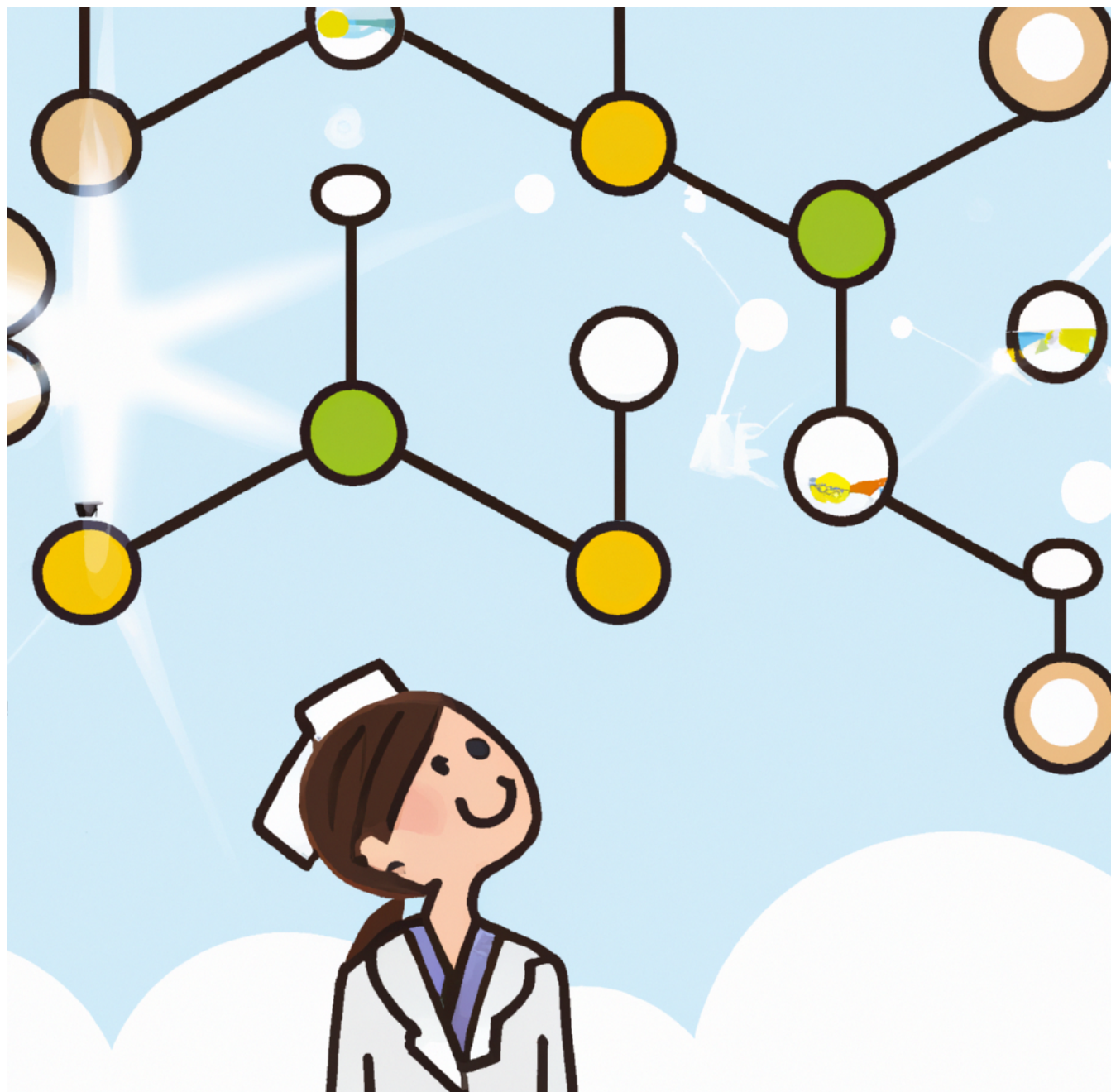


Drugs4Covid

Creación de grafos de conocimientos con evidencias clínicas extraídas automáticamente de literatura científica.
Consultas en lenguaje natural y búsquedas complejas guiadas por medicamentos, enfermedades, genes y proteínas.



Información de contacto

Dirección: Laboratorio 3206
Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF)
Universidad Politécnica de Madrid
Avda. Montepríncipe, s/n, Boadilla del Monte, 28660, Spain
Teléfono: 910673520
Página web: drugs4covid.oeg.fi.upm.es
Correo electrónico: carlos.badenes@upm.es

Tipo de oferta tecnológica

Soluciones tecnológicas

Áreas de investigación e innovación

- Salud y bienestar
- Tecnologías digitales, Inteligencia Artificial, ciberseguridad, 5G, robótica

ODS



Disponible desde: 2022

¿Dónde?

Ontology Engineering Group

Documentación

[Descargar documentación adicional \(pdf\)](#)

Palabras clave: | [Anotaciones Biomédicas](#) | [Catálogo Medicamentos](#) | [Evidencias](#) | [Grafos de Conocimientos](#) | [Ontología Biomédica](#) | [Pregunta-Respuesta](#)

Descripción breve conjunta de la solución y valor añadido que aporta

El objetivo principal de Drugs4Covid es crear recursos, siguiendo los principios de la Ciencia Abierta, que **faciliten la extracción de conocimiento a partir de literatura científica** relacionada con el Coronavirus. Estos recursos pueden aprovecharse por comunidades científicas que realizan investigación en relación con SARS-CoV-2/COVID-19 y también por comunidades terapéuticas, laboratorios, etc., que deseen encontrar y entender relaciones entre síntomas, medicamentos, principios activos y sus evidencias documentales.

Descripción de la base tecnológica

Se ha creado una **guía de buenas prácticas** que revisa y documenta los pasos necesarios para construir **grafos de conocimientos** a partir de conjuntos de artículos científicos. En las primeras etapas es necesario estructurar los datos que se recogen en forma de texto escrito y para ello hemos desarrollado **modelos de reconocimiento de entidades nombradas** que identifican los medicamentos, enfermedades, genes y proteínas mencionados en los artículos científicos. Se basan en modelos del lenguaje ya existentes que se han ajustado con vocabularios específicos para normalizar las referencias mediante códigos estandarizados (e.g. MeSH, ATC, ICD-10, SNOMED).

El descubrimiento de relaciones entre las entidades clínicas, ya sea explícito cuando se mencionan en los mismos textos o implícito a partir de sus representaciones semánticas, se ha abordado mediante **modelos de representación basados en tópicos probabilísticos**. En concreto, se ha realizado un estudio de las capacidades que ofrecen las representaciones basadas en tópicos dinámicos para capturar la relevancia de los medicamentos en el tratamiento del coronavirus. En esta línea, y para mejorar el rendimiento de los modelos del lenguaje para reconocer y relacionar entidades biomédicas en textos médicos en español, hemos **anotado manualmente 200 casos clínicos en español** con siete tipos de relaciones (analiza, altera, causa, diagnostica, manifestación_de, produce, refiere_a, y trata) y entre doce tipos de entidades (Enfermedad/Síndrome, Gen, Parte del cuerpo/Órgano, Glúcido, Procedimiento de Diagnóstico, Proteína, Procedimiento Terapéutico, Síntoma/Signo, Sustancia Farmacológica, Lípido, Organismo, Químico Orgánico y Abreviatura/Sigla/Alias).

Para añadir significado a las anotaciones y describir las evidencias que extraemos automáticamente al procesar los artículos científicos hemos creado la **ontología EBOCA**, donde se modelan las asociaciones entre conceptos biomédicos con respaldo en la literatura científica. Como resultado final publicamos un **grafo de conocimientos Drugs4Covid** con evidencias entre los medicamentos y enfermedades mencionados en el corpus CORD-19, que contiene publicaciones científicas sobre coronavirus en los últimos 50 años.

Finalmente, y para facilitar el acceso a toda esta información sin necesidad de ser experto en tecnologías semánticas, hemos creado una **interfaz de pregunta-respuesta en lenguaje natural** que permite consultar su contenido junto con otras bases de conocimiento externas, como por ejemplo DBpedia y Wikidata.

Necesidades de negocio / aplicación

- Necesidad de creación eficiente de relaciones entre entidades clínicas para optimizar las búsquedas complejas de información en este ámbito.

Referencias previas de prestación

- Como resultado final y caso de uso de la solución, se ha publicado un **grafo de conocimientos Drugs4Covid** con evidencias entre los medicamentos y enfermedades mencionados en el corpus CORD-19, que contiene publicaciones científicas sobre coronavirus en los últimos 50 años.

Ventajas competitivas

- Implementación de buenas prácticas para agilizar la construcción del grafo de conocimientos.
- No se requiere conocimiento en lenguajes formales: acceso a la información mediante consultas en lenguaje natural.
- Desarrollo basado en estándares y sistemas existentes.

Protección

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

Grado de desarrollo

CONCEPTO

INVESTIGACIÓN

PROTOTIPO - LAB

PROTOTIPO
INDUSTRIAL

PRODUCCIÓN