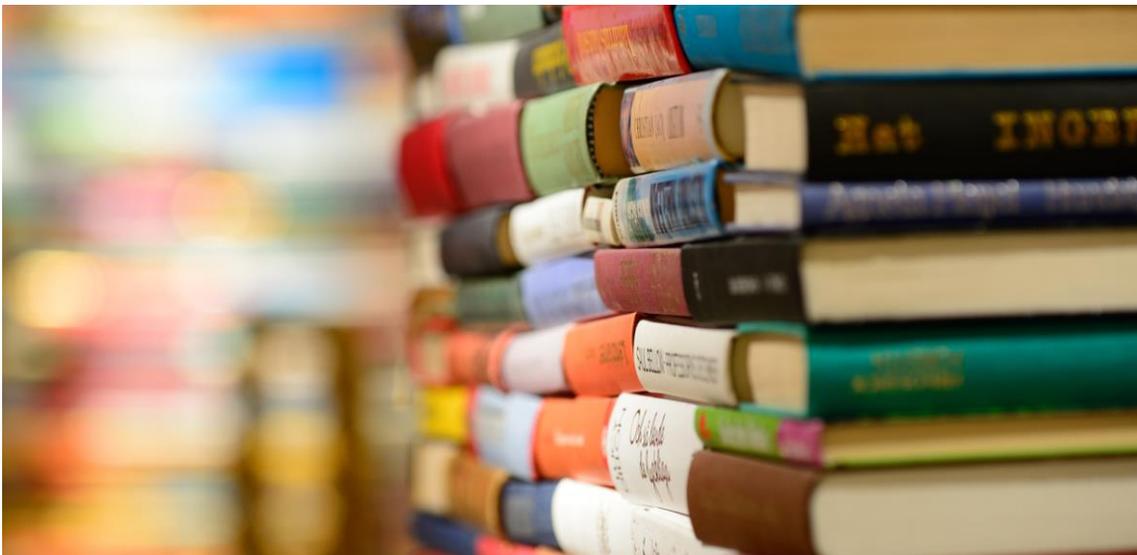


Madrid, lunes 11 de diciembre de 2023

El proyecto TeresIA recuperará y fomentará la terminología en español aplicando inteligencia artificial y conocimiento experto

- La iniciativa, coordinada por el CSIC con la participación de varias instituciones, ha sido dotada con 1,4 millones de euros por la Secretaría de Estado de Digitalización e Inteligencia Artificial
- Se enmarca en el Proyecto Estratégico (PERTE) de la Nueva Economía de la Lengua del Plan de Recuperación



El proyecto TeresIA creará un corpus de literatura científica en español. /iStock

El CSIC coordina el proyecto de terminología en español TeresIA, que generará un metabuscador de acceso a terminologías en español de alcance panhispánico y desarrollará una metodología innovadora para la generación de terminología en español, basándose en inteligencia artificial pero también en las mejores capacidades humanas y en el conocimiento experto.

El proyecto se presenta hoy, lunes 11 de diciembre, en la Delegación institucional del CSIC en Bruselas, con la asistencia de la Comisionada Especial para el PERTE Nueva Economía de la Lengua, **Cristina Gallach**, y la investigadora que dirige el proyecto, **Elea Giménez Toledo**, científica titular del Grupo de Investigación sobre el Libro Académico del Instituto de Filosofía del CSIC y coordinadora de la plataforma ES CIENCIA. El acto, que se enmarca en la Presidencia española del Consejo de la UE, contará con la presencia de representantes de la Dirección General del Español en el Mundo.

La iniciativa, denominada TeresIA: Portal de acceso a terminologías en español y servicios de Inteligencia Artificial, ha sido dotada con 1.450.000 de euros por la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA), del Ministerio de Transformación Digital.

TeresIA se enmarca en el Proyecto Estratégico para la Recuperación y Transformación Económica (PERTE) de la Nueva Economía de la Lengua, cuya Comisionada ha apoyado esta iniciativa, sensible a la relevancia del trabajo en terminología y multilingüismo en el contexto de la ciencia abierta y que, además, es de especial interés para el trabajo de traducción en instituciones multilaterales que trabajan en ámbitos multilingües.

La concesión económica de la SEDIA forma parte de la Estrategia Nacional de Inteligencia Artificial (ENIA), componente 16 del Plan de Recuperación, Transformación y Resiliencia.

Corpus de literatura científica en español

TeresIA creará un corpus de literatura científica en español; diseñará una metodología de extracción de información aplicando técnicas de inteligencia artificial (IA) y Procesamiento del Lenguaje Natural (PLN) para extraer términos, mostrarlos en su contexto, validarlos, sancionarlos, transformarlos a formatos de la web de datos y hacerlos visibles y recuperables en distintos entornos, incluido el metabuscador.

“Todo ello generará un enorme conjunto de datos terminológicos abiertos, siguiendo principios FAIR; y favorecerá la interacción con comunidades de expertos a los que se implicará en los procesos de validación”, indica Elea Giménez. “Desarrollar terminología en español en el marco de la ciencia abierta representa un avance para el multilingüismo”, señala la investigadora.

“Al trabajar con estándares tecnológicos, las palabras de la ciencia en español se conectarán con las palabras de la ciencia en otros idiomas. De este modo, se multiplican las posibilidades de recuperar y descubrir contenidos científicos en distintos idiomas. Además, se preserva la riqueza terminológica de cada lengua, se favorecen las traducciones especializadas (tanto la humanas como las automáticas), se entrenan algoritmos, y se mejora la enseñanza del español de especialidades, entre otras aplicaciones”, detalla Giménez Toledo.

Interinstitucional

El proyecto se desarrollará conjuntamente por un consorcio conformado por seis instituciones: la [Plataforma Temática Interdisciplinar \(PTI\) ES CIENCIA](#), del CSIC, el Instituto Cervantes, el Centro Nacional de Supercomputación (BSC), el Grupo de

Ingeniería Ontológica de la Universidad Politécnica de Madrid, la Asociación Española de Terminología (AETER) y la Real Academia de Ingeniería (RAI), que contarán con el apoyo del Departamento de Lengua Española de la Dirección General de Traducción de la Comisión Europea.

La idea primigenia de TeresIA surgió hace más de quince años, pero no había podido desarrollarse hasta el momento por distintas razones. El nombre del proyecto quiere reconocer el destacado trabajo de Teresa Cabré en la definición y planteamiento de esta iniciativa. La colaboración interinstitucional, los procesos innovadores introducidos en el proyecto, la visión política para apoyarlo y la oportunidad de los fondos europeos NextGenerationEU del Plan de Recuperación, han hecho que, finalmente, pueda ejecutarse. “En este momento hay que reconocer la perseverancia de Luis González (DGT) y Miguel Marañón (Instituto Cervantes) y de los distintos presidentes de AETER al no abandonar la idea de desarrollar el proyecto”, indica la investigadora.

La Plataforma Temática Interdisciplinar ES CIENCIA nació en 2019 con el reto de impulsar la calidad, la visibilidad, la descubribilidad y el reconocimiento de los contenidos científicos en español. El trabajo de colaboración interinstitucional de la plataforma ha hecho posible la configuración del consorcio que desarrollará TeresIA como acción estratégica del español y el multilingüismo en la comunicación científica, eje prioritario de investigación de ES CIENCIA. El proyecto se alinea con el grupo de trabajo sobre multilingüismo de OPERAS, infraestructura en la que el CSIC se ha integrado este año. Dentro del CSIC, además, el proyecto contará con el apoyo de la Unidad de tratamiento de la información del Centro de Ciencias Humanas y Sociales (CCHS), cuyo equipo forma parte de ES CIENCIA. Al asumir el liderazgo de este proyecto, el CSIC quiere recordar a Amelia de Irazazábal y al equipo de investigadoras pioneras que pusieron en marcha el grupo TERMESP.

Uno de los objetivos de la plataforma es impulsar la integración de información/recursos en infraestructuras nacionales e internacionales, de forma que los contenidos científicos en español sean visibles y tengan proyección en distintas comunidades científicas. Este objetivo confluye con el eje estratégico 3 de la ENIA, consistente en “Desarrollar plataformas de datos e infraestructuras tecnológicas que den soporte a la IA”, encuadrado en la medida 14: «Impulso al Plan Nacional de Tecnologías del Lenguaje», cuyo objetivo es impulsar el desarrollo específico de tecnologías de inteligencia artificial en el ámbito del procesamiento del lenguaje natural.

CSIC Comunicacióncomunicacion@csic.es