1. **Analítica y análisis de textos, búsqueda y recuperación de información**

Texto y multimedia → tipo de dato no estructurado → estructura impredecible.

``Datos no estructurados" son datos que no tienen una estructura clara, semánticamente abierta y fácil de usar. Por ello no se pueden almacenar en una base de datos tradicional o RDBMS. Casi ningún dato es realmente ``no estructurado´´.

Para poder analizar un texto tenemos que transformarlo en vectores, matrices o cosas parecidas (lenguaje que entienda el ordenador).

Podemos hacer algunas distinciones →

* La analítica de textos → proceso de análisis de texto no estructurado, se coge info importante, proporciona información cuantitativa.
* Minería o análisis de texto → se obtiene información cualitativa mediante texto no estructurado, proporciona información cualificada.

Búsqueda y recuperación de información; es encontrar material (datos no estructurados) para satisfacer unas necesidades.

**El proceso del lenguaje natural (NLP)**

Es una rama de la IA que ayuda a los ordenadores a entender, interpretar y manipular el lenguaje humano.

1. **Tipos de tareas y aplicaciones**

Podemos realizar varias 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.
* Clasificación de documentos en varios grupos.
* Búsqueda de los grupos o documentos de agrupación.
* Comprender el sentimiento de un texto y contexto.
* Procesamiento avanzado del lenguaje natural (NLP).

Traducir a otro idioma, comprender y responder preguntas.

* Gestión de riesgos.
* Comentarios de los clientes, escucha social y aplicaciones centradas en las personas.
* Publicidad mediante medios digitales.
* Enriquecimiento de contenido.
* Filtrado de spam.
* Prevención de delitos cibernéticos.

1. **Clasificación de textos, análisis de sentimientos y recuperación de información. Clasificador de Bayes.** 

