Tecnologías emergentes y datos abiertos:

PROCESAMIENTO DEL LENGUAJE NATURAL

# Resumen ejecutivo

El **Procesamiento del Lenguaje Natural** es hacer que las máquinas (los ordenadores) **entiendan el lenguaje humano**: hablado o en forma de texto. De manera más formal, el procesamiento del lenguaje natural es un **campo híbrido entre la informática y la lingüística**, que utiliza diferentes técnicas, algunas de ellas basadas en **Inteligencia Artificial**, para interpretar el lenguaje humano.

En este informe, veremos cómo el procesamiento del lenguaje natural está mucho más cerca de nuestro día a día de lo que podemos pensar inicialmente. Aplicaciones como la **traducción automática de textos**; el **análisis de sentimiento en redes sociales**; las **búsquedas que realizamos en Internet**; la **generación de resúmenes** meteorológicos o las sencillas peticiones que hacemos a nuestro **altavoz *inteligente****,* tienen una fuerte componente tecnológica de procesamiento del lenguaje natural.

El peso específico que el procesamiento del lenguaje natural tiene (y va a tener) en **la indústria y en la economía** es cada vez mayor, puesto que la mayoría de los datos que se producen en el mundo (fundamentalmente a través de Internet) son datos en forma de textos y voz (datos desestructurados). Los **datos abiertos** **juegan un papel crucial para esta tecnología**. Los algoritmos de inteligencia artificial que se usan para analizar y *entender* el lenguaje natural, requieren de una ingente cantidad de datos de calidad para ser entrenados. Multitud de estos datos proceden de los repositorios de datos abiertos de instituciones tanto públicas como privadas.

A lo largo de este informe, se repasa **la historia del procesamiento del lenguaje natural**, desde sus inicios hasta nuestros días. Además, en la sección *Inspire,* se describen algunos de los **casos de uso** más representativos que aprovechan el potencial del procesamiento del lenguaje natural. La predicción de texto al escribir un nuevo email, la clasificación de textos en categorías o la generación de noticias falsas, son solo algunos de los casos que se repasan en este informe.

Finalmente, para aquellos lectores más entusiastas, en la sección *Action,* se **desarrolla un caso de uso completo** (utilizando herramientas de programación) sobre el **análisis de sentimiento en conversaciones sobre debates públicos** ciudadanos.