## Analítica de textos, análisis y analítica de textos, búsqueda y recuperación de información y NLP.

Texto 🡪 tipo de dato no estructurado 🡪 su estructura no es clara es impredecible.

Los datos no estructurados proporcionan información que no está organizada con un modelo de datos ya establecido, por lo que no se puede almacenar en una base de datos tradicional (RDBMS). La multimedia y el texto son dos tipos comunes de contenido no estructurado.

**¿Cómo analizamos un texto?** Transformaremos el texto en esquemas manejables por los ordenadores como vectores, matrices o estructuras similares.

* **Analítica de textos 🡪** obtiene información cuantitativa mediante el texto no estructurado.
* **Minería o análisis de texto 🡪** obtiene información cualitativa mediante el texto no estructurado.

**Búsqueda y recuperación de información.**

Encontrar material no estructurado que satisfaga una necesidad de información en grandes colecciones, en big data, los datos no estructurados deben almacenarse en bases de datos NoSQL o datalakes y procesarse de manera diferente a los datos estructurados, usando algortimos especiales para ellos.

El procesamiento del lenguaje natural es una rama de la inteligencia artificial que ayuda a los ordenadores a comprender y manipular el lenguaje humano.

Usa disciplinas como:

* Ciencia de computación.
* Lingüística computacional.

La comprensión del lenguaje natural ayuda a las máquinas a leer texto o habla simulando la capacidad humana para entender un idioma natural (Siri, Alexa y la búsqueda por voz de Google que utilizan NLP para comprender y responder a las solicitudes de los usuarios.)

También se ha desarrollado aplicaciones sofisticadas de minería de texto en campos diversos como la investigación médica, la gestión de riesgos, la atención al cliente, los seguros y la publicidad contextual.

## Tipos de tareas y aplicaciones

Podemos realizar acciones sobre un conjunto de documentos o textos.

* Indexación de documentos.
* Resumen de documentos o un grupo de ellos.
* Etiquetado.
* Recuperar documentos o textos de una base de datos que cumpla con nuestra información.
* Clasificación de documentos en varios grupos.
* Búsqueda de los grupos o documentos de agrupaciones.
* Entender el sentimiento de un texto y contexto.
* Procesamiento avanzado del lenguaje natural (NLP).
* Traducir a otro idioma.
* Comprender y responder preguntas.