



UNIVERSIDAD DE COLIMA
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA
Ingeniería en Computación Inteligente

Componentes en el PLN

Profesor: Carrillo Zepeda Oswaldo.
Alumno: Guerrero Gómez Brandon Yair.

6° B

Fecha de entrega: Jueves, 19 de mayo de 2022

¿Qué es el procesamiento de lenguaje natural?

El Procesamiento del Lenguaje Natural (PLN) es el campo de conocimiento de la Inteligencia Artificial que se ocupa de investigar la manera de comunicar las máquinas con las personas mediante el uso de lenguas naturales, como el español, el inglés o el chino.

Virtualmente, cualquier lengua humana puede ser tratada por los ordenadores. Lógicamente, limitaciones de interés económico o práctico hace que solo las lenguas más habladas o utilizadas en el mundo digital tengan aplicaciones en uso.

Pensemos en cuántas lenguas hablan Siri o Google Assistant. El inglés, español, alemán, francés, portugués, chino, árabe y japonés (no necesariamente en este orden) son las que cuentan con más aplicaciones que las entienden. Google Translate es la que más lenguas trata, superando el centenar... pero hay entre 5000 y 7000 lenguas en el mundo.

Las lenguas humanas pueden expresarse por escrito (texto), oralmente (voz) y también mediante signos. Naturalmente, el PLN está más avanzado en el tratamiento de textos, donde hay muchos más datos y son más fáciles de conseguir en formato electrónico.

Los audios, aunque estén en formato digital, hay que procesarlos para transcribirlos en letras o caracteres y, a partir de ahí, entender la pregunta. El proceso de respuesta es el inverso: primero se elabora la oración y luego se "sintetiza la voz". Cabe destacar que la voz artificial cada vez suena más humana, con inflexiones tonales y prosódicas que imitan la producción humana.

Modelos para procesamiento del lenguaje natural

Tratar computacionalmente una lengua implica un proceso de modelización matemática. Los ordenadores solo entienden de bytes y dígitos, y los informáticos codifican los programas empleando lenguajes de programación como C, Python, Java, etc.

Los lingüistas computacionales se encargan de la tarea de "preparar" el modelo lingüístico para que los ingenieros informáticos lo implementen en un código eficiente y funcional. Básicamente, existen dos aproximaciones generales al problema de la modelización lingüística:

Modelos Lógicos: gramáticas

Los lingüistas escriben reglas de reconocimiento de patrones estructurales, empleando un formalismo gramatical concreto. Estas reglas, en combinación con la información almacenada en diccionarios computacionales, definen los patrones que hay que reconocer para resolver la tarea (buscar información, traducir, etc.).

Estos modelos lógicos pretenden reflejar la estructura lógica del lenguaje y surgen a partir de las teorías de N. Chomsky en los años 50.

Modelos probabilísticos del lenguaje natural: basados en datos

La aproximación es a la inversa: los lingüistas recogen colecciones de ejemplos y datos (corpus) y a partir de ellos se calculan las frecuencias de diferentes unidades lingüísticas (letras, palabras, oraciones) y su probabilidad de aparecer en un contexto determinado. Calculando esta probabilidad, se puede predecir cuál será la siguiente unidad en un contexto dado, sin necesidad de recurrir a reglas gramaticales explícitas.

Es el paradigma de “aprendizaje automático” que se ha impuesto en las últimas décadas en Inteligencia Artificial: los algoritmos infieren las posibles respuestas a partir de los datos observados anteriormente en el corpus.

Componentes del procesamiento del lenguaje natural

A continuación, vemos algunos de los componentes del procesamiento del lenguaje natural. No todos los análisis que se describen se aplican en cualquier tarea de PLN, sino que depende del objetivo de la aplicación.

1. **Análisis morfológico o léxico:** Consiste en el análisis interno de las palabras que forman oraciones para extraer lemas, rasgos flexivos, unidades léxicas compuestas. Es esencial para la información básica: categoría sintáctica y significado léxico.
2. **Análisis sintáctico:** Consiste en el análisis de la estructura de las oraciones de acuerdo con el modelo gramatical empleado (lógico o estadístico).
3. **Análisis semántico:** Proporciona la interpretación de las oraciones, una vez eliminadas las ambigüedades morfosintácticas.
4. **Análisis pragmático:** Incorpora el análisis del contexto de uso a la interpretación final. Aquí se incluye el tratamiento del lenguaje figurado (metáfora e ironía) como el conocimiento del mundo específico necesario para entender un texto especializado.

Un análisis morfológico, sintáctico, semántico o pragmático se aplicará dependiendo del objetivo de la aplicación. Por ejemplo, un conversor de texto a voz no necesita el análisis semántico o pragmático. Pero un sistema conversacional requiere información muy detallada del contexto y del dominio temático.