



¿Qué es el **procesamiento natural de texto**?



El desarrollo del lenguaje y la comunicación a través de textos marcó un hito en el desarrollo de las civilizaciones. Ahora, con el desarrollo en la capacidad de procesamiento de los equipos de cómputo y el avance en la inteligencia artificial, surge la posibilidad de utilizar las computadoras para analizar, entender y generar voz y texto, de la misma manera que una persona lo haría. A este auge tecnológico se vincula el procesamiento natural de lenguaje o natural language processing (NLP, por sus siglas en inglés).

NLP busca la manipulación o respuesta automática al lenguaje, en voz o texto, por medio de software, de la misma manera que un ser humano lo haría. Los inicios de NLP se remontan a la década de 1950 con el trabajo de Alan Turing. Sin embargo, el crecimiento más acelerado se ha generado a partir del 2010, con el aprendizaje automático y las redes neuronales profundas. Imagínese el potencial que tiene NLP, podríamos darle a las computadoras la habilidad de procesar y entender el significado de un texto o el sentimiento asociado con él.

A través de NLP se desarrollan las herramientas para traducción de idiomas, los sistemas para responder a comandos de voz, o las aplicaciones para resumir grandes cantidades de texto escrito. De hecho, las aplicaciones del NLP ya están entre nosotros, por ejemplo, los sistemas de ubicación GPS operados por voz, los asistentes de voz como Siri en iPhone, Cortana en Windows o Alexa en Amazon, las herramientas para dictar en los procesadores de texto o detectar *emails* mal intencionado o *spam*, y los agentes virtuales o *chatbots* para atención a clientes.

Dado que el lenguaje humano está lleno de ambigüedades, el campo de investigación en NLP está en continuo crecimiento, engloba la lingüística, las ciencias computacionales y la inteligencia artificial. En general, dentro de NLP existen las siguientes tareas básicas que ayudan en el procesamiento de voz y texto: (i) reconocimiento de voz a texto escrito, (ii) etiquetado gramatical, (iii) análisis de sentimiento, (iv) generación de lenguaje natural, (v) identificación del sentido de una palabra y (vi) resolución de co-referencias.

Las compañías pioneras en el desarrollo de aplicaciones de NLP son Google, Yahoo, IBM, Microsoft, Amazon y

recientemente OpenAI. En especial, esta última es una empresa sin fines de lucro fundada en 2015 por Sam Altman y Elon Musk dedicada a la investigación y el desarrollo de la inteligencia artificial, que busca un beneficio común para la humanidad. Dentro de las aplicaciones más impresionantes en OpenAI remarcó tres. La primera asociada con los sistemas GPT-1, GPT-2 y GPT-3, los cuales fueron entrenados para capturar unas ciertas oraciones de entrada, y generar todo un texto de salida que profundice las ideas ingresadas, en función de un entrenamiento previo, considerando los textos disponibles en internet. La evolución de GPT-1 hasta GPT-3 estableció un incremento exponencial en la complejidad del sistema de inteligencia artificial entrenado. La segunda aplicación es Codex, que se basa en una reorientación de GPT-3 y permitirá a los programadores definir lo que quieren realizar en su código como lenguaje natural. Es decir, la herramienta es capaz de codificar un texto que describe las instrucciones y acciones a realizar, en líneas de código, considerando diversos tipos de lenguajes de programación. Finalmente, la tercera y la aplicación más reciente, publicada en septiembre pasado, utiliza la estructura computacional de GPT-3 para resumir un libro completo.

Todos estos avances en NLP permitirán que, en corto tiempo, las computadoras procesen, entiendan y generen textos tal como lo hacemos los seres humanos. Aunque esto abre nuevas posibilidades de aplicación, la explotación del NLP en un futuro inmediato requiere considerar los aspectos éticos y de responsabilidad social, así como las regulaciones gubernamentales, lo cual es crucial para guiar los desarrollos de NLP en beneficio de la sociedad. **UP**

