

MINERÍA DE TEXTOS

Salvador López Mendoza

Mayo de 2018

DEFINICIONES

- La *minería de textos* también es conocida como *análisis de textos*.
- Objetivo: Convertir datos textuales en **información de calidad** o **conocimiento accionable**.
 - Minimizar el esfuerzo humano al consumir datos textuales.
 - Proporcionar conocimiento para la toma de decisiones óptima.
- La minería de textos se relaciona con la *recuperación de textos*.
Preprocesador para la minería de textos.

DATOS TEXTUALES VS. DATOS NO TEXTUALES

Datos no textuales:

- Los datos no textuales generalmente son el resultado de utilizar algún tipo de *sensor* ante cosas que ocurren en el mundo real.
- El sensor *reporta* los datos observados.

Datos textuales:

- Los datos textuales son lo que percibe un ser humano (*sensor humano*).
- El ser humano *expresa* lo que ha percibido.

En la minería de textos se utilizan también datos no textuales para generar el conocimiento que permita realizar algún tipo de acción.

TAREAS DE LA MINERÍA DE TEXTOS

- ① Minar conocimiento acerca del lenguaje.
- ② Minar el contenido de los datos textuales.
- ③ Minar el conocimiento acerca del observador.
- ④ Inferir otras variables del mundo real.

TÓPICOS A TRATAR

- ➊ Procesamiento de lenguaje natural y representación de textos.
- ➋ Minado de asociación de palabras.
- ➌ Minado de tópicos.
- ➍ Minado de opiniones y análisis de sentimientos.
- ➎ Predicción basada en textos.