

La RAE y Amazon crean un radar para cazar gazapos lingüísticos en internet

► Es una inteligencia artificial entrenada con casi 9 millones de documentos

BRUNO PARDO PORTO
MADRID

Hay una inteligencia artificial que lleva meses leyendo lo que escribes en Twitter. También lo que dices en los foros y hasta en Wallapop. No quiere venderte un nuevo champú ni saber a dónde vas de vacaciones para anunciarte hoteles. Tampoco quiere espiarte para vender tus datos. Solo quiere saber cómo te expresas, en qué fallas, qué extranjerismos usas y cuáles son tus erratas más comunes. Es capaz de medir tu riqueza léxica y valorar tu claridad comunicativa. ¿Para qué? Para saber cuál es la salud del español en internet.

Esta innovadora máquina, desarrollada por Amazon Web Services (AWS) y la Real Academia Española (RAE), y que forma parte del proyecto LEIA, es capaz de analizar grandes cantidades de documentos en muy poco tiempo, y tras meses de trabajo y pruebas ya está lista para servir a los intereses de los académicos.

Una revolución

Santiago Muñoz Machado, director de la institución, explicó durante la presentación de la herramienta que ésta les ayudará en su tarea centenaria, que es, a la vez, su razón de ser: vigilar el español, escuchar a sus hablantes, registrar sus usos y gazapos y demás ejercicios de diagnóstico. «Queremos saber con qué problemas nos encontramos en la lengua en las redes, en internet, y esto nos permite hacerlo de un modo automático, y no a mano, como antes -celebró-. Estamos en un momento refundacional del cuidado del idioma que está provocado por la gran revolución de la inteligencia artificial».

La versión beta de esta inteligencia artificial, que fue la que se hizo pública ayer en la Docta Casa, ha sido entrenada con casi nueve millones de archivos o, lo que es lo mismo, con ciento cincuenta gigas de texto (esto es el triple de lo que ocupan todas las

entradas de Wikipedia en español). Esta ingente base de datos representa, según aseguran los expertos, el «español digital espontáneo actual». De hecho, el primer informe que la RAE hará con este apoyo, y que conoceremos antes de que acabe el año, versará sobre el estado del español en internet. Pero el plan es utilizar esta tecnología con bases de datos muy diversas. Por ejemplo, con documentos jurídicos. O periodísticos. O institucionales. Y así emitir estudios con el sello de la Academia valorando la calidad del español en esos ámbitos, o incluso enfrentarlos para ver cómo varía la lengua según el contexto.

Análisis por países

De momento, esta inteligencia artificial, tal y como detalló Carlos Carús, director de tecnología de AWS en España y Portugal, está dividida en tres campos: el estudio del uso de extranjerismos y su proporción en un texto (uno de los más extendidos, por cierto, es 'stop'), el análisis de la riqueza léxica (a través del sistema MTLD) y el radar de errores lingüísticos, que los identifica y clasifica en ortográficos, gramaticales, léxicos, de estilo y tipográficos. Los resultados de estos cálculos pueden visualizarse en mapas y gráficos sencillos y organizarse según diferentes

criterios, como el tipo de texto, la fecha o el lugar de origen, lo que permite localizar cuáles son los países en los que se utilizan más extranjerismos o cuáles presentan una mayor riqueza léxica, por poner un caso.

En un futuro no muy lejano estos parámetros serán más y se podrá analizar la claridad del lenguaje administrativo, comparar la calidad del español por épocas o detectar las incorrecciones de los asistentes de voz y otros dispositivos tecnológicos. Todo esto, claro, siguiendo siempre los criterios fijados por la RAE: los algoritmos que detectan los errores han sido diseñados siguiendo las obras académicas. «Esta herramienta acaba de nacer, es el primer paso, esperamos que haya muchos más. El objetivo es aplicar la inteligencia artificial para cuidar nuestra lengua y que siga evolucionando a los nuevos usos», aseveró Miguel Álava, director general de AWS España.

