

Jerónimo Molina Molina

Procesamiento del Lenguaje Natural

6. Tareas NLP

Como ya sabemos, el PLN incluye diferentes técnicas que se orientan a interpretar el lenguaje humano, las cuales van desde métodos estadísticos y aprendizaje automático hasta métodos basados en reglas y métodos algorítmicos.

Ya conocemos algunas tareas básicas de NLP, como los análisis PoS, por ejemplo, la lematización/derivación la detección del idioma o la identificación de relaciones semánticas - desambiguación-.

Se puede afirmar que, para procesar la información y el conocimiento, el PLN combina modelos de lingüística computacional y aprendizaje automático -ML y DL-, empleando técnicas de preprocesamiento, como la tokenización, derivación o lematización o eliminación de stopwords.

Pero más allá de lo que ya sabemos, en este tema vamos a centrarnos en la descripción de diferentes tareas NLP que son relevantes y en algunas librerías o servicios online que se pueden emplear a tal efecto.

Tareas NLP

Si hay tres tareas que han marcado una diferencia importante en el avance desde NLP a NLU, cabe destacar las siguientes:

- **Análisis sintáctico** o etiquetado PoS. Ya hemos explicado en qué consiste, y su relevancia resulta clave al profundizar en el entendimiento del lenguaje.
- **Análisis semántico**: Tiene como objetivo **encontrar significado a los mensajes**, y existen diferentes acciones a tener en consideración:
 - **Extracción de entidades** (NER), proceso sobre el que ya se ha profundizado y que consiste en identificar lugares, fechas y sujetos.
 - **Análisis del sentimiento** y detección de emociones
 - **Segmentación de temas**: Identificar los diferentes temas que se abordan en un texto
 - **Extracción de relaciones** o identificación de las relaciones entre las diferentes partes de un texto.
- **Análisis del discurso**: En esta tarea se analizan **estructuras de mayor tamaño**, con el objetivo de extraer una **visión o conocimiento más global del texto**.

Existen **otras muchas tareas** a tener en consideración, y que, individualmente, pueden generar valor añadido, tales como:

- **Text to Speech o Speech to Text**: convertir texto en mensajes de voz o viceversa. Este servicio lo usamos con frecuencia sin ser conscientes de ello.
- **Resumen o sumariazación de documentos**: Resulta de mucha utilidad
- **Traducción automática** y detección del idioma.

Pero,...

¿Qué hay de casos concretos o aplicaciones reales de PLN?

Bien, pasemos entonces a abordar diferentes casos de uso del Procesamiento o comprensión del lenguaje que puedan aplicarse de forma más o menos directa.

Casos prácticos básicos del PLN

Clasificación de textos.

Consiste en **asignarle una o más categorías a un texto**, de modo que posteriormente puedan realizarse búsquedas sobre estas categorías.

Igualmente entraría dentro de la clasificación de textos el **filtrado de correos electrónicos** (SPAM, Notificaciones, Promociones/Publicidad, etc...)

La clasificación de textos igualmente puede servir para categorizar el contenido en función del perfil de usuario que puede tener acceso (gestión, financiero, técnico, comercial...) Esto permite, por ejemplo, **proporcionar documentos a un usuario en función de su perfil**.

Análisis de sentimientos.

El análisis de sentimiento es una tarea que **devuelve un valor numérico**, en el intervalo [-1 y 1], que identifica el **sentido negativo (-1), neutro (0) o positivo (1)** de un texto.

Este ejercicio resulta de mucha utilidad en análisis de textos, y tiene numerosísimas aplicaciones, como por ejemplo el análisis del valor percibido de un producto por parte de los clientes o usuarios.

Detección de emociones.

Parecido al análisis de sentimientos, pero mucho **más detallado**, **identifica** en el texto diferentes **emociones** (duda, felicidad, amor, odio, inseguridad, temor, miedo...) permitiendo implementar filtros y clasificaciones emocionales de los mismos.

Este tipo de tareas cobran un valor importantísimo en el análisis de relaciones con clientes, por ejemplo. Permitiendo incluso el registro de la evolución emocional del cliente a lo largo de la relación comercial, posibilitando, entre otras cosas, un análisis en tiempo real en el caso de chats o videoconferencias, y proporcionando una visión de alto nivel en mensajes de texto.

Detección de similitud entre textos.

Es una tarea que consiste en **analizar** el **nivel de semejanza** o diferencia **entre dos frases, textos o documentos**, y en esta tarea están basados los detectores de plagio, por ejemplo.

Respuestas a preguntas.

Es la **capacidad de un sistema NLP** de **responder a preguntas en lenguaje natural**. Existen diferentes formas de enfocar estas aplicaciones, si bien actualmente se imponen los modelos grandes del lenguaje, como ChatGPT de OpenAI.

Traducción automática.

Permite **traducir un texto de un idioma a otro**. Es una tarea que lleva perfeccionándose mucho tiempo.

Actualmente la traducción simultánea, implementada en lo que se denomina "Edge computing", permite disponer de auriculares que traducen en tiempo real un mensaje de un idioma a otro, facilitando de este modo la comunicación.



Generación de texto.

El nivel de desarrollo de esta actividad es muy alto, y actualmente son muchos los sistemas que generan textos que bien podrían pasar por textos escritos por un humano.

redacta un texto de estilo periodístico que explique la experiencia de OESIA en ciberseguridad, omitiendo detalles técnicos



OESIA, una reconocida empresa especializada en soluciones tecnológicas, ha destacado recientemente por su experiencia en el campo de la ciberseguridad. A medida que los ciberataques se vuelven más frecuentes y sofisticados, OESIA ha demostrado estar a la vanguardia en la protección de las organizaciones contra estas amenazas digitales.

La preocupación por la seguridad informática ha aumentado en los últimos años, y OESIA ha sabido adaptarse a esta creciente demanda. Con un enfoque orientado a los objetivos de sus clientes, la empresa ha investigado y desarrollado las mejores posibilidades y opciones científico-tecnológicas para combatir los ciberataques.

Una de las fortalezas de OESIA en este campo es su capacidad para analizar e implementar sistemas específicos de análisis de datos. Mediante el uso de técnicas avanzadas de análisis, OESIA ha logrado identificar patrones y comportamientos sospechosos en las redes de sus clientes, permitiendo una detección temprana de posibles amenazas.

Además, OESIA ha diseñado un conjunto de agentes inteligentes, conocidos como Smart Agents, capaces de tomar decisiones por sí mismos para mejorar los resultados presentados al usuario. Estos agentes inteligentes han demostrado ser una herramienta eficaz en la lucha contra los ciberataques, ya que pueden adaptarse y responder rápidamente a las nuevas amenazas que surgen en el entorno digital.

La experiencia de OESIA en ciberseguridad ha sido ampliamente reconocida en el mercado. Grandes organizaciones confían en sus soluciones para proteger sus sistemas y datos sensibles. OESIA ha logrado resolver los problemas que el mercado actualmente enfrenta en términos de gestión del conocimiento y explotación de modelos de lenguaje a gran escala.

En resumen, OESIA se ha posicionado como un referente en el campo de la ciberseguridad, ofreciendo soluciones innovadoras y

Resumir textos.

Esta es otra **tarea muy relevante** que **actualmente realizan** con un **nivel muy próximo a la perfección** los modelos grandes del lenguaje.

Combinar esta tarea con Speech To Text da resultados asombrosos que pueden ahorrar bastante trabajo, por ejemplo, elaborando resúmenes de reuniones. Otra aplicación es la presentación breve de un documento que se puede proponer para descarga.

Simplificación de textos.

Consiste en **elaborar un texto**, **equivalente a otro dado**, pero **explicando de forma más sencilla** aquellas partes que presenten dificultades para su comprensión.

Librerías de alto nivel que se aplican en PLN

Reconocimiento del Lenguaje Conversacional de Azure

<https://azure.microsoft.com/es-es/products/ai-services/conversational-language-understanding>

Microsoft lleva ya mucho tiempo trabajando en el ámbito del **reconocimiento del lenguaje**, y ya tiene lista su nueva generación de soluciones, se trata de su suite para el reconocimiento del lenguaje, con soluciones tan eficientes como las siguientes.



Microsoft permite una amplísima variedad de tareas, desde el **análisis de sentimientos** a la **traducción**, **sumarización**... Merece mucho la pena detenerse a consultar las diferentes posibilidades. Además, ofrecen una documentación amplísima, que permite explotar los servicios sin problema, además de un paquete de prueba gratuito.

También dispone de un workflow de orquestación que permite diseñar procesos NLU basados en diferentes soluciones de las que ofrecen.

API de Lenguaje Natural de Google

https://cloud.google.com/natural-language?hl=es_419

Al igual que Microsoft, Google no se ha quedado atrás, y ofrece un abanico de soluciones empresariales basado en NLP sorprendente

Tres soluciones de lenguaje natural que funcionan con tu texto

AutoML

Entrena tus propios modelos personalizados de aprendizaje automático de alta calidad para clasificar, extraer y detectar opiniones con el mínimo esfuerzo y sin la necesidad de contar con experiencia en el área gracias a Vertex AI para lenguaje natural, con tecnología de [AutoML](#). Puedes usar la IU de [AutoML](#) para subir tus datos de entrenamiento y probar tu modelo personalizado sin una sola línea de código.

API de Natural Language

Los modelos potentes y previamente entrenados de la [API de Natural Language](#) les permiten a los desarrolladores aplicar con facilidad la comprensión del lenguaje natural (CLN) a sus aplicaciones mediante funciones como el análisis de opiniones, de entidades, de opiniones por entidad y de sintaxis, y la clasificación de contenido.

IA de Healthcare Natural Language

Obtén análisis en tiempo real de las estadísticas almacenadas en textos médicos no estructurados. La [API de Healthcare Natural Language](#) te permite sintetizar estadísticas médicas procesables de documentos médicos, y AutoML Entity Extraction for Healthcare simplifica la creación de modelos de extracción de conocimiento personalizados para las apps de salud y ciencias biológicas, sin necesidad de tener habilidades de programación.

Ejercicio propuesto

Ejercicio de negocio

En este caso, se propone que analices de manera minuciosa las diferentes soluciones NLP que ofrecen Microsoft, Google y Amazon, y hagas una tabla comparativa de los diferentes servicios que ofrece cada uno.

Termina el informe con una memoria explicativa que indique en qué destaca cada uno de los proveedores.

Ejercicio técnico

Elige un proveedor (Amazon, Microsoft o Google), genera una cuenta gratuita y elige un servicio entre:

- Sumarización

- Análisis de sentimiento
- Análisis NER

Implementa una aplicación Python, que, dado un texto, realice la función seleccionada. Muestra el resultado.

Bibliografía para completar el aprendizaje

1. <https://azure.microsoft.com/es-es/products/ai-services/conversational-language-understanding>
2. https://cloud.google.com/natural-language?hl=es_419
3. <https://planetachatbot.com/33-tareas-nlp/>