



REAL ACADEMIA ESPAÑOLA

NOTA DE PRENSA

La RAE, inmersa en la nueva fase del proyecto LEIA (Lengua Española e Inteligencia Artificial)

- La Academia trabaja en la creación de un observatorio de neologismos, tecnicismos, términos y variaciones del español, así como un verificador ortográfico, gramatical y léxico, y una herramienta de respuesta a dudas lingüísticas, entre otras cuestiones.
- El acceso a estos recursos basados en inteligencia artificial será libre y abierto mediante una plataforma visible desde la página web de la RAE.
- Se pondrá también el foco en la diversidad de las variedades geográficas del español y la accesibilidad de las herramientas de IA.
- Para el desarrollo de estos proyectos serán clave los materiales y obras en los que la RAE lleva trabajando más de 300 años, fuentes cruciales para un buen entrenamiento de la IA.

14 DE MARZO DE 2024

La Real Academia Española (RAE) se encuentra inmersa en la segunda fase del proyecto LEIA (Lengua Española e Inteligencia Artificial), centrada en la creación de herramientas, tecnologías y aplicaciones para la proyección y uso de la lengua española en el ámbito de la IA. Un trabajo que se encuentra enmarcado en el Plan de Recuperación, Transformación y Resiliencia, financiado por la Unión Europea mediante los fondos NextGenerationEU.

Esta nueva etapa del proyecto, ideado y liderado por la RAE, se traducirá en la creación de un observatorio de neologismos, tecnicismos, términos y variaciones del español, y de herramientas de verificación ortográfica, gramatical y léxica y de respuesta a consultas lingüísticas; la recopilación de material basado en la diversidad de las variedades geográficas del español; el análisis y la aplicación de la accesibilidad de las herramientas de IA; la mejora de herramientas de regulación de la lengua, y labores de digitalización de fondos propios de la Academia. Además,

para apoyar las pruebas y el desarrollo de estas tareas, está prevista la realización de diversos retos, entre otras cuestiones.

Los servicios resultantes de esta fase del proyecto LEIA estarán integrados en una plataforma a la que podrá acceder de manera libre y gratuita para su consulta todo aquel que lo desee a través de la página web de la RAE (www.rae.es). Para su construcción serán clave los materiales y obras en los que la RAE lleva trabajando más de 300 años, fuentes cruciales para un buen entrenamiento de la IA.

LA NUEVA FASE EN EL MARCO DE LEIA

Todas las actividades anteriores, que ya han comenzado a acometerse, cumplen con el objetivo fundamental de LEIA: la defensa, proyección y buen uso del español en el universo digital y, especialmente, en el ámbito de la inteligencia artificial y las tecnologías actuales. De esta manera, se siguen las líneas principales del proyecto LEIA, que son, por un lado, velar por el buen uso de la lengua española en las máquinas, en los sistemas y las aplicaciones instaladas en los dispositivos y que utilizan el lenguaje, y, por el otro, aprovechar la inteligencia artificial para crear herramientas que fomenten el uso correcto del español en los seres humanos.

Desde su presentación por parte del director de la RAE y presidente de la ASALE, Santiago Muñoz Machado, en un gran acto celebrado en Sevilla en 2019 y presidido por SS. MM. los Reyes, se ha trabajado en esta iniciativa desde dos vertientes. De manera bilateral con las grandes empresas tecnológicas a nivel nacional y mundial para procurar que sus desarrollos utilicen un español correcto, lo que ya ha dado sus primeros resultados, y ahora en el marco del RD 632/2022 de 26 de julio de 2022, en el que se regula la concesión directa de una subvención a la Real Academia Española de cinco millones de euros por parte de la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA), dependiente del Ministerio para la Transformación Digital y de la Función Pública, dentro del Plan de Recuperación, Transformación y Resiliencia, con las actividades anteriormente mencionadas.

HERRAMIENTAS LINGÜÍSTICAS BASADAS EN IA AL ALCANCE DE TODOS

La RAE, dentro de esta segunda fase del proyecto LEIA, está trabajando en la creación de un **observatorio de neologismos, términos y variaciones del español**. Este consistirá en una herramienta capaz de detectar automáticamente, en el universo digital y a partir de un buen número de fuentes (obras académicas, corpus, prensa, redes sociales...), palabras y expresiones que por distintos motivos no están registradas en el *Diccionario de la lengua española (DLE)* u otras obras académicas, así como nuevos usos de palabras. Estos términos serán

principalmente neologismos, derivados, tecnicismos, regionalismos y extranjerismos.

Esta herramienta permitirá a los equipos lingüísticos de la RAE observar diariamente palabras que puede ser conveniente estudiar. Posteriormente, la plataforma mostrará públicamente una selección de esas palabras en observación con información sobre su uso.

El observatorio ofrecerá tanto datos sobre términos de uso actual como sobre nuevos significados de palabras que ya contaban con un amplio recorrido en nuestra lengua. Estos comentarios siempre serán provisionales, ya que al no estar incluidos en las obras académicas podrán verse modificados en el futuro.

Por otra parte, se creará un **verificador lingüístico** en abierto que se alojará, asimismo, en la página de LEIA, accesible desde la web de la RAE, y que permitirá a los usuarios introducir un texto para comprobar si es correcto desde un punto de vista ortográfico, gramatical y léxico.

El verificador hará sugerencias de cambio que expliquen el problema detectado y envíen al lugar donde se trata la cuestión, dentro de las obras de la RAE. Para el perfeccionamiento durante la creación de este recurso, que estará interconectado con los otros desarrollos de la plataforma, el trabajo interno se complementará con encuentros (como *hackathons*) para poner a prueba el funcionamiento de la herramienta y fomentar su conocimiento entre el público especializado o con inquietudes por la aplicación de las nuevas tecnologías en la lengua.

Otra parte del proyecto es la creación de una **herramienta para dar respuesta a las dudas lingüísticas** de los hispanohablantes. Se podrá introducir una cuestión en un buscador y el sistema seleccionará la respuesta de entre los cientos de miles almacenadas en la plataforma. En caso de que el sistema aún no tenga una respuesta exacta para la pregunta planteada, el usuario podrá enviarla a través de un cuestionario y recibirá la contestación personalizada de los lingüistas de la RAE.

UNA IA ENTRENADA EN EL PANHISPANISMO

Dentro de esta fase del proyecto, otra línea es la **recopilación de material de las distintas variedades geográficas del español** (especialmente léxico y oral). Para hacer esta radiografía de la diversidad de nuestra lengua, se creará una sección interactiva en la que los hispanohablantes puedan aportar información en relación con imágenes, textos u otros elementos que se les muestren. La participación ciudadana será clave en este apartado del proyecto, en el que, por ejemplo, los usuarios podrán ver una imagen y describirla con su propia voz, u observar un objeto e indicar cómo lo llaman en su lugar de procedencia.

Con las grabaciones, se pretende crear un corpus oral que permita entrenar a los

sistemas o aplicaciones en los distintos acentos para que puedan reconocerlos con facilidad.

LEIA persigue que la consulta de sus resultados sea accesible para todos. Por ello, se va a trabajar en que así sea con la implementación de medidas de **accesibilidad** que se aplicarán a los espacios digitales públicos de esta rama del proyecto.

Otros puntos en los que se está trabajando es en la **mejora de herramientas tecnológicas de la RAE** que se emplean en las labores de regulación de la lengua, con el fin de optimizarlas para la mejor difusión y aplicación de sus contenidos, y en la continuación de la **digitalización** de parte de los fondos y obras de la Academia.

Por último, se pretende organizar **retos** (*hackathons* y *datathons*) en los que se plantee una cuestión relacionada con el proyecto a un gran número de participantes y se pidan posibles opciones, entre otros aspectos. La idea principal es compartir los resultados y avances con la comunidad investigadora y universitaria, y que su talento sea partícipe de LEIA.

Junto a estos servicios, también se pondrán a disposición pública, de manera abierta en los formatos adecuados, y siempre que sea posible, los materiales generados para la creación de este proyecto (como códigos fuente, datos o corpus de entrenamiento). De esta manera, tal y como se anunció en el Real Decreto 632/2022, de 26 de julio de 2022, las infraestructuras lingüísticas y los materiales desarrollados gracias a esta vertiente del proyecto contribuirán a impulsar la industria de las Tecnologías del Lenguaje en español.

Las tareas explicadas anteriormente se han dividido en distintos paquetes de trabajo. Para su ejecución, la RAE cuenta con su personal interno, así como con la participación de empresas externas subcontratadas en los términos establecidos por el real decreto. Hasta el momento, estas compañías son Fujitsu, con tecnología de Amazon Web Services (AWS), para los paquetes correspondientes al observatorio, verificador, diversidad y accesibilidad, y VASS, para la elaboración de la herramienta de consultas y respuestas.

Prensa Real Academia Española
comunicacion@rae.es
91 420 14 78 297

