

Inteligencia Artificial

Unidad 4: Procesamiento de Lenguaje Natural

TEMA 4: Algoritmos de IA Moderna-II

Módulo 1: Introducción al Procesamiento del Lenguaje Natural



Profesora

Ing. Patricia Reyes Silva

Unidad 4

Procesamiento de Lenguaje Natural

TEMA 4: Algoritmos de IA Moderna-II

Sesión 22

MÓDULO 1: Introducción al Procesamiento de Lenguaje Natural



Contenido

1. ¿Qué es el procesamiento del lenguaje natural (NLP) ?
2. Componentes del NLP
3. Técnicas - Pasos para comprender el NLP
4. CASOS DE USO: ¿Dónde se aplica NLP?
5. Herramientas disponibles para NLP



Preguntas

1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

LENGUAJE NATURAL

LENGUAJE = Es el medio hablado y/o escrito por humanos para propósitos generales de comunicación.

Ejemplos:

- Las lenguas (ingles, español, chino, ruso, etc.).
- Los lenguajes de programación.
- Los lenguajes usados en el estudio de la lógica formal: la lógica matemática.

NATURAL = Adquirido de manera espontanea (lenguas maternas) con la finalidad de comunicarnos.



1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

¿Qué es el NLP? Sus aplicaciones



Video: <https://youtu.be/nn-yhjVJZQE> (Duración: 3 min)

1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

¿POR QUE PROCESAR EL LENGUAJE NATURAL?

Porque la IA requiere entender muy bien cada lenguaje natural para poder:

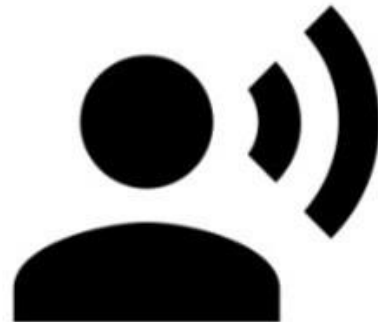
- ❖ Crear sistemas como reconocimiento de voz.
- ❖ Generalizar documentos.
- ❖ Traducir automáticamente.
- ❖ Detectar Spam.
- ❖ Reconocer entidades con nombre.
- ❖ Ofrecer respuestas a preguntas.
- ❖ Generar autocompletado de sentencias.
- ❖ Ofrecer el ingreso de texto predictivo, etc.



1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

Procesamiento del lenguaje natural (en adelante, NLP):

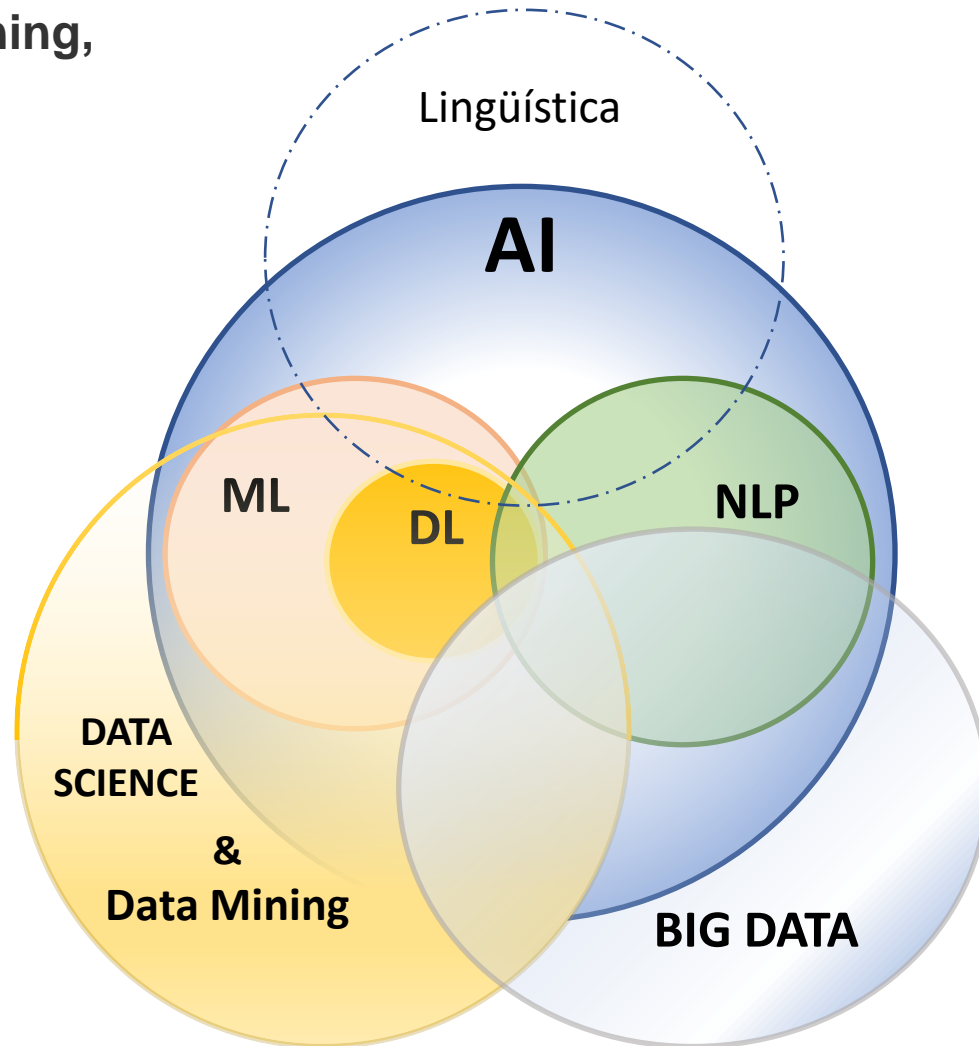
- El **procesamiento del lenguaje natural** es una subsección de la informática y la inteligencia artificial dedicada a cómo las computadoras analizan los lenguajes naturales (humanos).
- **NLP** permite el uso de algoritmos de aprendizaje automático para texto y voz.



1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

IA, NLP, Machine Learning, Deep Learning, Big Data y Data Science

- **NLP** aprovecha las herramientas, técnicas y algoritmos de IA y aprendizaje automático (ML) para comprender los datos del lenguaje natural no estructurados y entender su significado.
- **NLP** permite el uso de algoritmos de aprendizaje automático (ML) tanto para texto y voz.

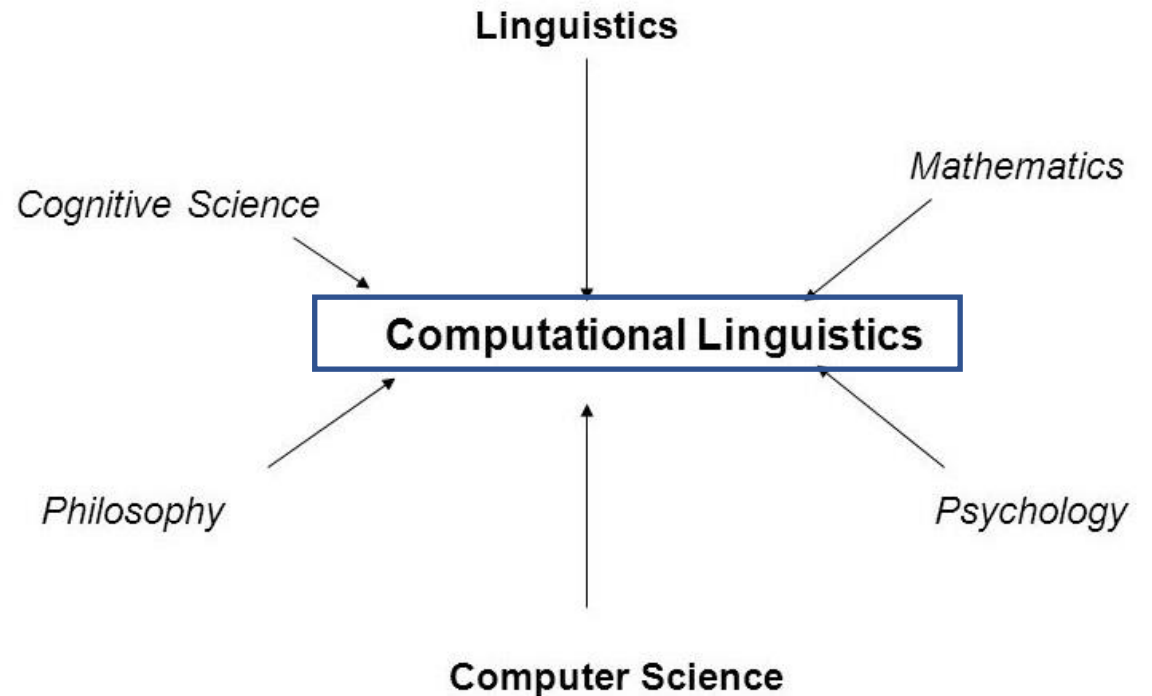


AI: Inteligencia Artificial
ML: Machine Learning
DL: Deep Learning
NLP: Procesamiento de Lenguaje Natural
BD: Big Data
DS y DM: Data Science y Data Mining
Lingüística

1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

LINGÜÍSTICA COMPUTACIONAL

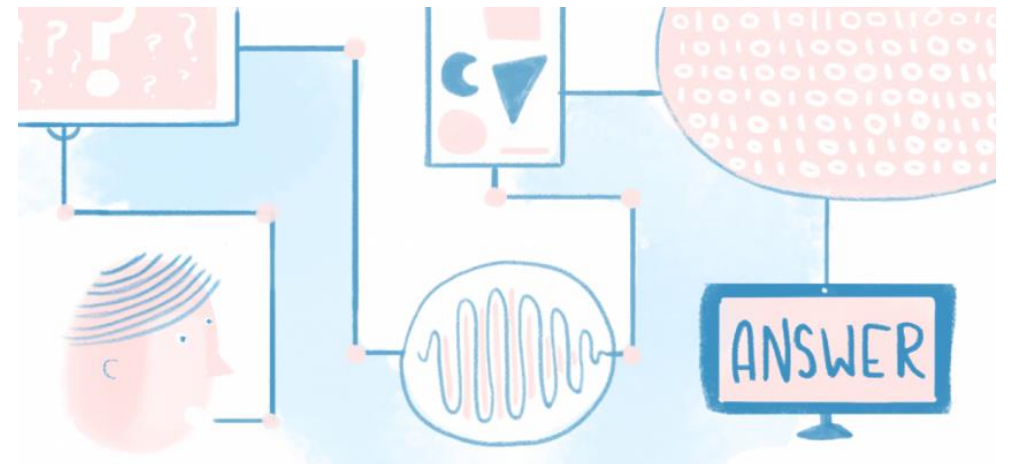
- Campo multidisciplinar de la lingüística y la informática.
- Utiliza la informática para estudiar y tratar el lenguaje humano (análisis, síntesis y comprensión del lenguaje escrito y hablado).
- Participan lingüistas, informáticos especializados en inteligencia artificial, psicólogos cognoscitivos y expertos en lógica, entre otros.



1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

LINGÜÍSTICA COMPUTACIONAL: Objetivos Comerciales

- ❖ Traducir texto de un idioma a otro.
- ❖ Recuperar texto relacionado con un tema específico.
- ❖ Analizar texto o lenguaje hablado en busca de contexto, sentimiento u otras cualidades afectivas.
- ❖ Responder preguntas, incluidas aquellas que requieran inferencia y respuestas descriptivas o discursivas.
- ❖ Resumen de texto.
- ❖ Construir agentes de diálogo capaces de realizar tareas complejas como realizar una compra, planificar un viaje o programar un mantenimiento.
- ❖ Creación de chatbots capaces de pasar la prueba de Turing .



1. ¿Qué es el procesamiento del lenguaje natural (NLP)?

Google Duplex 2018 - Sistema de reconocimiento de voz



Video: <https://youtu.be/904c2RB1BJo> (Duración: 5 min)

2. Componentes del NLP

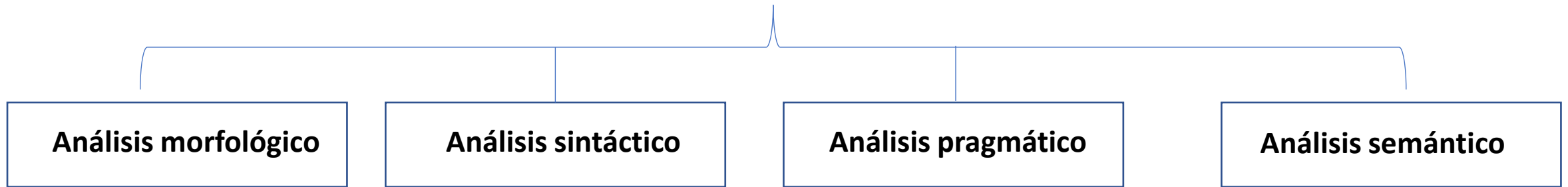
COMPONENTES DEL NLP

Análisis morfológico

Análisis sintáctico

Análisis pragmático

Análisis semántico



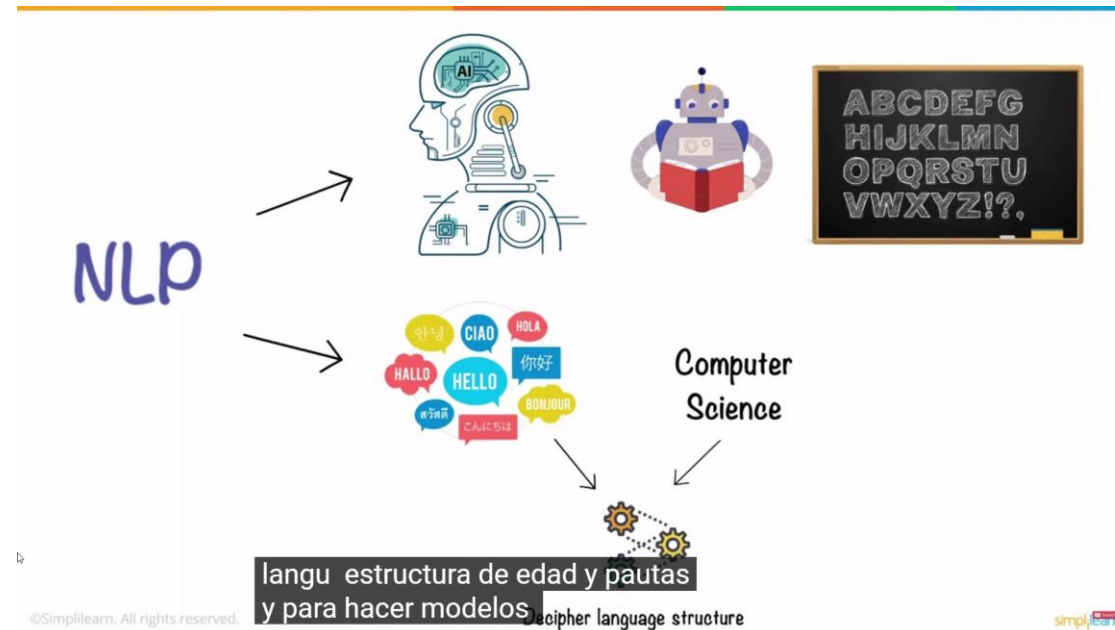
3. Técnicas - Pasos para comprender el NLP

Ejemplo de procesamiento de un texto

Milan is a city in northern Italy, capital of Lombardy, and the second-most populous city in Italy after Rome. Milan served as the capital of the Western Roman Empire, the Duchy of Milan and the Kingdom of Lombardy–Venetia. The city proper has a population of about 1.4 million while its metropolitan city has 3.26 million inhabitants. Its continuously built-up urban area, that stretches well beyond the boundaries of the administrative metropolitan city, is the fourth largest in the EU with 5.27 million inhabitants.

3. Técnicas - Pasos para comprender el NLP

¿En qué consiste el Procesamiento del Lenguaje Natural? Técnicas o pasos en NLP

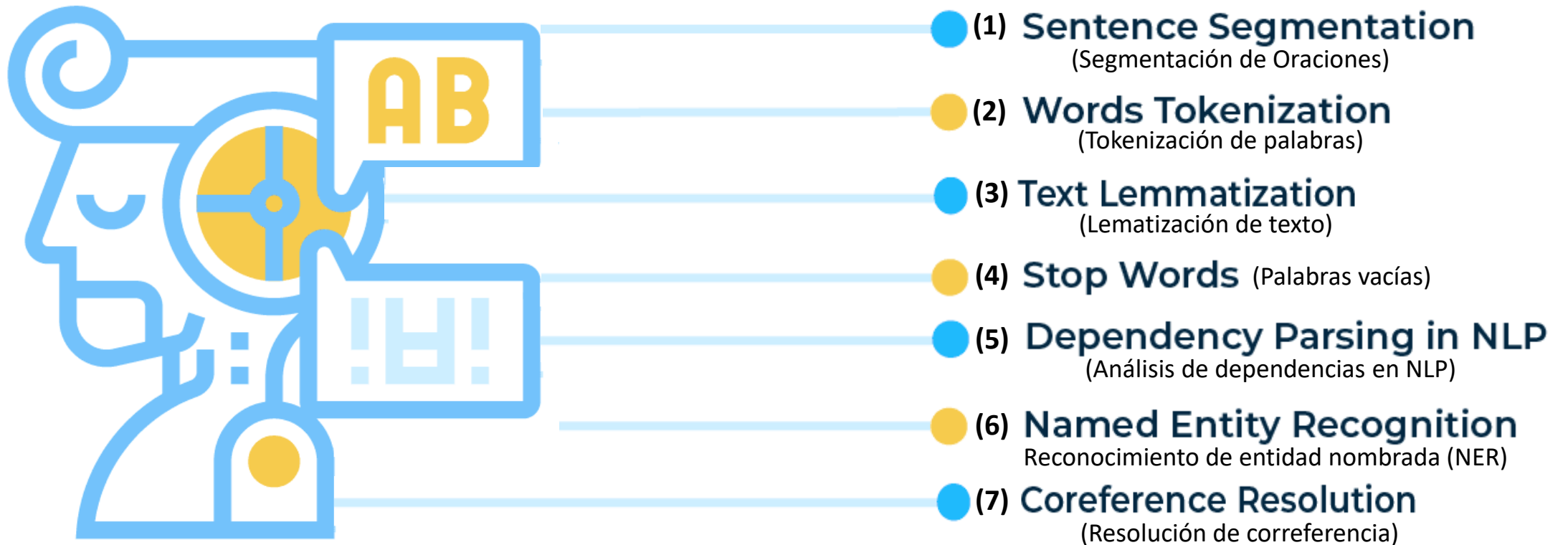


Video: <https://youtu.be/yx-YgLMCxZM> (Duración: 4 min)

3. Técnicas - Pasos para comprender el NLP

Pasos para comprender el lenguaje natural

- Existen siete (7) pasos básicos que los científicos de datos emprendemos para derivar el significado del texto o de cualquier otro dato no estructurado:



3. Técnicas - Pasos para comprender el NLP

(1) Sentence Segmentation (Segmentación de Oraciones)

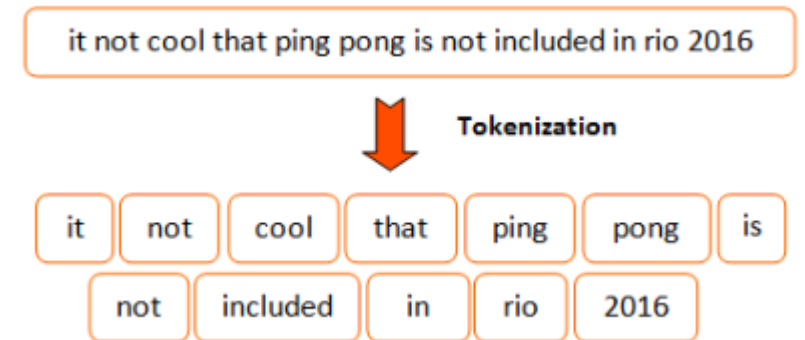
- El primer paso en el procesamiento del lenguaje natural es dividir las oraciones en objetos separados.
- Esta etapa es bastante fácil.
 - Un algoritmo de IA inteligente analiza los conjuntos de datos y define los signos de puntuación.
 - Cada vez que nota un punto, da por finalizada la frase y la separa del texto completo.
- Esta etapa es importante ya que permite al modelo de NLP derivar el significado de la oración y luego pasar al análisis de todo el párrafo.
- Dividir el texto en oraciones puede ser muy sencillo cuando los datos vienen en un formato más o menos estructurado. Sin embargo, la información puede presentarse sin signos de puntuación o sin otros elementos del texto. En tales casos, se aplican técnicas complejas para identificar partes significativas.



3. Técnicas - Pasos para comprender el NLP

(2) Words Tokenization (Tokenización de palabras)

- Una vez que hayamos dividido el texto en oraciones, es hora de separar las palabras y determinar sus partes del discurso.
- En inglés, esto es fácil de hacer identificando espacios entre las palabras o tokens (los signos de puntuación también se consideran tokens separados, ya que tienen ciertos significados y pueden cambiar la idea completa del texto).
- El siguiente paso en NLP es mirar cada token por separado y definir su parte del discurso.
- Los algoritmos de IA analizan cada palabra y aplican un cierto conjunto de criterios para clasificarla en adjetivos, sustantivos, verbos, etc. (esto ayudará a una máquina a comprender el papel de cada token en la oración o el texto).
- Para este propósito, se utiliza un modelo de clasificación de partes del discurso previamente entrenado. Este modelo ha sido entrenado procesando millones de textos en inglés previamente etiquetados y marcados para proporcionar a los algoritmos datos esenciales.
- Analiza grandes conjuntos de datos, lo que le ayuda a desarrollar estadísticas que se utilizan además para definir a qué parte del discurso pertenece una palabra.



3. Técnicas - Pasos para comprender el NLP

(3) Text Lemmatization (Lematización de texto)

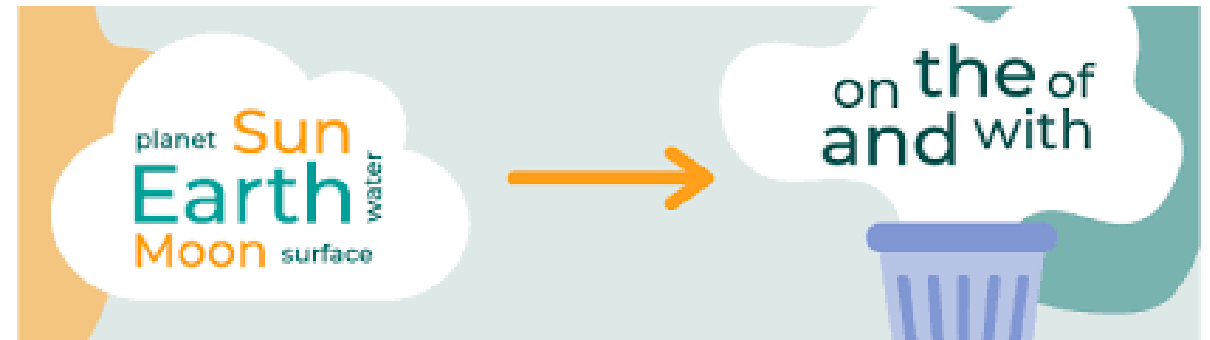
- La mayoría de los textos y oraciones contienen palabras raíz, así como palabras con diferentes formas gramaticales.
- El procesamiento del lenguaje natural se utiliza aquí para ayudar a la máquina a identificar el significado y categorizar estas palabras.
- Por ejemplo, es posible que se vean las palabras "población" y "poblado" en el mismo texto. Aunque pertenecen a diferentes partes del discurso, el significado de estas palabras es bastante similar.
- Los modelos de NLP se aplican aquí para averiguar el "lema" de cada token, que es la forma básica de cada palabra. Este paso ayuda a un sistema de inteligencia artificial a comprender el concepto central del texto.



3. Técnicas - Pasos para comprender el NLP

● (4) Stop Words (Palabras vacías)

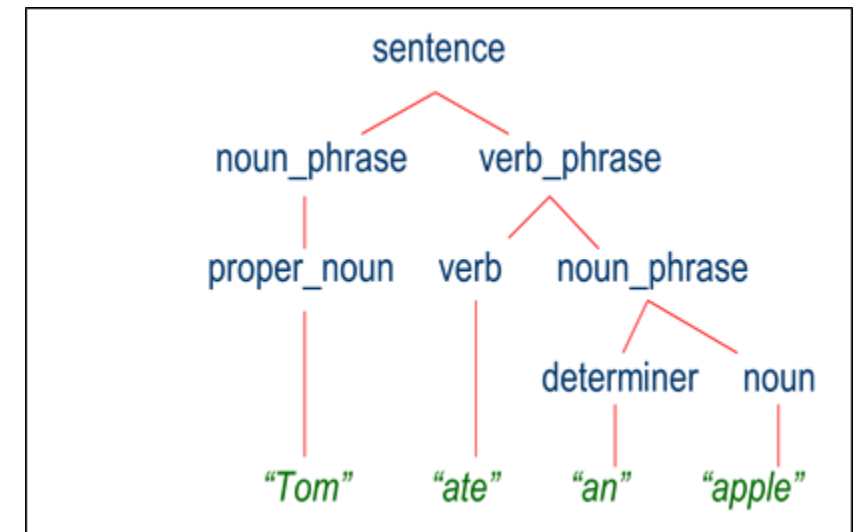
- El siguiente paso esencial en el procesamiento del lenguaje natural es identificar las **palabras vacías** y filtrarlas antes de decodificar el significado central del texto.
- Cada idioma tiene varios enlazadores y palabras de "relleno" que no agregan ningún significado adicional al texto, pero aparecen con frecuencia en el habla o en un texto escrito de manera informal.
- Estos objetos pueden producir un tipo de ruido que dificultará que un sistema de NLP obtenga información a partir de los datos.
- Por lo tanto, las canalizaciones de NLP generalmente marcan estos tokens como "**palabras vacías**" y las omiten al analizar su texto o cualquier otro dato.



3. Técnicas - Pasos para comprender el NLP

(5) Dependency Parsing in NLP (Análisis de dependencias en NLP)

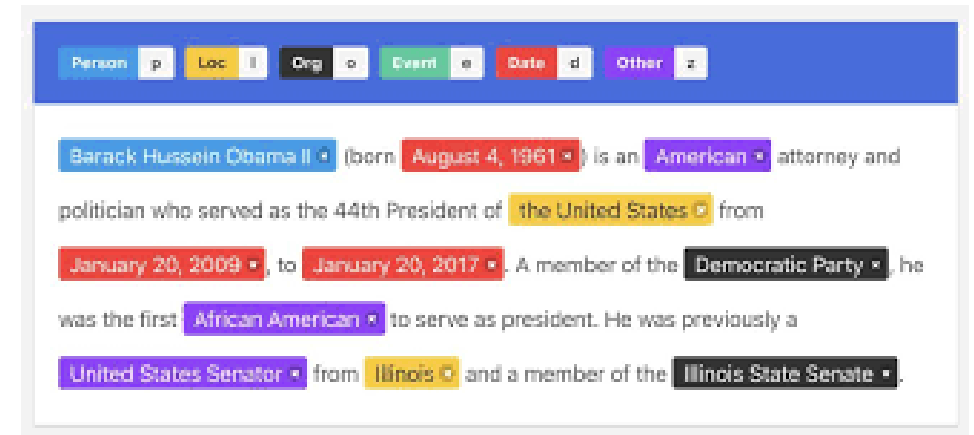
- La tarea principal en esta etapa es descubrir las relaciones entre todas las palabras de un texto.
- Por esta razón, los algoritmos NLP construyen un árbol de análisis que define la palabra raíz en la oración y cierra la brecha entre otros tokens.
- También pueden definir una palabra madre para cada token para obtener más información y, posteriormente, comprender el concepto central.



3. Técnicas - Pasos para comprender el NLP

(6) Named Entity Recognition Reconocimiento de entidad nombrada (NER)

- Cuando ya se tiene un árbol de análisis listo para usar, es hora de pasar al **Reconocimiento de entidad con nombre (NER)**.
- En esta etapa del procesamiento del lenguaje natural, los científicos de datos comenzamos a extraer ideas del texto relacionando tokens con objetos de la vida real.
- Si miramos nuestros ejemplos de texto, podemos ver que las palabras “Milán” e “Italia” se refieren a lugares geográficos, mientras que “Imperio Romano Occidental” es el nombre histórico.
- Al extraer esta información, un modelo de NLP obtiene un significado adicional del texto que puede usarse para realizar un análisis exhaustivo.



3. Técnicas - Pasos para comprender el NLP

● (7) Coreference Resolution (Resolución de coreferencia)

- Concluido el Reconocimiento de entidades nombradas, tenemos mucha información al alcance de la mano:
 - hemos dividido el texto en oraciones y palabras,
 - hemos derivado su significado y
 - hemos construido relaciones entre los objetos principales del texto.

Pero, ¿qué hacemos con las coreferencias?



3. Técnicas - Pasos para comprender el NLP

(7) Coreference Resolution (Resolución de correferencia)

Continuación...

- Todavía tenemos un obstáculo que impide que nuestro modelo de NLP comprenda completamente el lenguaje natural:

Cada idioma tiene muchas entidades, como pronombres y otras partes del habla, que pueden relacionarse con otra palabra en la oración y ampliar su significado.

- La resolución de correferencia se realiza para agrupar todas las menciones en el texto que se refieren a un concepto o entidad de la vida real. Por lo tanto, un modelo de NLP comprenderá a qué se refieren palabras como "él", "su" o "así".
- Para comprender mejor cómo funciona la resolución de correferencia, visitar: <https://huggingface.co/coref/>

4. CASOS DE USO: ¿Dónde se aplica NLP?

CASOS DE USO - NLP



Chatbots

- Las empresas de todos los dominios utilizan chatbots para mejorar la experiencia del cliente y analizar los comentarios de los clientes.



Sentiment Analysis

- Se aplica para interpretar y clasificar emociones en base a resúmenes de texto disponibles, comentarios, etc.



Marketing

- Con chatbots simples y pasando a redactores publicitarios inteligentes de IA que generan lemas, los modelos de NLP facilitan el trabajo de los especialistas en marketing.



Banking

- Ayuda a los bancos a mitigar los riesgos, automatizar los procesos comerciales o verificar la calidad de los servicios al cliente.



Fake News Detection

- Se aplica con frecuencia para detectar información falsa y proporcionar estadísticas sobre su exposición.



Healthcare

- Ayuda a los trabajadores médicos a procesar los datos de los pacientes, mejorar la calidad de la atención médica, y a identificar a los pacientes que necesitan atención especial.

PREGUNTAS

Dudas y opiniones