

# Tecnologías emergentes y datos abiertos: PROCESAMIENTO DE LENGUAJE NATURAL

## INTRODUCCIÓN Y METODOLOGÍA

Contenido elaborado por Alejandro Alija, experto  
en Transformación Digital y datos abiertos.



## INTRODUCCIÓN

“Oye, Siri”, “Ok, Google”, o “Hey Mercedes” son expresiones cotidianas para una parte de la población y, sin embargo, muy pocas personas son conscientes de los fundamentos científicos y tecnológicos que residen debajo de estos comandos tan sencillos. Cuando la complejidad tecnológica es transparente al usuario final es síntoma de que la tecnología está en su punto de madurez óptimo. Así que, **si alguna vez has intentado dialogar con un asistente virtual, eso es fundamentalmente, Procesamiento del Lenguaje Natural**, en adelante NLP (del inglés, *Natural Language Processing*), una de las ramas de las Tecnologías del Lenguaje.

Las tecnologías para el Procesamiento del Lenguaje Natural están 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 los resúmenes** del tiempo o de eventos deportivos, o las peticiones simples que hacemos a nuestro **altavoz inteligente**, tienen una fuerte componente tecnológica de Procesamiento del Lenguaje Natural.

Sobre el **impacto del NLP en el mundo**, un [informe de 2017 de Tractica](#) estima que la oportunidad total de mercado de software, hardware y servicios de NLP será de más de **22.000 millones de dólares (americanos) para 2025**. El documento también pronostica que las soluciones de software de NLP que aprovechan las **capacidades de la Inteligencia Artificial** verán un crecimiento del mercado de **5.400 millones de dólares para 2025**.

En este contexto, desde la Iniciativa Aporta hemos elaborado el informe “Tecnologías emergentes y datos abiertos: Procesamiento del Lenguaje Natural”, donde **veremos en profundidad algunos de los casos de uso más relevantes del NLP en la actualidad**. Además, entenderemos la historia del Procesamiento del Lenguaje Natural, sus comienzos y los hitos más importantes de su evolución hasta la actualidad. En la última parte de este informe **analizaremos, de forma práctica, un caso de NLP** (el análisis de comentarios ciudadanos) **aplicado a un conjunto de datos abiertos** con resultados muy interesantes. ¡Te esperamos en las próximas secciones!

## METODOLOGÍA

Este informe se enmarca dentro de una colección más amplia de recursos sobre tecnologías emergentes y datos abiertos, cuyo objetivo es **introducir en la materia al lector mediante el empleo de casos de uso prácticos, sencillos y reconocibles**. Al mismo tiempo, se pretende facilitar **una guía de aprendizaje práctica** para aquellos lectores con conocimientos más avanzados, que, mediante el desarrollo de un caso práctico, puedan experimentar de forma autodidacta con herramientas reales para el análisis y explotación de datos abiertos.

Para conseguir este doble objetivo, el informe se estructura en tres partes bien diferenciadas: *Awareness*, *Inspire* y *Action* (Figura 1), que pueden ser abordadas de forma independiente en cualquier momento y sin necesidad de haber realizado una lectura previa de las otras secciones.



**Figura 1.** Metodología de la colección Awareness, Inspire, Action.



La primera sección, **Awareness**, sirve de introducción al tema en cuestión (en este informe, el Procesamiento del Lenguaje Natural). Esta sección está indicada para aquellas personas que se inician en el tema por primera vez y tratan de abordar la temática de forma sencilla, clara y sin el uso de tecnicismos que dificulten la lectura.



La segunda sección, **Inspire**, pretende servir de inspiración a aquellas personas que se han iniciado en la materia y que se preguntan cómo les afecta a ellas en su vida diaria o en su trabajo el tema que se aborda. La forma de identificarse con una tecnología, un campo de la ciencia o cualquier otra materia es verse reflejado en ella. De esta forma, la sección *Inspire*, contiene ejemplos y casos de aplicación de una cierta tecnología en situaciones, más o menos, cotidianas que favorece que los lectores se identifiquen y comiencen a pensar en dicha tecnología como algo que también les afecta.



Por último, la sección **Action** selecciona alguno de los casos de usos explicados en la sección *Inspire* y lo desarrolla de forma práctica, utilizando para ello, datos y herramientas tecnológicas reales. El ejemplo, desarrollado en Action, se pone a disposición de las personas usuarias de esta guía en forma de código y datos abiertos (Anexo I) para que puedan experimentar y desarrollar con sus propios medios el caso de uso que se aborda en la sección *Action*.