

# Tema 5. Aplicaciones

# Aplicaciones

---

- Hay numerosas aplicaciones de las técnicas y tecnologías presentadas en los capítulos anteriores
- Nos centraremos en mostrar dos de ellas:
  - Análisis de sentimientos
  - Análisis de tendencias

- Minería de opinión y análisis de sentimientos
- Análisis de tendencias
- Bibliografía

# Minería de opinión: introducción

---

- Los términos de “minería de opinión” y “análisis de sentimientos” se usan de forma intercambiable
- Su objetivo es extraer información subjetiva (opinión, valoración, emoción, actitud) sobre un evento, producto, entidad y sus atributos
- Se aborda como un problema de clasificación de un texto en una polaridad positiva, negativa o neutra
- La polaridad a veces se puede expresar en una escala de 1-5 (estrellas en revisiones de películas, productos, ...)
- Es una tarea complicada incluso para los humanos: interpretaciones que dependen de factores culturales, experiencias propias de cada persona
- Cuanto más corto sea el texto y menos formal más difícil resulta

# Minería de opinión: introducción

- La minería de opinión se utiliza fundamentalmente para la toma de decisiones.  
Por ejemplo:
  - Elegir un servicio (restaurante, teléfono móvil, ...)
  - Determinar la política de mercadotecnia de una empresa
  - Predecir las tendencias en periodo electoral
  - Identificar nuevos eventos que resultan relevantes para los usuarios
  - Interacciones sociales y sus cambios (estudios de satisfacción por áreas geográficas, edad, ...)

# Minería de opinión: retos y dificultades

- Determinar la entidad de la que se está opinando (ambigüedad) y sus aspectos o atributos (en un hotel la ubicación, limpieza, servicio, ruido, habitaciones, ...)
- Una misma palabra se puede considerar positiva o negativa dependiendo del aspecto que se evalúa:
  - “largo” puede ser una valoración positiva para la vida de una batería de un portátil, pero negativa si se refiere al tiempo que tarda en arrancar el mismo portátil
- En una misma valoración pueden aparecer comentarios positivos y negativos. Ejemplo:

*“aunque el servicio no es bueno, me gusta este restaurante”*

  - La oración es positiva con respecto al restaurante pero negativa con respecto al servicio

# Minería de opinión: enfoques

Hay principalmente dos enfoques [1]:

## **1. Semántico:**

- Basado en el uso de diccionarios de términos con valoración semántica de polaridad u opinión
- Los sistemas preprocesan el texto, lo dividen en palabras, eliminan las palabras vacías de contenido (stop words), normalizan a nivel léxico y comprueban la aparición de los términos en el diccionario para asignar el valor de polaridad del texto mediante la suma de los valores de polaridad de los términos

# Minería de opinión: enfoques

## 1. Semántico:

- Estos sistemas incluyen un tratamiento más o menos avanzado de:
  - términos modificadores (*muy, poco, demasiado*) que aumentan o reducen la polaridad de los términos a los que acompañan
  - términos inversores o negadores (*no, tampoco*), que invierten la polaridad de los términos a los que afectan
- Diccionarios de sentimientos disponibles (para español [Sentiwordnet](http://sentiwordnet.isti.cnr.it/) (<http://sentiwordnet.isti.cnr.it/>), traducción a español de otro en inglés (<http://danigayo.info/PFCblog/index.php?entry=entry130117-183114>))



# Minería de opinión: enfoques

## **1. Semántico. Pros y contras:**

- Su ventaja principal es que algunos errores son relativamente sencillos de corregir enriqueciendo el diccionario
- Su inconveniente principal es el esfuerzo para construir un diccionario para cada dominio ya que se basa en un trabajo manual
- En [2] se describe un análisis de este enfoque aplicado a tuits en español

# Minería de opinión: enfoques

Hay principalmente dos enfoques [1]:

## 2. Aprendizaje supervisado:

- Se entrena un clasificador con un algoritmo de aprendizaje supervisado a partir de una colección de textos anotados con polaridad
- Cada texto habitualmente se representa con un vector de palabras (bag of words), n-gramas o skip-grams, con características semánticas que modelan la estructura sintáctica de las palabras o frases, la intensificación, la negación, la subjetividad o la ironía

# Minería de opinión: enfoques

## **2. Aprendizaje supervisado. Pros y contras:**

- La principal ventaja es que cuesta poco construir un analizador de sentimientos a partir de la colección de textos etiquetados o anotados
- Como desventaja, estos sistemas suelen ser una caja negra en la que corregir errores o añadir nuevo conocimiento suele requerir ampliar la colección de textos etiquetados y reentrenar el modelo
- La anotación de textos también requiere un gran esfuerzo humano, aunque los sistemas semisupervisados lo pueden paliar

# Minería de opinión: enfoques

## 2. Aprendizaje supervisado: *transformers*:

- La arquitectura de Transformers y los embeddings contextuales han mejorado los resultados de los sistemas supervisados previos de análisis de sentimientos
- Esta arquitectura ha permitido mejorar significativamente los resultados en análisis de polaridad de aspectos concretos de una entidad [5], ya que algunos aspectos pueden tener una polaridad diferente de la polaridad global que se asigna a la entidad
- Video con descripción de cómo hacer análisis de sentimientos con Bert:  
<https://www.youtube.com/watch?v=8N-nM3QW7O0>

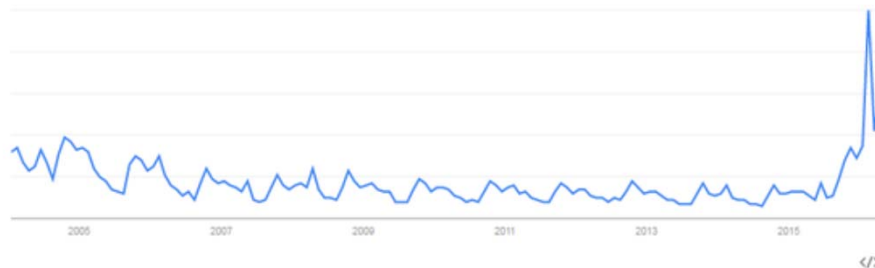
# Minería de opinión: demos

- En <https://text-processing.com/demo/sentiment/> hay una demo desarrollada en Python y una API que permite analizar la polaridad de textos
- En <https://text2data.com/Demo> hay otra demo que permite indicar la polaridad de textos cortos, tipo tuit
- Lectura para profundizar: textos formales [1], social media [3] y [4]
- La campaña de evaluación TASS 2020 (<http://tass.sepln.org/2020/>) es la más reciente para análisis de sentimientos de tuits en español. En [http://ceur-ws.org/Vol-2664/tass\\_overview.pdf](http://ceur-ws.org/Vol-2664/tass_overview.pdf) se puede encontrar el resumen de la campaña y los resultados

- Minería de opinión y análisis de sentimientos
- **Análisis de tendencias**
- Bibliografía

# Análisis de tendencias: introducción

- Las tendencias son **temas candentes que provocan un interés extraordinario**, generalmente de forma temporal, en la comunidad de usuarios de una red social.
- Es importante diferenciar un tema candente de un tema que es popular constantemente. Los primeros son tendencias, mientras que los segundos no.
- Una tendencia surge de forma repentina, como por ejemplo por una noticia que nace, dura un tiempo, y después se deja de hablar de ella:



# Análisis de tendencias: introducción

---

- La tarea de **detección de tendencias** consiste en descubrir temas que muchos usuarios comparten al mismo tiempo.
- Fuera de las redes sociales, también se puede asimilar a la tarea de **detección de eventos** (*event detection*).
- En lugar de detectar simplemente términos frecuentes, se trata de encontrar términos cuya **frecuencia crece de forma repentina** y, por tanto, no eran tan frecuentes en momentos anteriores.



# Análisis de tendencias: técnicas

- La mayoría de aproximaciones para la detección de tendencias se centran en técnicas de **clustering**. Por ejemplo:
  - La más utilizada es **LDA** (Latent Dirichlet Allocation), una aproximación para topic modelling.
  - También se han utilizado otras técnicas como **K-means** y **LSH** (Locality Sensitive Hashing).
- Las alternativas al clustering se centran en el análisis de frecuencias, es decir, descubrir términos **cuya frecuencia aumenta de forma repentina**.
  - Requiere mantener un historial de frecuencias en memoria. Si un término ya era frecuente en periodos anteriores, no será tendencia en el periodo actual.
  - Se puede usar Kullback-Leibler divergence (**KLD**) que compara la frecuencia del término en el momento actual,  $P(i)$ , y la frecuencia en momentos anteriores,  $Q(i)$ . Un alto valor de  $D$  indica que el término es tendencia.

$$D_{\text{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}.$$

# Análisis de tendencias: técnicas

---

Diferentes características a considerar:

- **Temporalidad:** Puede realizarse en directo sobre un stream de datos u offline sobre una colección pregenerada.
- **Aprendizaje:** Puede ser supervisado (utilizando un conjunto de tendencias para entrenamiento) o no supervisado.
- **Tipos de tendencias:** Detección general de todo tipo de tendencias o restringido a un cierto tipo de tendencias (p. ej. noticias).

# Análisis de tendencias: aplicaciones

- **Emergencias:** la detección de tendencias puede ayudar a conocer la situación (*situational awareness*).
- **Noticias:** un sistema de detección de tendencias puede ayudar a descubrir noticias de impacto de forma temprana.
- **Gestión de reputación:** se pueden detectar tendencias en sentimiento positivo o negativo hacia una marca o empresa.

# Análisis de tendencias

- Lectura recomendada: [5]
- Herramientas:
  - Twitter topic detection en Python: <https://github.com/heerme/twitter-topics>
  - Event detection and clustering for Twitter: <https://github.com/harshil93/Event-Detection-and-Clustering-for-Twitter>

- Minería de opinión y análisis de sentimientos
- Análisis de tendencias
- Enlaces de interés
- **Bibliografía**

# Bibliografía

---

- [1] B. Liu. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012. Versión previa online: <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>.
- [2] A. Moreno-Ortiz, C. Pérez Hernández. [Lexicon-Based Sentiment Analysis of Twitter Messages in Spanish](#). *Procesamiento del Lenguaje Natural*, vol 50, pág. 93-100, 2013.
- [3] P. Rey del Castillo. [Fuzzy sentiment analysis using spanish tweets](#). *European Conference on Quality in Official Statistics*, 2016.
- [4] [Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with 32nd SEPLN Conference \(SEPLN 2016\)](#)
- [5] Chi Sun, Luyao Huang, Xipeng Qiu. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. 2019. <https://arxiv.org/pdf/1903.09588.pdf>
- [6] Atefeh, Farzindar, and Wael Khreich. "A survey of techniques for event detection in twitter." *Computational Intelligence* 31.1 (2015): 132-164.