objective of Drugs4Covid is to create resources, following the principles of Open Science, that facilitate the extraction of knowledge from scientific literature related to Coronavirus. These resources can be used by scientific communities conducting research related to SARS-CoV-2/COVID-19 and also by therapeutic communities, laboratories, etc., wishing to find and understand relationships between symptoms, drugs, active ingredients and their documentary evidence.

We have created a good practice guide that reviews and documents the steps needed to build knowledge graphs from sets of scientific articles. In the first stages it is necessary to structure the data that is collected in the form of written text and for this purpose we have developed named entity recognition models that identify the drugs, diseases, genes and proteins mentioned in the scientific articles. They are based on since existing language models that have been adjusted with specific vocabularies to normalise the references by standardised codes (e.g. MeSH, ATC, ICD-10, SNOMED).

To add meaning to the annotations and describe the evidence we automatically extract when processing scientific articles we have created the **EBOCA ontology**, where the associations between biomedical concepts supported in the scientific literature are modelled. As a final result we published a **Drugs4Covidknowledge graph** with evidence between the drugs and diseases mentioned in the CORD-19 corpus, which contains scientific publications on coronaviruses over the last 50 years.

Finally, and to facilitate access to all this information without the need to be an expert in semantic technologies, we have created a **question-answer interface in natural language** that allows its content to be consulted together with other external knowledge bases, such as DBpedia and Wikidata.

---

**Market demands**

- Implementation of good practices to speed up the construction of the knowledge graph.
- No knowledge of formal languages required: access to information through natural language queries.
- Development based on existing systems and following accepted standards.

---

**Previous references**

---

**Intellectual property**

- Patente
- **Registro sw**
- Secreto industrial

---

**Development stage**

CONCEPT

RESEARCH

**LAB – PROTOTYPE**

INDUSTRIAL
PROTOTYPE

PRODUCTION

---