

[Inicio](#) > LangTech4Health Baritone: Desarrollo de una infraestructura interoperable de tecnologías de digitalización de datos de salud en español basados en procesamiento del lenguaje e inteligencia artificial

[LangTech4Health Baritone: Desarrollo de una infraestructura interoperable de tecnologías de digitalización de datos de salud en español basados en procesamiento del lenguaje e inteligencia artificial](#)

Description

El nuevo milenio impulsó la digitalización de las historias clínicas para la transición digital de la salud en España. Este proceso generó grandes repositorios de datos en hospitales, centros de salud regionales y bases de datos nacionales. Si bien los datos estructurados (p.ej. resultados de laboratorio, códigos de diagnóstico) ya se han explotado, los datos no estructurados en forma de texto libre que contienen información única del paciente (síntomas, comorbilidades, tratamiento) siguen sin utilizarse.

Es fundamental incluir datos no estructurados en el análisis masivo de datos para mejorar el manejo de los trastornos reumáticos inmunológicos (TRI). Utilizando infraestructuras informáticas de alto rendimiento (HPC), avances en Inteligencia Artificial (IA), en particular deep learning y modelos de lenguaje, se han generado mejores motores de búsqueda, de análisis de datos y aplicaciones informáticas semánticas y cognitivas. El desarrollo de soluciones lingüísticas en IA para los sistemas de salud requiere la evaluación e integración de componentes de procesamiento de texto, como la anotación semántica automática de conceptos clínicos, y su armonización con vocabularios controlados interoperables. La generación de repositorios de datos sintéticos que preserven la privacidad tienen que evitar sesgos relativos al sexo, género y otros durante la selección y anotación de datos para la capacitación de sistemas clínicos de IA y PLN.

LangTech4Health implementará un sistema de anotación semántica automática con AI avanzada y enfoques de PNL basados en deeplearning y potenciados por computación de alto rendimiento y datos clínicos reales y sintéticos. Este sistema reconocerá y extraerá automáticamente conceptos de las historias clínicas, con énfasis en fenotipos, enfermedades, síntomas, tratamientos, medicamentos y procedimientos relacionados con TRI (Objetivo 1). Para abordar la interoperabilidad semántica, la integración de datos y el análisis, implementaremos y validaremos módulos de normalización de conceptos basados en la clasificación. Los conceptos se armonizarán con terminologías médicas como SNOMED CT, ICD-10 y HPO (Objetivo 2). Estos módulos PLN se empaquetarán en dockers/conta.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 11 Mar 2025 - 15:31): <https://www.bsc.es/es/research-and-development/projects/langtech4health-baritone-desarrollo-de-una-infraestructura>