

8Belts y la Universidad de Granada colaboran en un proyecto de modelado del lenguaje basado en Machine Learning e IA

El proyecto tiene como objetivo desarrollar una plataforma inteligente que permita a los alumnos realizar ejercicios de escritura en el nuevo idioma, donde será el propio sistema quien corrija y realice recomendaciones para obtener unos resultados correctos

8Belts y la Universidad de Granada colaboran en un proyecto de Modelado del Lenguaje Humano. La iniciativa se enmarca en la estrategia de innovación de 8Belts orientada a la incorporación pionera de tecnologías como la Inteligencia Artificial y el Machine Learning a su método online de aprendizaje de idiomas.

Esta estrategia arranca en la Universidad de Granada ofertando una serie de temáticas de Trabajos de Fin de Grado que incluirán infraestructura, asesoramiento y prácticas por parte del equipo de 8Belts. Este año lectivo, comienza con el primer proyecto dentro del área del Procesamiento del Lenguaje Natural y Machine Learning. Al frente de esta colaboración por parte de la Universidad de Granada se encuentra Juan Francisco Huete Guadix, Catedrático y Director del Departamento de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Granada.

“Hasta ahora, el Procesamiento del Lenguaje Natural se ha utilizado en entornos muy específicos relacionados con la seguridad o la automatización y nuestro objetivo es aplicarlo en la mejora del aprendizaje de idiomas”, comenta Carlos Moreno, director de Innovación de 8Belts.

El proyecto, centrado inicialmente en el aprendizaje del inglés y que se extenderá paulatinamente al resto de idiomas que ofrece 8Belts, tiene como objetivo desarrollar una plataforma inteligente que permita a los alumnos realizar ejercicios y practicar la escritura en el nuevo idioma, donde será el propio sistema quien corrija y realice recomendaciones para obtener unos resultados correctos. El nuevo sistema sabrá qué palabra conecta con otras y podrá conectar hasta cuatro palabras hacia atrás y otras cuatro hacia delante, incluyendo, en el caso del inglés, los complicados *phrasal verbs*.

“*El objetivo* -señala Carlos Moreno- es facilitar y fijar el aprendizaje. Gracias a esta herramienta, los alumnos podrán practicar y crear frases fácilmente y de manera muy dinámica, hasta llegar al resultado correcto, realizando en segundos lo que antes tardaba minutos”.

Grafos

En el ámbito de la IA, el proyecto se sitúa en el área de “Procesamiento de Lenguaje Natural y teoría de grafos”. Un grafo está formado por nodos y aristas, los nodos son palabras y las aristas conexiones entre palabras, y el sistema guarda las estadísticas de cómo se van conectando una palabra con la

siguiente, midiendo el número de veces que aparecen conectadas de determinadas maneras. El grafo permite generar texto sintético y simular cualquier idioma.

El proyecto utiliza una base de datos de 8Belts formada por grafos, entrenada con decenas de miles de libros y que ha aprendido cómo se articulan miles y miles de palabras y cuál es la distribución de la siguiente. Actualmente esta base de datos incluye más de 50 millones de nodos y 80 millones de aristas.

En su aplicación al método 8Belts, el grafo tendrá diferentes utilidades como conectar conceptos, “cuál es el camino que hay de una palabra a otra” mediante palabras intermedias y también actuará como corrector. También ayudará al alumno a conectar dos conceptos mediante otras palabras intermedias y podrá generar sinónimos y estructuras gramaticales similares a otras.

A medida que el alumno escribe, el sistema le indicará lo que es incorrecto o no tiene sentido y le sugerirá una serie de palabras. “No tiene nada que ver – señala Ernesto Martínez, Arquitecto de IA y responsable del proyecto- con los correctores ortográficos, que se basan en un sistema de reglas, ya que analiza el contexto probabilístico entero donde se desarrolla la frase, y determina si tiene sentido lo que escribe el alumno”.

El proyecto aúna el Machine Learning y los grafos probabilísticos utilizando una enorme base datos orientada a grafos. El siguiente paso que contempla 8Belts a medio plazo es cambiar el concepto del “writing” y llevarlo a un contexto visual. “En la pantalla -explica Ernesto Martínez- habrá elementos gráficos que representan diferentes palabras que habrá que conectar de una determinada manera y añadir otros elementos gráficos para crear frases. Cuando la conexión no sea correcta, el elemento reaccionará indicándolo para que el alumno corrija. Una vez construida la frase, el alumno podrá verificar si es correcta. La idea es gamificar los ejercicios de escritura y desarrollar memoria visual”.
