



Curso:
Big Data
Unidad II: Minería de Texto.
Clase 1: Introducción al Text Mining

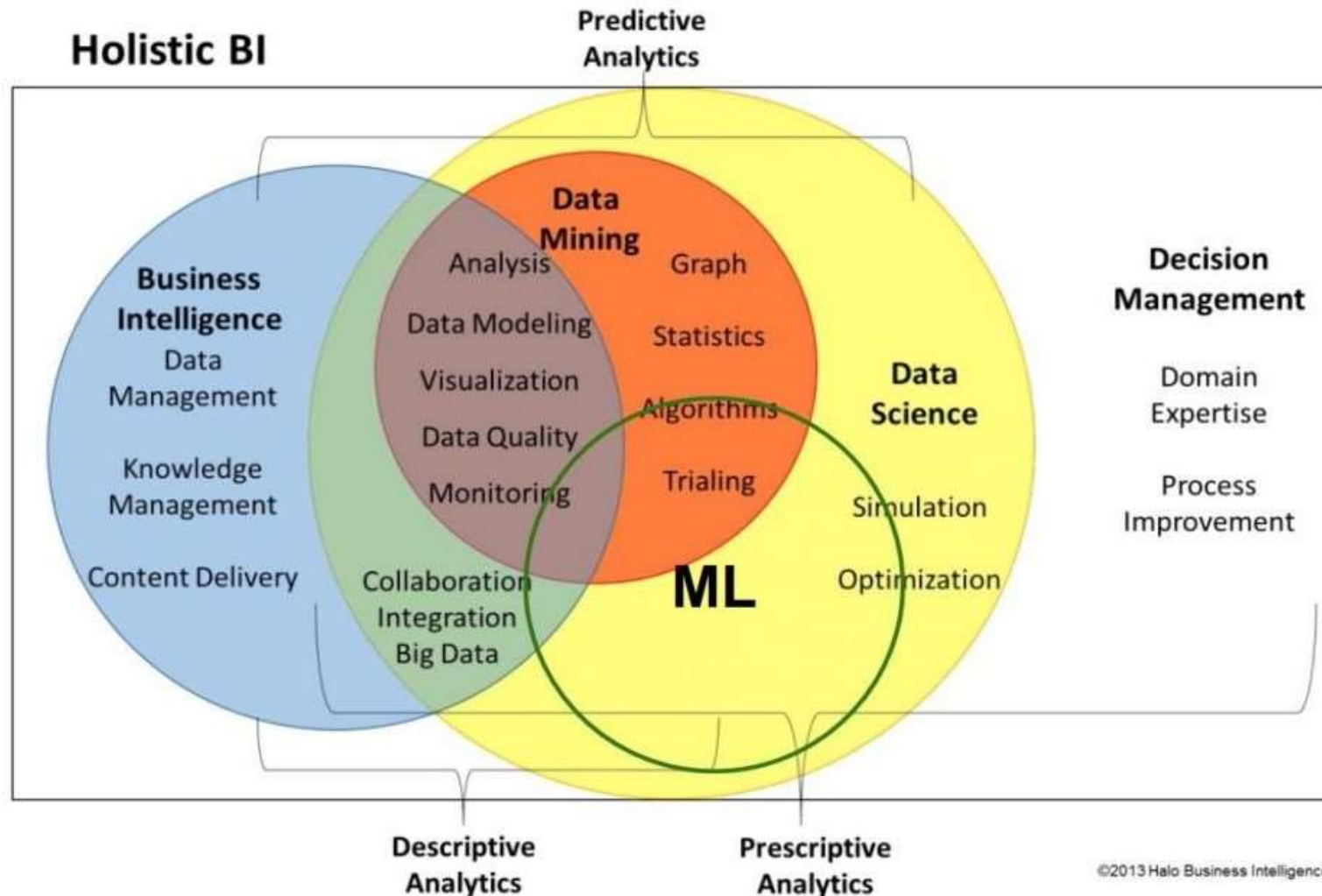
Profesor: Diego Miranda O.

Data Scientist

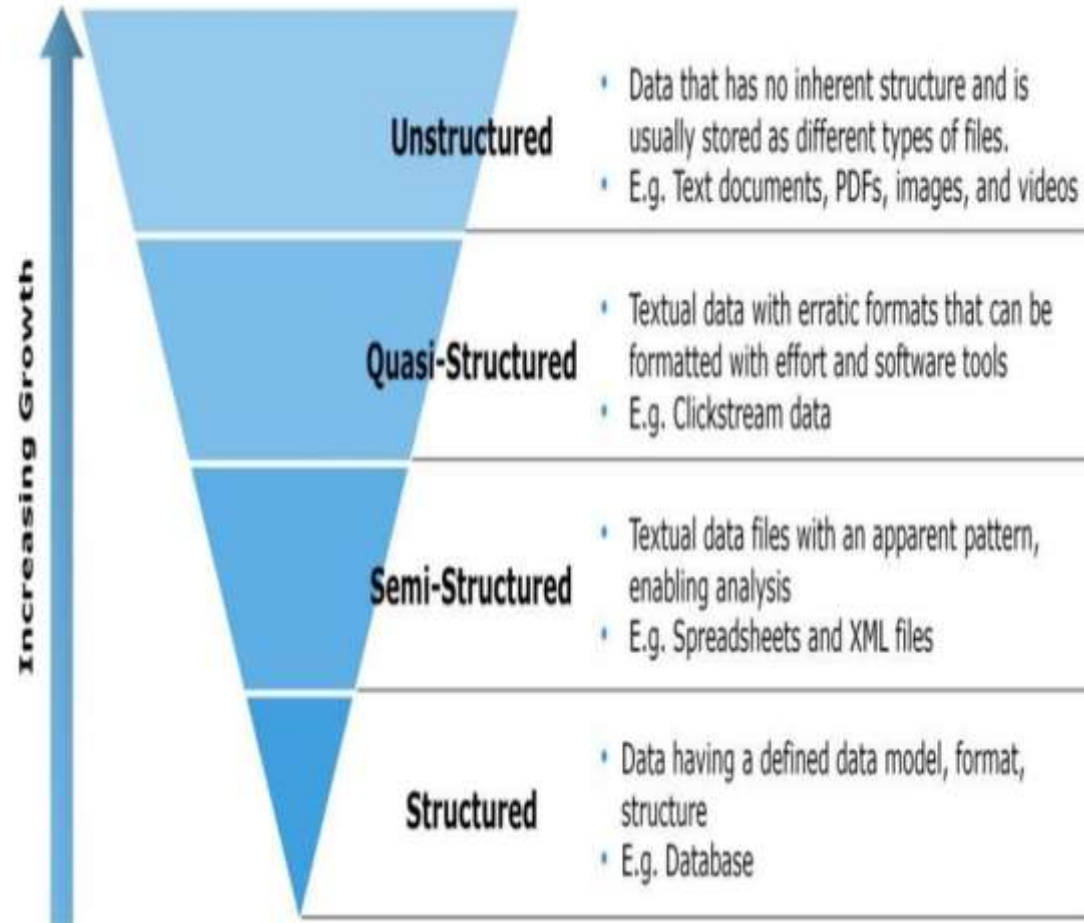
Motivación

Mucho del conocimiento en las organizaciones no es accesible a través de técnicas tradicionales de analítica de datos pues los datos están en forma de texto libre, o sea, sin estructura.

Data Science



Tipos de datos



Tipos de Datos

- ***Datos estructurados:*** corresponden a datos organizados en repositorios tales como una base de datos, de modo que sus elementos pueden ser accedidos por métodos de análisis y procesamiento efectivos (por ejemplo, una tabla SQL).
- ***Datos no-estructurados:*** corresponden a datos sin una estructura o modelo predefinido o que no está organizado de forma predefinida, por lo que se hace difícil entenderlos usando métodos computacionales tradicionales (por ejemplo, textos de noticias o quejas de clientes).

Fuentes Textuales: Noticias



Mundo » España | Rusia | China | Medio Oriente | África | Asia | Europa

Qatar 2022: denuncia por presunto maltrato a personas LGBTQ 2:12

Doha, Qatar (CNN) -- La Copa del Mundo está en marcha en Qatar, pero los problemas relacionados con [los derechos LGBTQ+ para el Estado del Golfo](#), el organismo rector del fútbol mundial FIFA, los equipos y los hinchas simplemente no desaparecen.

Este sábado, dos seguidores del fútbol alemanes le dijeron a CNN que funcionarios de seguridad en Qatar les pidieron que se quitaran los artículos con los colores del arcoíris que llevaban puestos mientras se dirigían a ver el partido de la Copa Mundial entre Francia y Dinamarca el sábado.

CNN fue testigo de la conclusión del incidente en la estación de metro de Msheireb, en Doha, cuando [Bengt Kunkel](#), que llevaba una banda con los colores del arcoíris, y su amigo, que lucía un brazalete de color similar, se negaron a entregar los artículos. El arcoíris es un símbolo del orgullo LGBTQ+.

Fuentes Textuales: Reviews



Adriaan



Comprehensive Introduction to the circular economy

Calificado en Estados Unidos us el 22 de junio de 2022

Compra verificada

I found this handbook to provide an excellent overview and serve as a reference tool as I studied the Circular Economy Handbook. It is a great book to anyone who wants to get an overview of the principles driving circular thinking, looking for ideas on how to approach the circular economy or in need to a framework to approach the topic.

Útil

Informar de un abuso

[Traducir opinión al idioma Español](#)

De otros países



PeterD



An Essential Guide to the Circular Economy

Calificado en Reino Unido gb el 23 de noviembre de 2020

Compra verificada

The 1st edition of "A Circular Economy Handbook for Business and Supply Chains" (published in 2017) was the best guide to the circular economy; the 2nd edition attains the same standard and goes beyond.

Catherine has developed a new Circular Economy Framework 2.0 and the book provides a deep dive into a variety of ways to explain and demonstrate this new framework. The case studies are broad in scope and give some very useful examples from the real world of circularity. We are seeing more and more every day, the negative impacts of the climate crisis on our planet. Catherine highlights the real effect that circular practices can have on mitigating the problems we are facing.

Fuentes Textuales: Quejas


Cliente: "Ya intenté todas las soluciones que me proporcionaron, pero sigo sin conexión, ¿qué pueden hacer al respecto?"

Representante: "Al no haber tenido éxito con ninguna de las soluciones, tendremos que hacer un cambio en su equipo electrónico. Comprendemos que esto puede ser un poco molesto, pero como una manera de agradecerle su paciencia y cooperación, le enviaremos hoy mismo a nuestro técnico para llevarle su nuevo equipo. Esto no tendrá costo alguno. Además, le haremos un descuento del 20% en su próxima mensualidad por las molestias causadas, ¿le parece bien?"

Fuentes Textuales: Opiniones



Fuentes Textuales: Patentes



[← Back to results](#) [self-driving;](#)

Managing self-driving behavior of autonomous or semi-autonomous vehicle based upon actual driving behavior of driver

Abstract

A system and method for measuring a driver's actual driving behaviors (e.g., acceleration, deceleration) in a manual driving mode to determine their preferred driving style, and then causing an autonomous or semi-autonomous vehicle to operate itself, within limits, in accordance with the drivers' driving style when operating in a self-driving mode, thereby providing a more familiar and comfortable driving experience for the driver. Data is collected on the actual driving behavior, any pre-existing data is accessed on the actual driving behavior, and the collected data and the pre-existing data are combined. A custom control is then created based upon the combined data, and the custom control is applied to manage the self-driving behavior of the autonomous or semi-autonomous vehicle in a self-driving mode. Additional data continues to be collected on the actual driving behavior, and the custom control is adjusted based upon the collected additional data.

Images (6)

US10689003B1
United States

[Download PDF](#) [Find Prior Art](#) [Similar](#)

Inventor: [Brian M. Fields](#), [Joseph P. Harr](#), [Jeremy Myers](#)
Current Assignee: [State Farm Mutual Automobile Insurance Co](#)

Worldwide applications
2015 · [US](#) 2018 · [US](#) 2019 · [US](#)

Fuentes Textuales: Leyes

The screenshot displays the official website of the Chilean Congress (Biblioteca del Congreso Nacional de Chile / BCN). The header is red with the BCN logo and navigation links: Ley Chile, BCN, Ley Chile, Catálogo, Información Territorial, Historia Política, and Formación Cívica. Below the header, there is a search bar with the text "Buscar en Ley Chile" and a magnifying glass icon, followed by a "BÚSQUEDA AVANZADA" link and a "Selección" button. The main content area has a red navigation bar with tabs: Texto, Versiones, Historia de la Ley, and Autores. The "Texto" tab is selected. On the left, a sidebar lists the structure of the law: LEY 17260, ENCABEZADO, ARTÍCULO 1, ARTÍCULO 2, ARTÍCULO 3, ARTÍCULO 4, and PROMULGACIÓN. The main text area shows the following content:

MODIFICA LA LEY N.º 16.724 Y por cuanto el H. Congreso Nacional ha dado su aprobación al siguiente

Proyecto de ley:

"Artículo 1.º- Agrégase el siguiente inciso al artículo 24 de la ley N.º 16.724:

"Las disposiciones de los incisos primero, segundo y quinto del artículo anterior son aplicables a los trabajadores enunciados en la letra a) del artículo 19 del decreto supremo N.º 153 (M) de 22 de Febrero de 1966, como asimismo a las funciones de los Embaladores, en los puertos donde existen de hecho y en aquellos

Text Mining

Tarea de examinar automáticamente textos en lenguaje natural, con el fin de descubrir patrones y relaciones que puedan ser útiles y accionables para la toma de decisiones (***text mining***).



Data Science: Text Mining



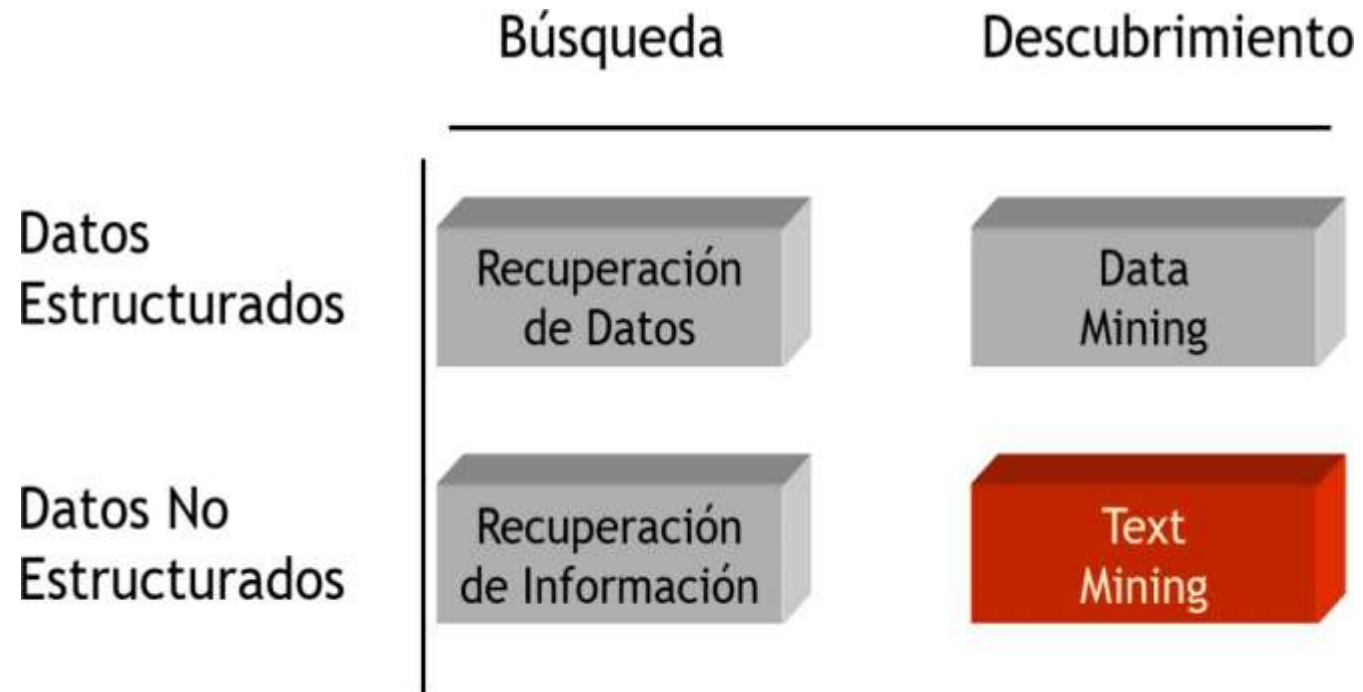
Text Mining

- Las aplicaciones en ámbitos industriales y de negocios son muchas:
 - ✓ Agrupación de documentos (*clustering*)
 - ✓ Categorización de textos
 - ✓ Extracción de información para poblar bases de datos
 - ✓ Generación de textos
 - ✓ Descubrimiento de asociaciones entre otros.

Text Mining

- Los métodos computacionales de análisis de texto deben ser capaces de abordar estos tres problemas:
 - Ambigüedad
 - Dimensionalidad
 - Conocimiento lingüístico

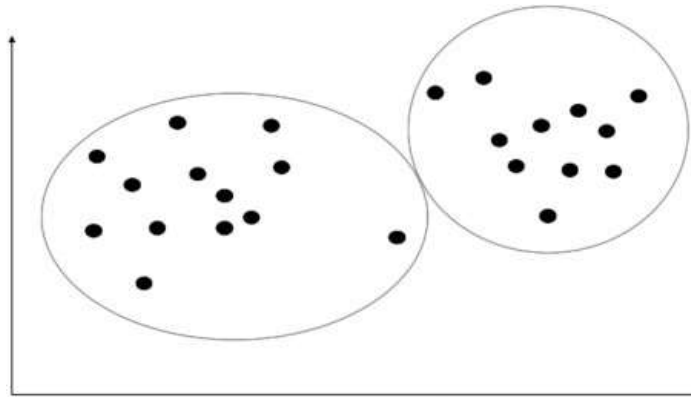
Búsqueda versus descubrimiento



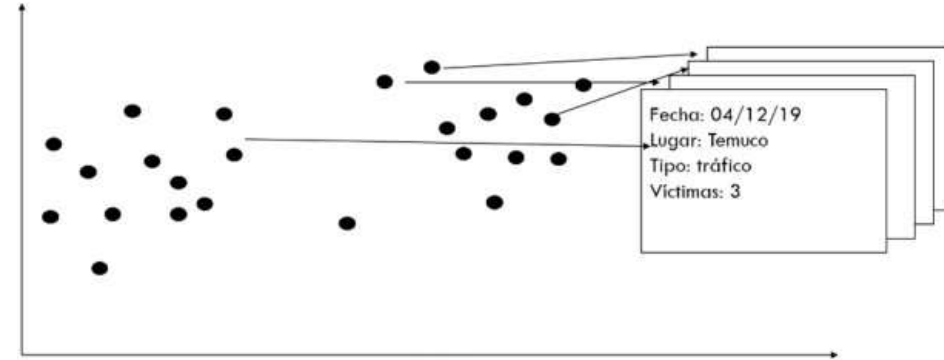
Procesamiento del Lenguaje Natural (Natural-Language Processing NLP)

- Proporciona teorías, modelos y métodos.
- Se enfocan en crear sistemas que procesan información textual.
- Usualmente, muchas de estas requieren recuperar información desde fuentes específicas de datos no estructurados.
- Muchas tareas de NLP se pueden resolver utilizando enfoques tradicionales basados en reglas y métodos probabilísticos.

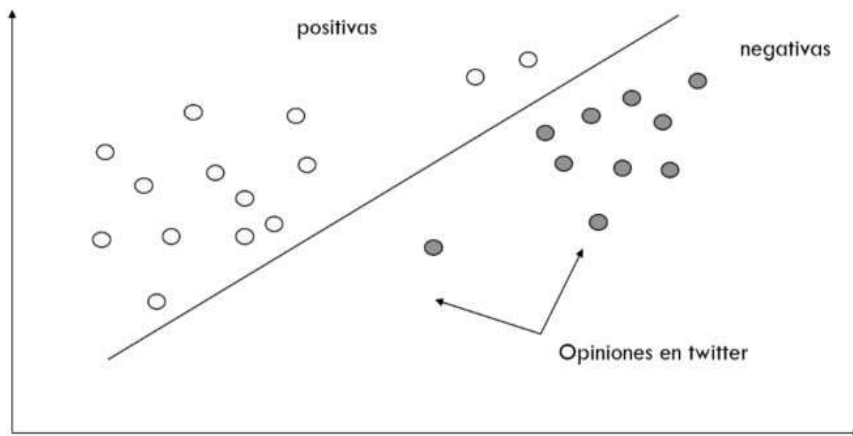
Tareas y Aplicaciones



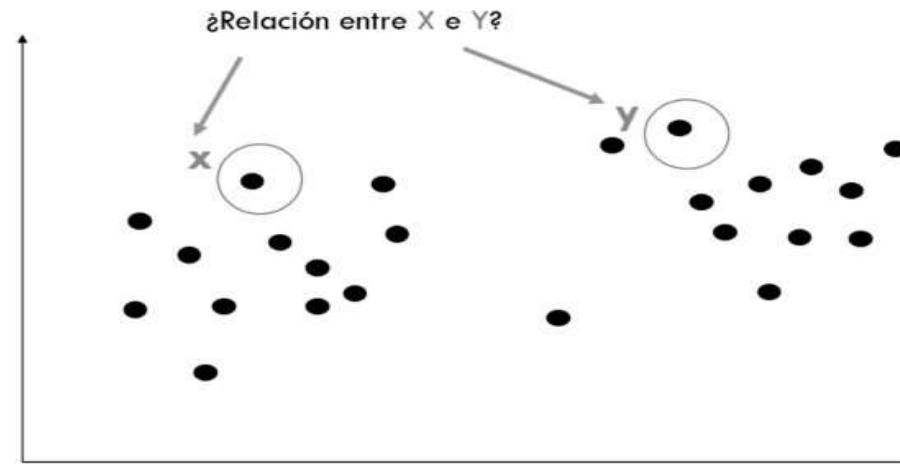
Agrupación de información textual



Extracción de información

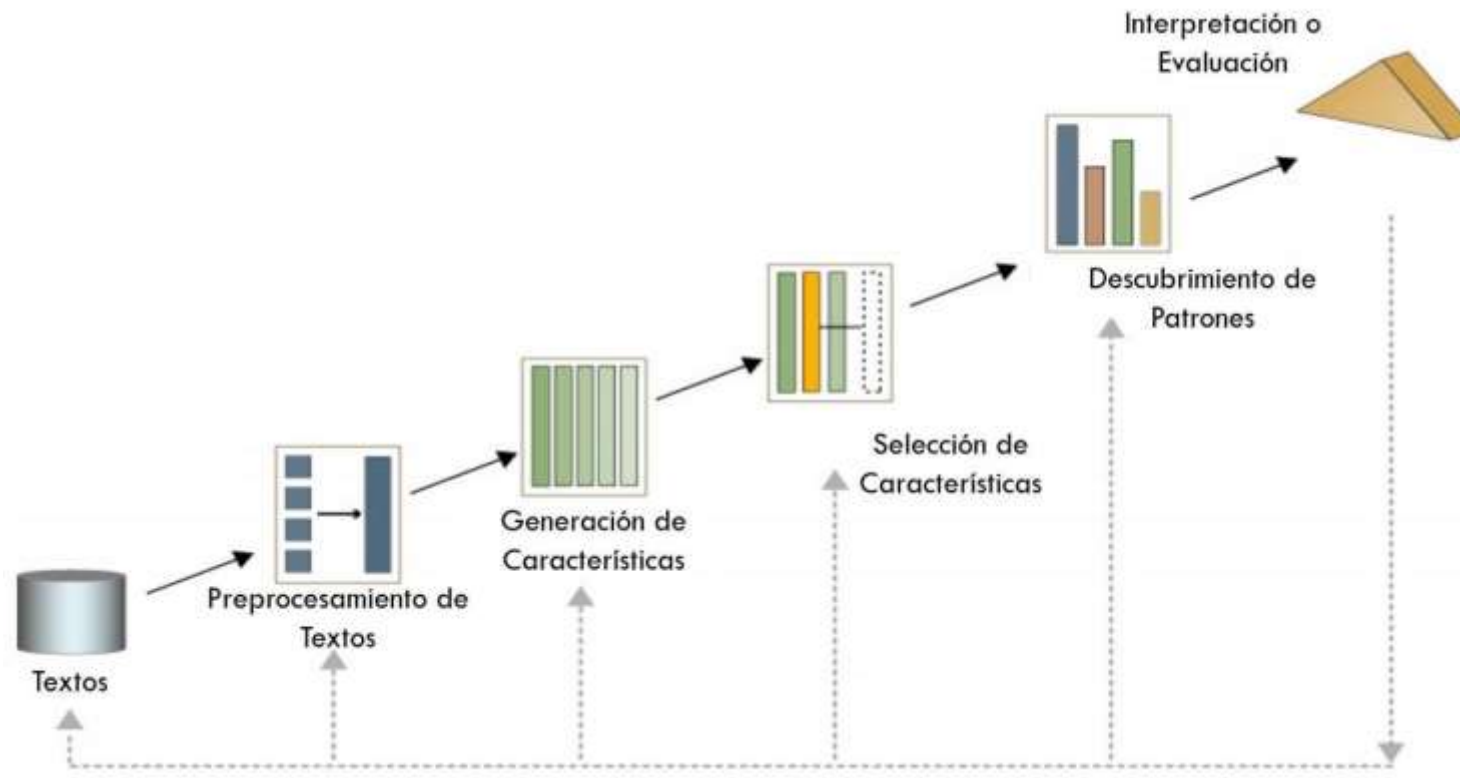


Categorización de textos



Inferencia de relaciones

Proceso de la Analítica Textual



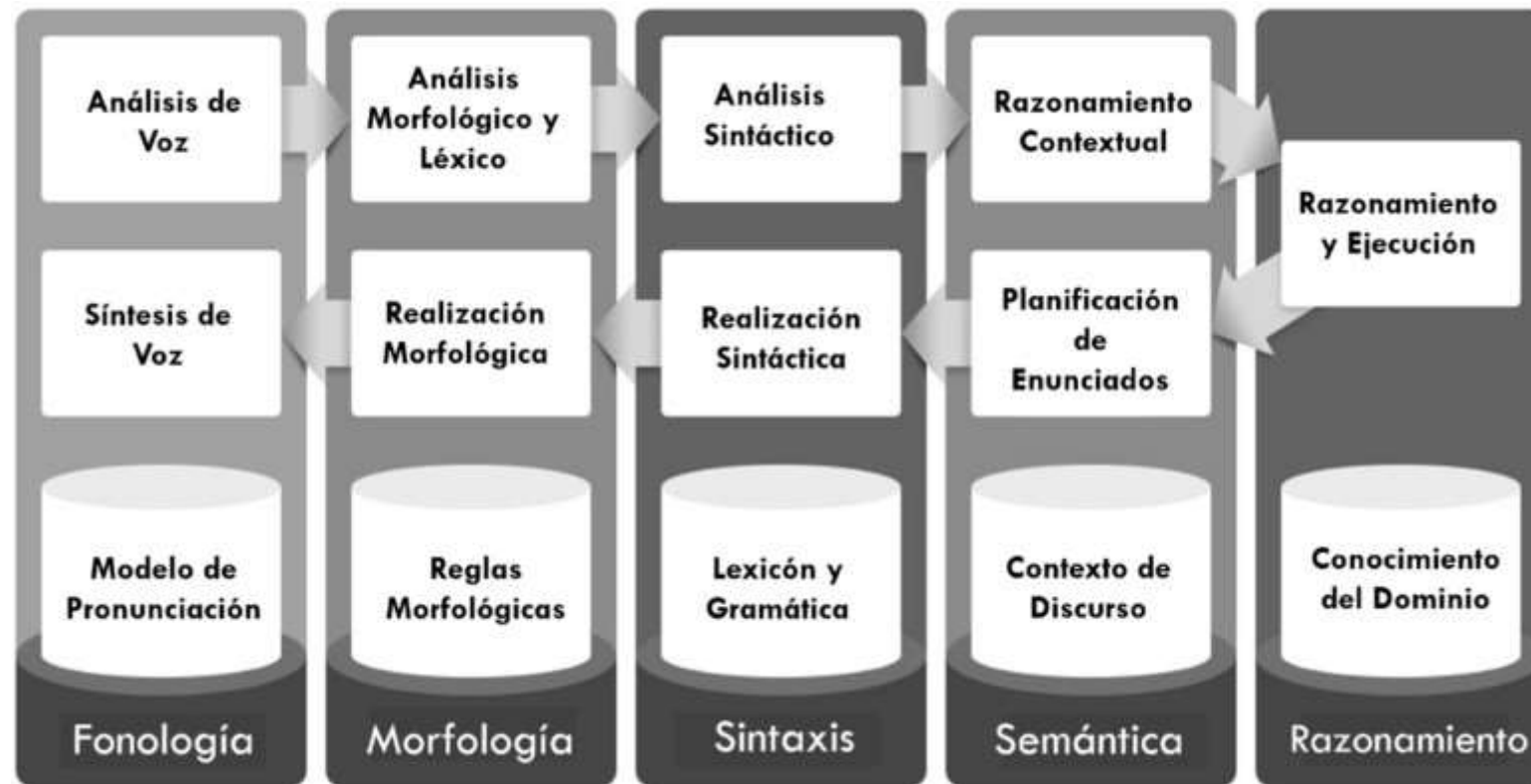
Text Mining

Ciencia basada en examinar y descubrir patrones interesantes e idealmente accionables, desde grandes colecciones de documentos escritos en lenguaje natural.

Para esto se combinan técnicas y modelos provenientes desde el NLP, el ML y la recuperación de información.

Lo anterior permite realizar tareas como: categorización de documentos, agrupación de información textual, extracción de información específica desde documentos, descubrimiento de asociaciones, detección de tópicos.

Estas tareas constituyen la base de aplicaciones transversales a todos los dominios, tanto en el ámbito privado como público, desde aplicaciones científicas hasta las industriales y de negocios.



Niveles, tareas y recursos lingüísticos en el NLP

VISUALIZACIONES DE COLECCIONES DE TEXTO

Nube de palabras

La forma más tradicional de **presentar** en **forma resumida y visual** un **documento** o una **colección de texto**.



Word Tree

**Martin Wattenberg -
Fernanda Viegas**
Creadores de técnica
WordTree

Permite:

- Tomar un texto
- Generar un árbol de patrones.



Texto y reducción de dimensionalidad

Reducción de dimensionalidad



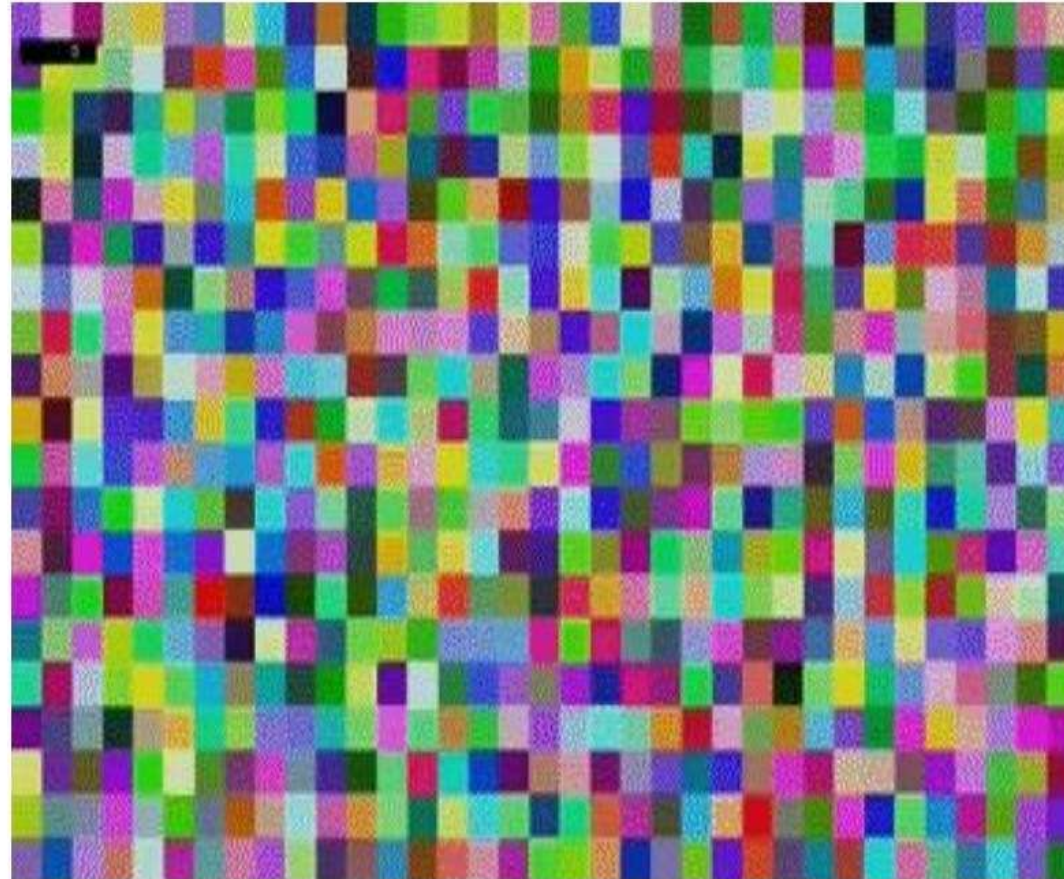
Representar noticias (textos) de elección presidencial (2013) como puntos en 2 dimensiones.

El canal color representa distintos medios.

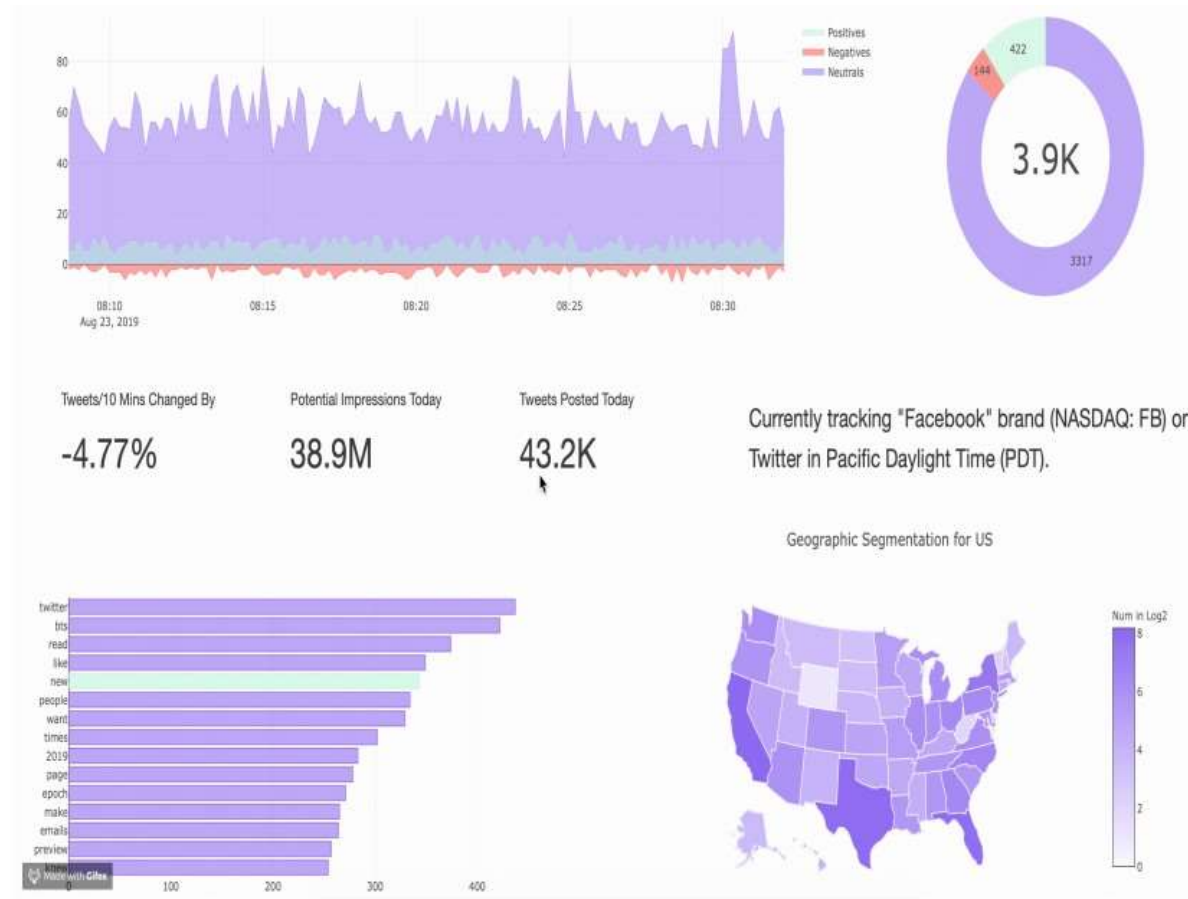


Aplicaciones del Text Mining

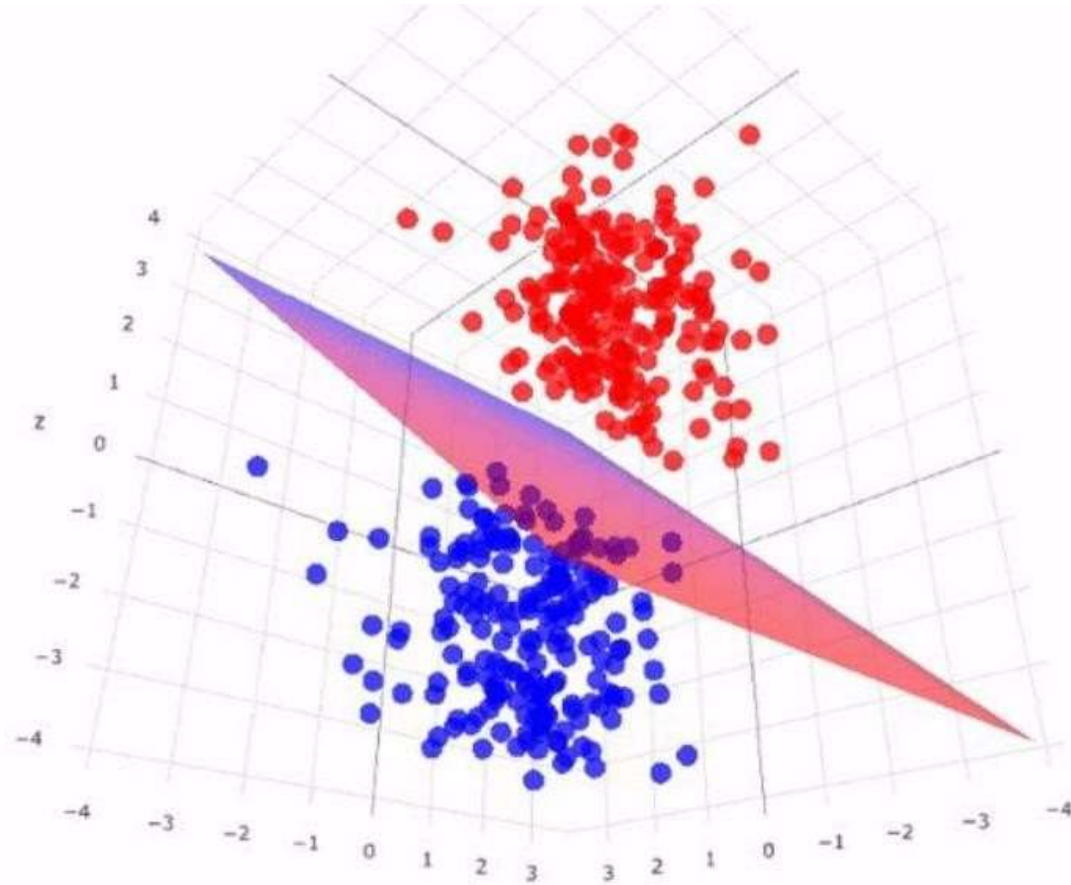
Clustering de quejas



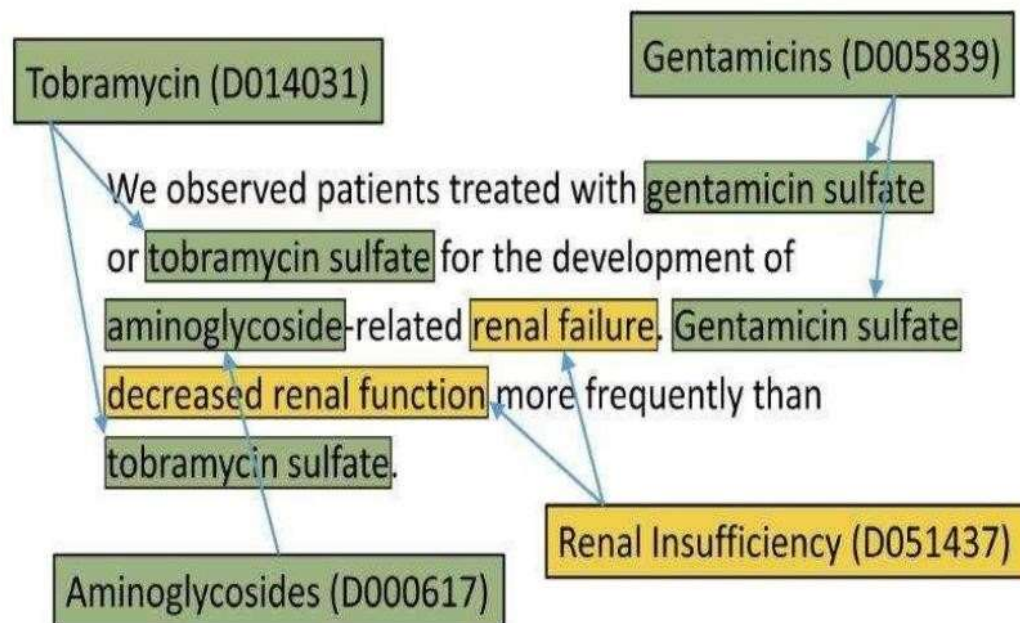
Análisis de Sentimientos



Clasificación de Clientes



Resolución de
Patologías
Médicas



F-Score	Dataset	Task
90.30%	CLEF	Disease & problem norm.
92.29%	NCBI	Disease norm. in literature

DIFICULTADES O PITFALLS DEL TEXT MINING



Radio ADN @adnradiochile · 16h

...

AUDIO | Director del Sernac sobre demanda colectiva contra Falabella y Paris: "Estas empresas deberán responder ante la justicia" bit.ly/3362rZ8



30

12

57



Falabella Ayuda CL ✓ @Falabell... · 45s

...

Hola @adnradiochile Por favor envíanos por interno el N° de tu orden de compra, RUT, mail y teléfono de contacto para poder revisar tu solicitud. ¡Quedamos atentos!





Sole&Smile @Solefija · 41m

Soy Celiaca y estoy embarazada.

@LATAM Airlines no me puede dar mi vianda porque como tiene los aviones en reparacion subcontrato una empresa y no previo casos asi. Vuelo 20 horas.



LATAM Airlines ✓ @LATAM Airlines · 26m

Hola Celiaca, con el fin de brindarte mayor asistencia, brindanos los nombres de los pasajeros, código de reserva o número de ticket, vía DM. Estaremos atentos a tus comentarios. Saludos. [twitter.com/messages/compo...](https://twitter.com/messages/compose)

