

6

SENTIMENT ANALYSIS

Minería de Opinión



Contenidos



- 1. Las emociones y la Web**
- 2. Definición y descripción**
- 3. Tipos de análisis**
- 4. Opiniones y emociones**
- 5. Métodos y técnicas**
- 6. Desarrollos**
- 7. Desafíos**
- 8. Modelos avanzados**
- 9. Evaluación de resultados**

Objetivos



- A. COMPRENDER** los **fundamentos básicos** de la técnica de Análisis de sentimiento.
- B. CONOCER** las principales **ventajas e inconvenientes** de esta técnica.
- C. APLICAR** esta técnica a un conjunto de contenidos para **evaluar de forma crítica los resultados** obtenidos.

LAS EMOCIONES Y LA WEB



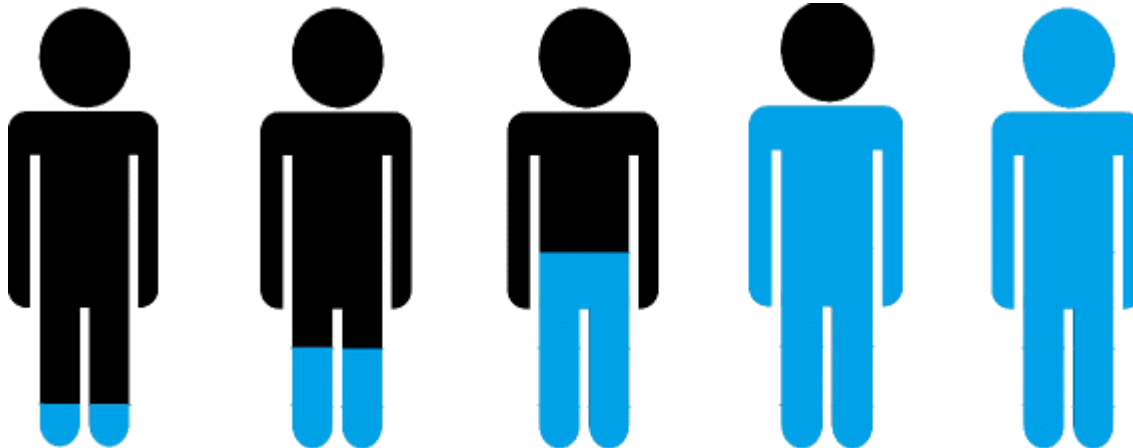


Los estímulos del usuario

Básicos
VS
Auxiliares

Internos
VS
Externos

Naturales
VS
Artificiales





Las emociones y sentimientos

- ☐ Las emociones son **el motor para nuestra inmediata reacción** ante **estímulos**.
- ☐ El **sentimiento** es la **conciencia** de una cierta **emoción**.
- ☐ Las **emociones y sentimientos** son el origen de la **conducta humana**.



Las emociones y sentimientos

Emociones Positivas

Físicas

- Placer

Mentales

- Recompensa
- Alegría
- Satisfacción

Emociones Negativas

Físicas

- Dolor

Mentales

- Castigo
- Enfado
- Insatisfacción

La motivación

- ☐ El cerebro (en tanto que procesador de información), busca desde $t=0$ la **supervivencia** del sistema.
- ☐ No hay nada más esencial para la supervivencia del ser humano que el éxito en **obtener todo aquello que es beneficioso y evitar lo dañino**.
- ☐ Por tanto, nuestra motivación nos guía continuamente hacia la obtención de **estímulos positivos** y el **rechazo a los negativos**.

¿Por qué es importante **medir el sentimiento**?



☐ **Empresas y organizaciones**

- ✓ Marketing Intelligence



☐ **Individuos**

- ✓ Decisiones en compras de productos o servicios
- ✓ Decisiones en asuntos sociales (ej., elecciones)

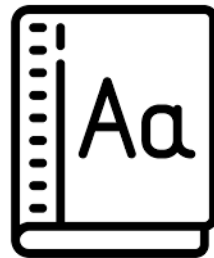
❓ ¿Por qué es importante **medir el sentimiento**?

- ☐ El auge de las plataformas de redes sociales ha despertado el interés en el **análisis del sentimiento**.
- ☐ Con la proliferación de revisiones, calificaciones, recomendaciones y otras formas de expresión online, la opinión de los usuarios web se ha convertido en un tipo de moneda virtual para las empresas que buscan **comercializar sus productos, identificar nuevas oportunidades y administrar su reputación**.

¿? ¿Por qué es importante **medir el sentimiento**?

- ☐ Existe un **efecto del discurso público** en la reputación de las empresas, productos, marcas e incluso personas.
- ☐ Filtrar el ruido, **comprender las conversaciones**, identificar el contenido relevante y finalmente actuar de forma apropiada.

DEFINICIÓN Y DESCRIPCIÓN



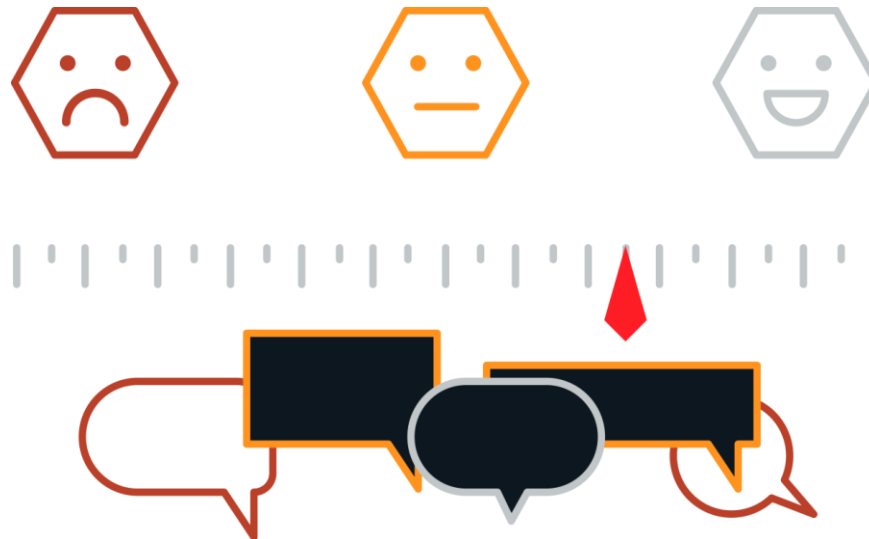
What is sentiment analysis?



El análisis de sentimiento trata de **identificar, extraer, cuantificar y estudiar los estados afectivos** e información subjetiva de una forma sistemática.



Se basa en la **identificación** de palabras individuales o frases en un texto y en su posterior **vinculación** con diferentes escalas emocionales.





El análisis de sentimiento puede igualmente ser aplicado a **contenido visual** (imágenes y vídeos)





Y, por supuesto: emoticonos y *emojis*



<http://www.emojitracker.com/>



Y, por supuesto: emoticonos y *emojis*



Search Emojipedia



eg heart 🍷 arrow beach 👍 fireworks birthday 🌱



<https://emojipedia.org/>

Áreas implicadas

- Implica el uso del procesamiento natural del idioma (NPL), análisis de texto, lingüística computacional y biométrica, análisis métrico de datos...



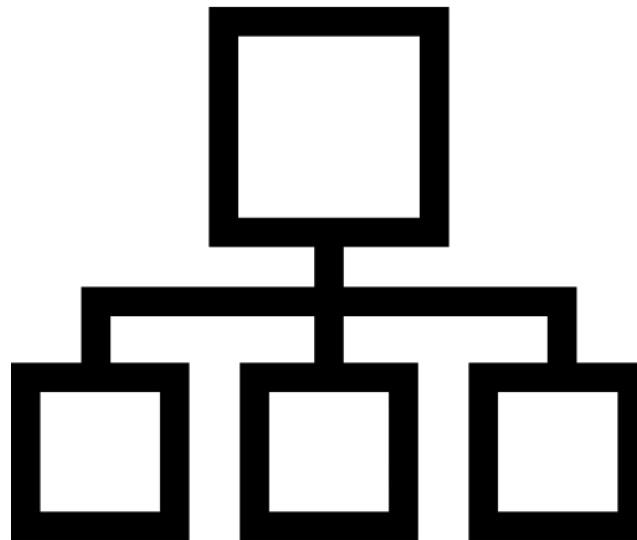
Cyberemotions Project

- “To understand the role of collective emotions in creating, forming and breaking-up ICT mediated communities as a spontaneous emergent behaviour occurring in complex techno-social networks”



<http://www.cyberemotions.eu>

TIPOS DE ANÁLISIS



Básico

- Se basa en clasificar la polaridad de una opinión expresada en un texto como **positiva**, **negativa** o **neutra**.



Negative



Neutral



Positive



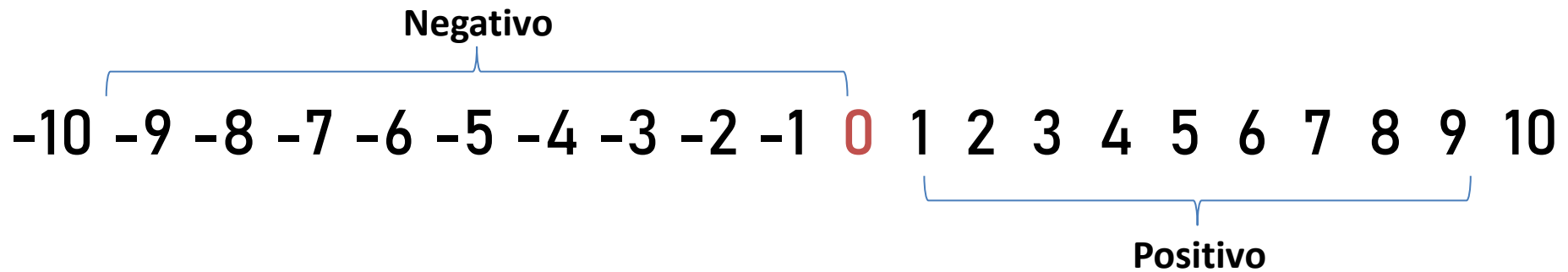
Básico

- ☐ Se recomienda el uso de un **sistema de escalamiento**.
- ☐ Las **palabras comúnmente asociadas con tener un sentimiento negativo, neutral o positivo** reciben un número asociado en una escala o intervalo.
- ☐ Esto hace posible ajustar el sentimiento de un término determinado **en relación con su entorno** (generalmente a nivel de frase).



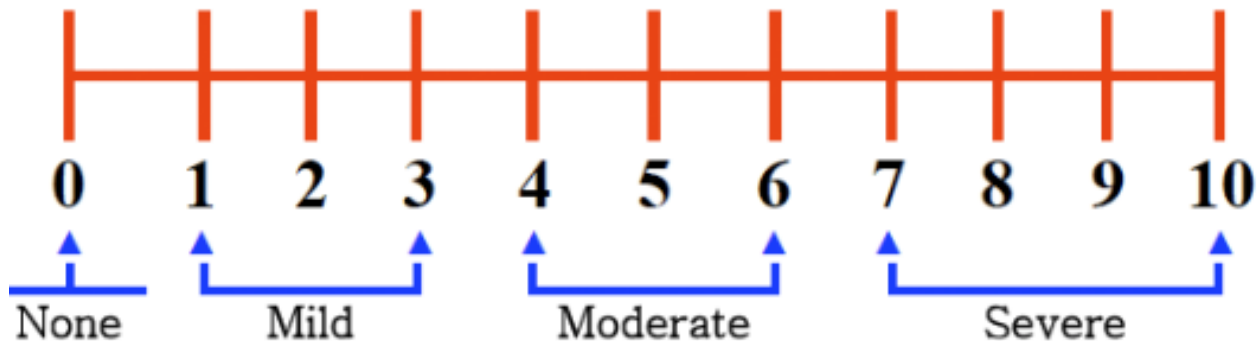
Básico

□ Escala **simétrica** (por ejemplo, de -10 a $+10$):



Básico

- Escala **asimétrica** (de 0 a un límite superior positivo; por ejemplo +10):





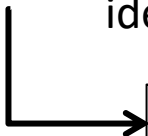
Básico



¿Y cómo operamos con el nivel neutro?

A

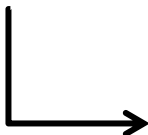
El algoritmo procede primero a identificar y filtrar el lenguaje neutro, y luego evalúa el resto en términos respecto al neutro identificado.



Cuando los datos están claramente agrupados

B

El algoritmo construye una clasificación a tres niveles en un solo paso directamente.



Cuando los datos son en su mayoría neutrales con pequeñas desviaciones hacia afectos positivos y negativos

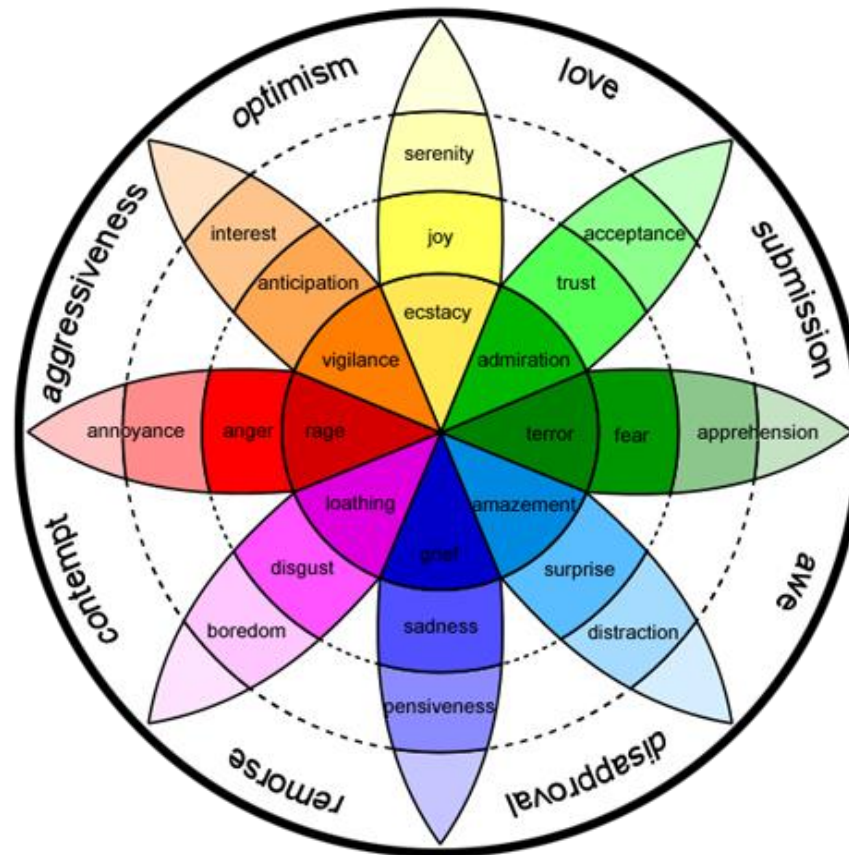


Avanzado

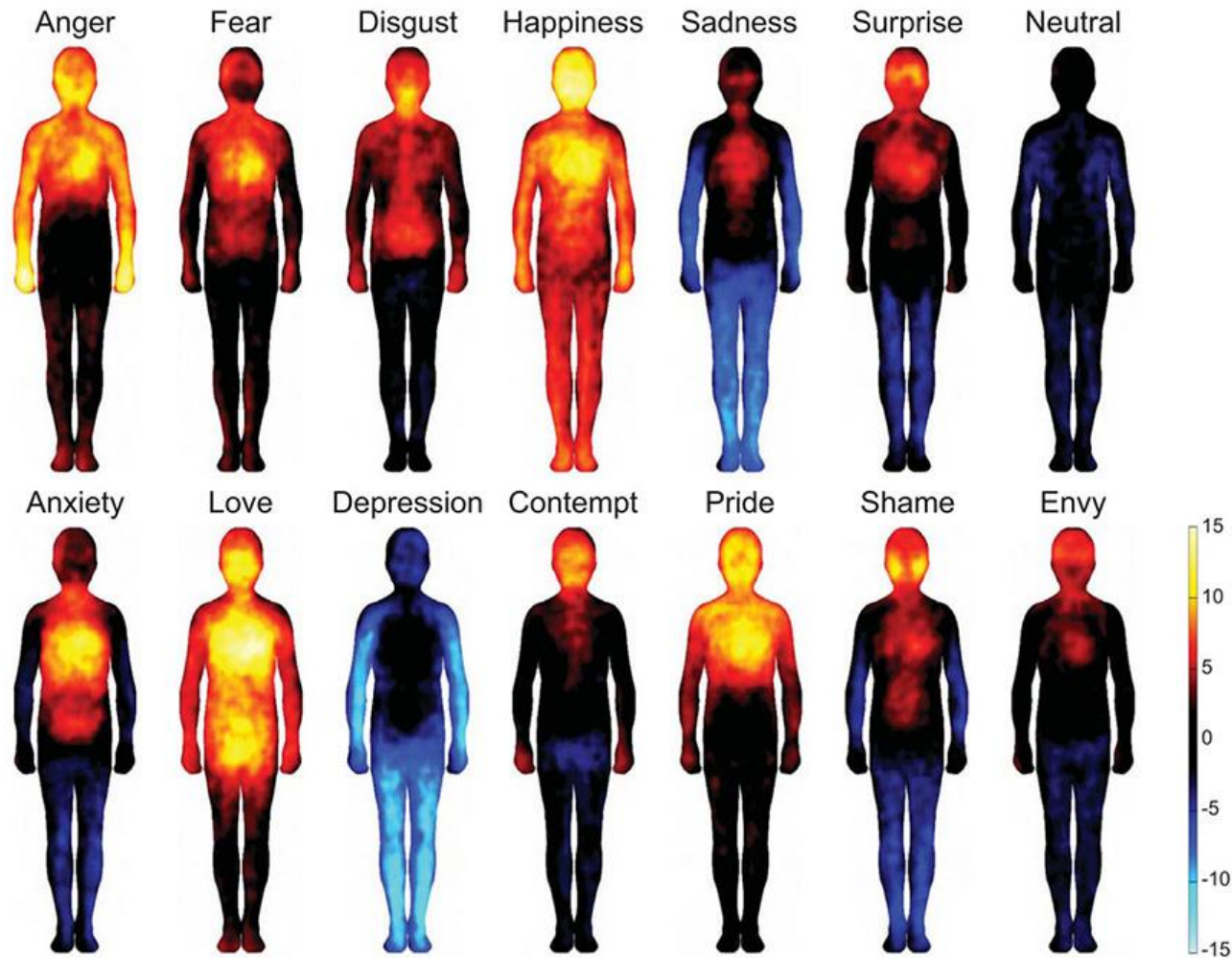
- ☐ Busca más allá de la polaridad para adentrarse en los **distintos estados emocionales** (enfadado, triste, feliz, etc.) tanto de la persona que genera el contenido (*source*) como del referente de sus contenidos (*target*).

Avanzado

LA RUEDA DE LA EMOCIÓN (PLUTCHIK):



Avanzado



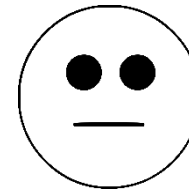


Avanzado



Emociones básicas (Parrot)

1. LOVE
2. JOY
3. SURPRISE
4. ANGER
5. SADNESS
6. FEAR



Emociones faciales básicas (Ekman)

1. HAPPINESS
2. DISGUST
3. SURPRISE
4. ANGER
5. SADNESS
6. FEAR



OPINIONES Y EMOCIONES



Opiniones regulares

- ☐ Opiniones sobre un único target
- ☐ Pueden ser **directas** o **indirectas**



“The touch screen is really cool.”



“After taking the drug, my pain has gone.”



Opiniones comparativas

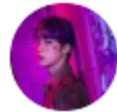
- ☐ Opiniones sobre más de un target



“iPhone is better than Galaxy.”

Emociones

- ☐ Emociones sobre uno o más targets (explícitos o implícitos)



Cari Stars⁷
@StarsCari

Why is Elon musk so fucking stupid.

5:32 PM · Mar 7, 2020 · [Twitter for iPhone](#)

Dimensiones

☐ El origen, el portador y el destino:



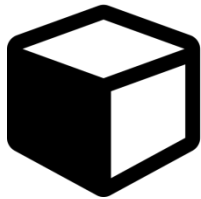
Dimensiones

- Atributos de la dimensión “Destino” (featured-based):



Dimensiones

☐ La **opinión** como un quintuplo:



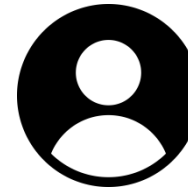
Entity



Aspect



Sentiment



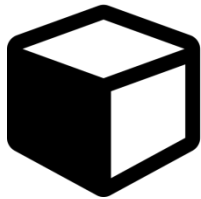
Holder



Time

Dimensiones

☐ La **emoción** como un quintuplo:



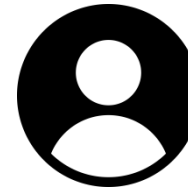
Entity



Aspect



Emotion
Type



Feeler



Time

MÉTODOS Y TÉCNICAS

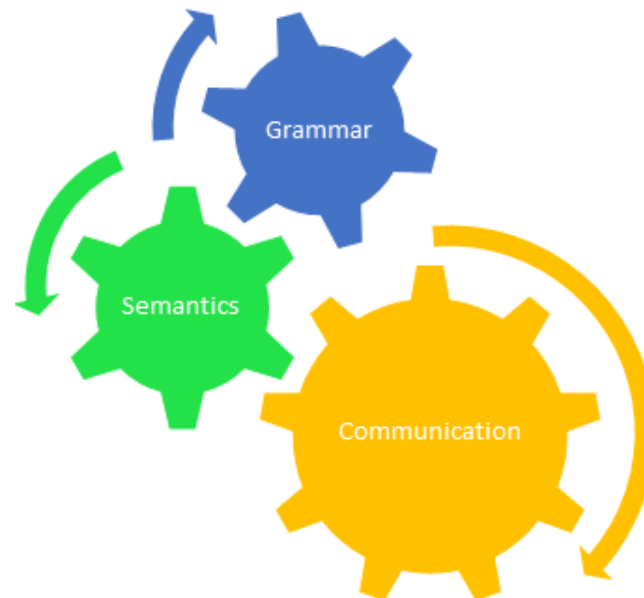


Técnicas basadas en el conocimiento

- ☐ Clasifican el texto por categorías de afecto según la **presencia explícita de palabras de afecto no ambiguas** (feliz, triste, asustado y aburrido).
- ☐ Adicionalmente, asignan a palabras arbitrarias una **"afinidad" probable** a emociones particulares.

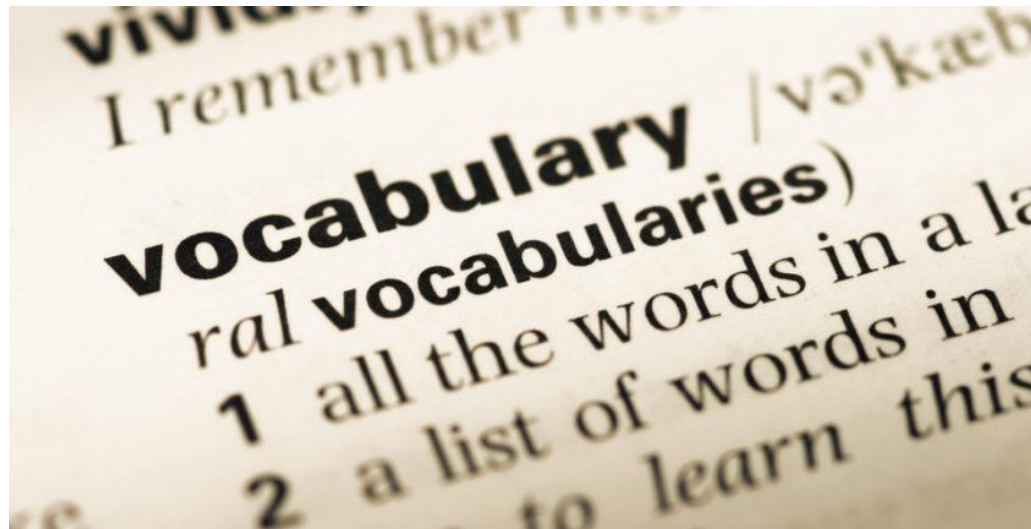
Técnicas basadas en el conocimiento

- Los sistemas basados en el conocimiento hacen uso de los recursos públicos disponibles (*lexicons*) para **extraer la información semántica y afectiva** asociada con los conceptos del lenguaje natural.



Técnicas basadas en el conocimiento

- ☐ Un *lexicon* es un diccionario, vocabulario o un **libro de palabras** que se ha creado para analizar los sentimientos.
- ☐ La mayoría tienen una lista de palabras polares positivas y negativas con algún puntaje asociado.



Técnicas basadas en el conocimiento

- ☐ **AFINN**

<https://github.com/fnielsen/afinn>

- ☐ **Bing Liu's lexicon**

<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

- ☐ **MPQA**

http://mpqa.cs.pitt.edu/lexicons/subj_lexicon

- ☐ **SentiWordNet**

<http://sentiwordnet.isti.cnr.it>

- ☐ **VADER**

<https://github.com/cjhutto/vaderSentiment>

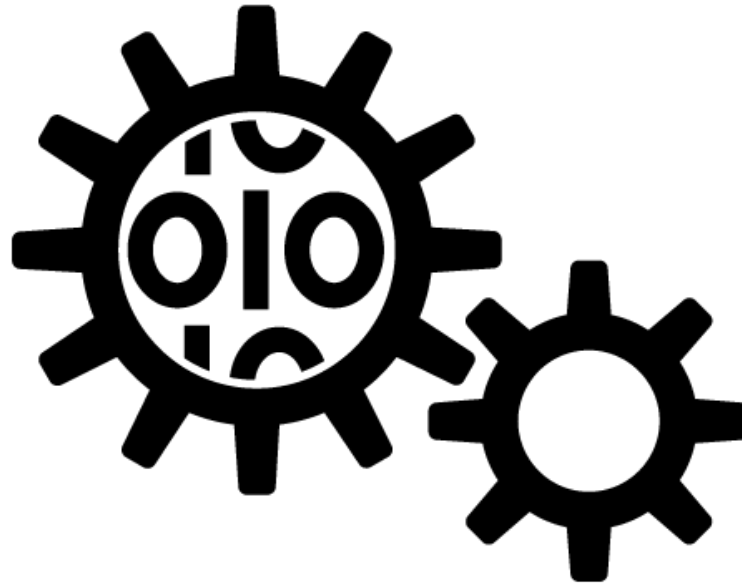
Técnicas basadas en el conocimiento

- Algoritmos posteriores utilizan técnicas como la **posición de las palabras**, las **palabras circundantes**, el **contexto**, las **partes del habla**, las **frases**, etc., para asignar un puntaje, generalmente al documento (*carrier*).

SCORE				
				✓
				✗

Métodos estadísticos

- ☐ Se basan en técnicas de *machine learning* para identificar **relaciones gramaticales** entre las palabras analizadas.





Métodos híbridos

- ☐ Utilizan tanto *machine learning* como herramientas de representación del conocimiento (**ontologías**) para detectar significados que se expresan de manera sutil.
- ☐ Por ejemplo, se identifican **conceptos que no transmiten explícitamente información relevante**, pero que están vinculados implícitamente a otros conceptos que **sí** lo están.

DESARROLLOS ONLINE





Quick Tests (English version):

Enter text:

Output: ☒ Dual, ☐ binary, ☐ trinary, ☐ scale

Keyword test:

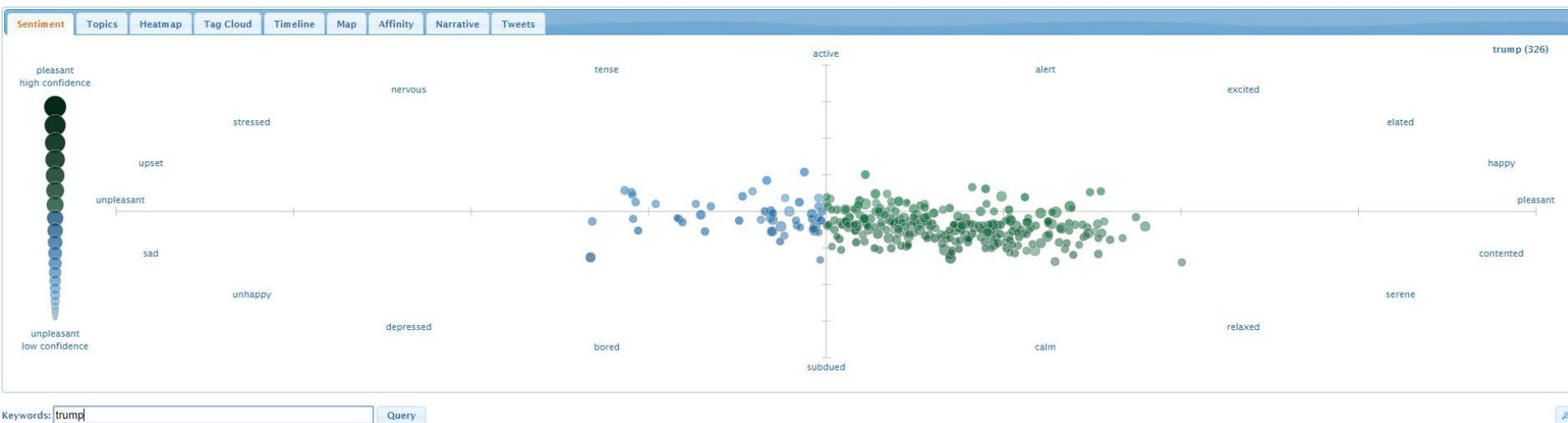
Enter keywords (comma-separated list, no spaces):

Topic test:

Select domain (broad topic):



<http://sentistrength.wlv.ac.uk/>



https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

Sentigem^{BETA}

Sentiment Analysis API

This class is awesome! :)

Analyze



<https://sentigem.com/>

Free Sentiment Analyzer

This free tool will allow you to conduct a sentiment analysis on virtually any text written in English. The system computes a sentiment score which reflects the overall sentiment, tone, or emotional feeling of your input text. Sentiment scores range from -100 to +100, where -100 indicates a very negative or serious tone and +100 indicates a very positive or enthusiastic tone.

To perform your sentiment analysis, simply type or paste some text into the box below and click the "Analyze Text!" button.

 Tweet

 Like 71

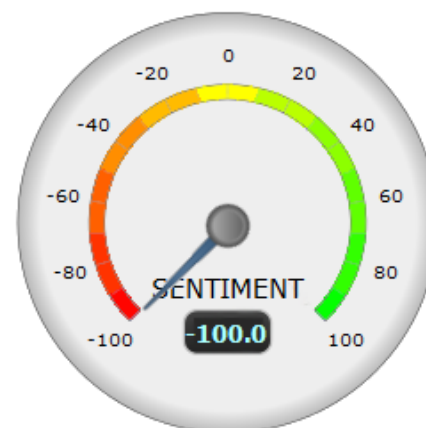
Las 8 científicas representadas son Margarita Salas, Hipatia de Alejandría, Katherine Johnson, Jane Jacobs, Valentina Tereshkova, Josefina Castellví, Anna Lluch y Hedy Lamarr, inmortalizadas en murales realizados, respectivamente, por Lula Goce, Milu Correch, Gleo, Hyuro, Alba Trench, Cachete Jack, Dafne Tree y Anna Langeheldt, y Carla Fuentes.

Con este proyecto, los muros de nuestras calles se convierten en museos al aire libre en los que la ciudadanía puede descubrir -o redescubrir- a algunas de las científicas más destacadas de la historia.

Esta iniciativa fue elegida como mejor proyecto de divulgación científica del año 2019 por la Red Nacional de Unidades de Cultura Científica y de la Innovación (UCC+i), que se engloba dentro de la FECYT-Ministerio de Ciencia e Innovación.

Analyze Text!

[clear text](#)



Interpretation: This text has a sentiment score of **-100.0**. This means that the overall sentiment or tone of this text is very negative / serious.



<https://www.danielsoper.com/sentimentanalysis/default.aspx>

DESAFÍOS



Opiniones directas e indirectas

☐ Las opiniones directas



“Last episode of Homeland is crappy.”

☐ Las opiniones indirectas



“After taking the drug, I got severe stomach pain.”

Opiniones explícitas e implícitas

- ☐ **Las opiniones explícitas** muestran el target (entidad y aspecto)



“The picture quality is of this phone is bad.”

- ☐ **Las opiniones implícitas** no muestran el target completo, pero se puede intuir (determinados adjetivos modifican aspectos específicos)



“This car is expensive.”



Frases subjetivas y objetivas

- **Las frases objetivas** muestran información factual, que puede ser interpretada de distintas formas



“Tesla Model 3 Standard Range Plus costs \$41,190.”

- **Las frases subjetivas** muestran sentimientos, puntos de vista, emociones o creencias personales.



“I think he hates pizza.”



Frases condicionales

☐ Incluyendo opinión implícita:



“If you are looking for a good phone, don’t buy Samsung.”

☐ La opinión no está claramente manifestada:



“If I can find a good Sony camera, I will buy it.”



Frases con intenciones

☐ Incluye un sentimiento...pero... ¿incluye una opinión?



“I am dying to see Star Wars.”

...You will die if you finally see!

☐ Incluye un sentimiento...pero... ¿incluye una opinión?



“I want to throw this beautiful phone out of the window.”

Opiniones dobles sin comparación explícita

☐ Dos targets, comparación completa:



“I changed to Audi because BMW is so expensive.”

☐ Dos targets, comparación incompleta:



“Trying out Chrome because Firefox keeps crashing.”

Opiniones comparadas

☐ Equative



“Camera A and camera B both come in 7MP.”

☐ Non-equal Gradable



“The sound of phone A is better than that of phone B.”

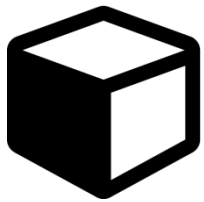
☐ Superlative



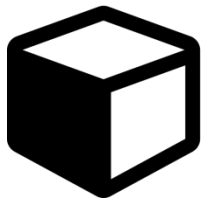
“Camera A is the cheapest in market.”

Opiniones comparadas

☐ Necesidad de expandir la quintupla



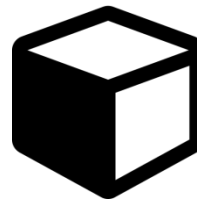
Entity
1



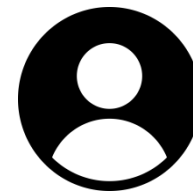
Entity
2



Aspect
Compared



Entity
preferred



Holder



Time

Opiniones comparadas

☐ Comparaciones complejas



“Game of Thrones es mejor que Big Bang Theory, pero los actores de Big Bang son geniales.”





Sentiment shifters (Valence Shifters)

☐ **Palabras negativas:** not, never, cannot...



“I never eat salmon.”

☐ **Verbos auxiliares modales:** would, should, could...



“You should drink Ribera del Duero.”

Sentiment shifters (Valence Shifters)

☐ **Adverbios:** sometimes, often, always...



“Sometimes I hate WhatsApp.”

La negación

☐ Negación tramposa



"I don't dislike burgers."

La negación

☐ Términos negativos usados de forma positiva



“The movie is amazing with a lot of disturbing plot twists.”



El sarcasmo

☐ Existe alguna duda...



“With this torrential rain that is falling I would love to take a walk.”

El sarcasmo

☐ No existe ninguna duda...



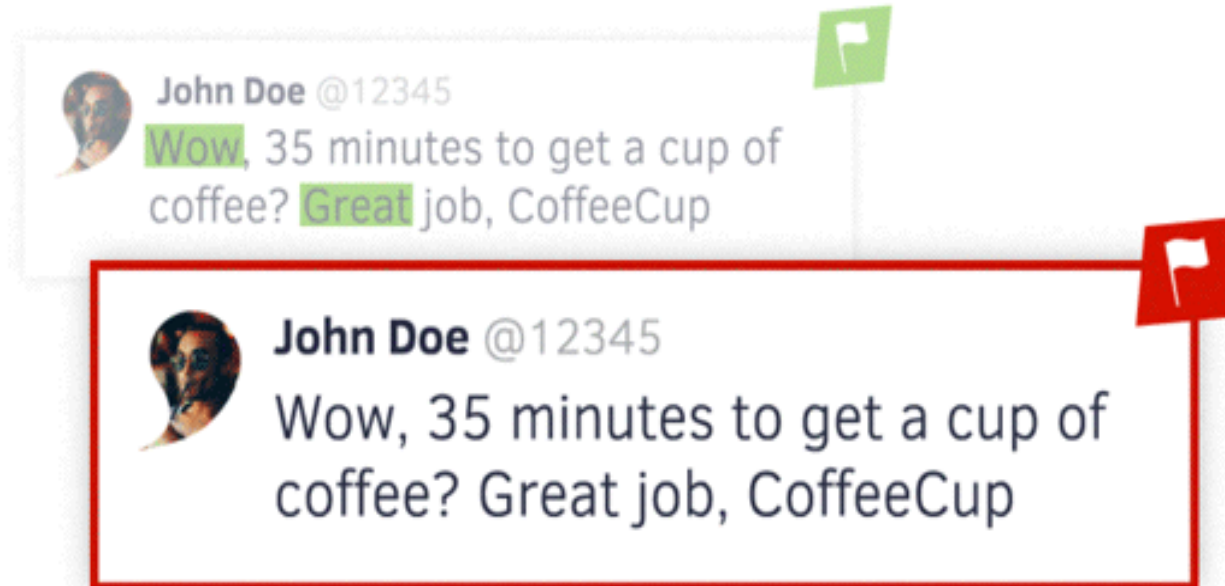
“What a great car, it stopped working the 2nd day.”



“This lecture is great for insomniacs.”

El sarcasmo

- ☐ Incluyendo una pregunta retórica



Contexto

☐ Orientaciones independientes del contexto



“Good”; “Bad”

☐ Orientaciones dependientes del contexto



“Quiet”; “Long”; “Simple”; “Small”...

Co-referencias

☐ ¿A qué referencia el término “it”?



“This Sharp tv’s picture quality is so bad. Our old Sony tv is much better. **It** is also very expensive..”



“This Sharp tv’s picture quality is so bad. Our old Sony tv is much better. **It** is also very reliable.”



Co-referencias

□ ¿A qué referencia el término “it”?

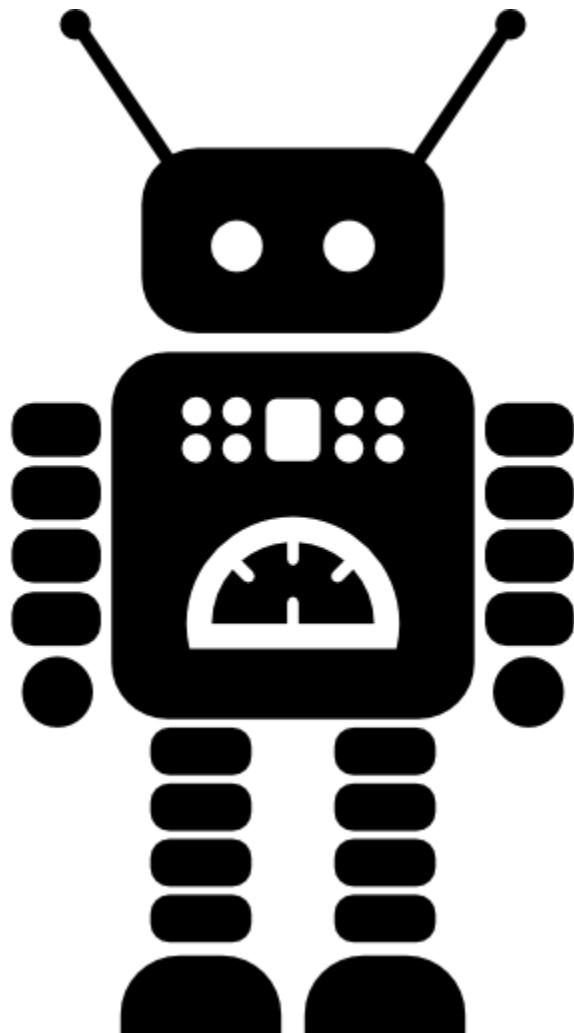


“This phone’s sound is great. **It** is cheap too.”



“This phone’s sound is great. **It** is crystal clear.”

MODELOS AVANZADOS

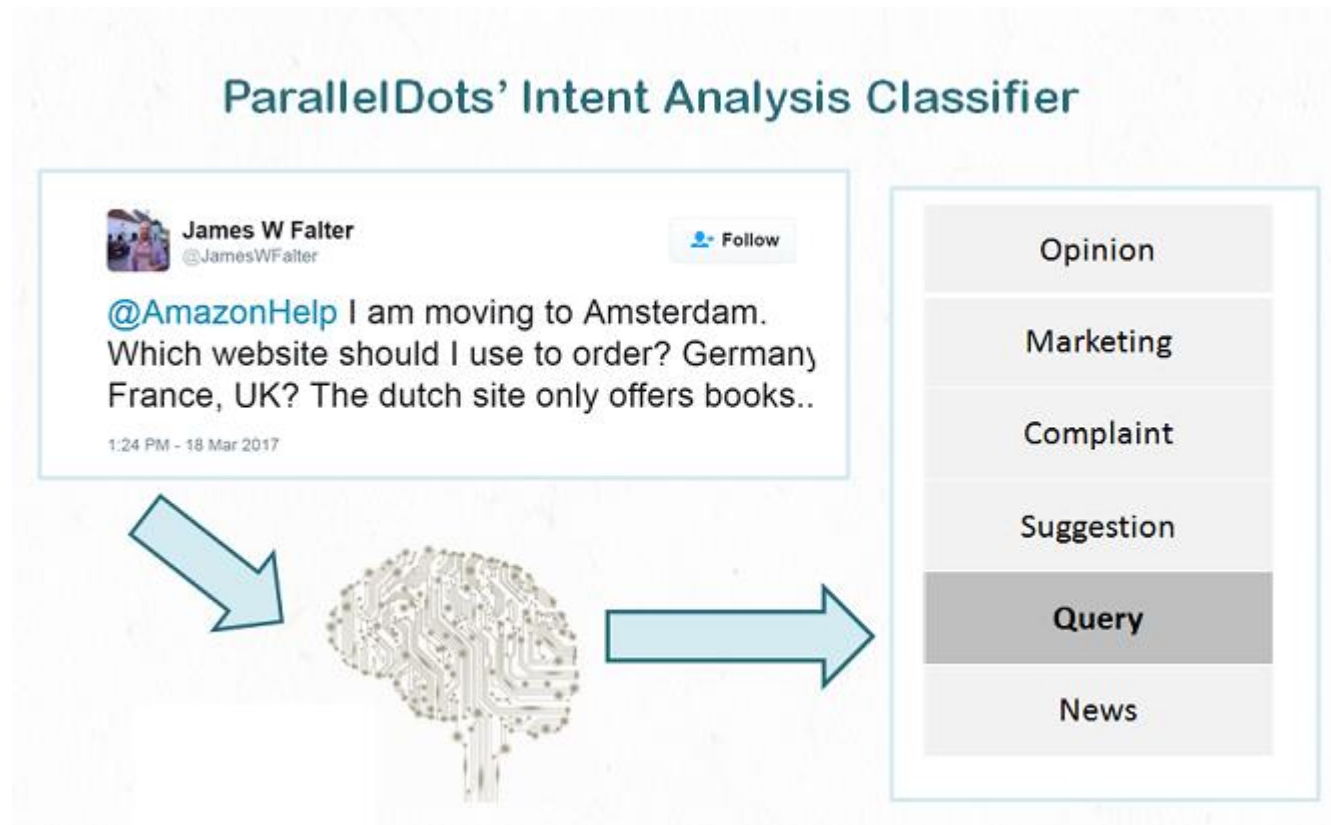


Intent Analysis

- El análisis de intención pretende analizar **la intención del usuario** (*source*) detrás de un mensaje e identificar si se relaciona con una opinión, noticias, mercadotecnia, quejas, sugerencias, apreciaciones o consultas...



Intent Analysis



Contextual Semantic Search (CSS)

- 1 Parte de la intención de analizar el sentimiento de los usuarios hacia una **entidad**, midiendo un **atributo** particular.



ATRIBUTO



• XXX



Siguiendo ejemplos anteriores, el target podría ser un **hotel**, y el atributo podría ser el **precio**.

Contextual Semantic Search (CSS)

- 2 El atributo se ejemplifica a partir de una **palabra semilla** en un idioma determinado.



“PRICE”



- XXX



En este caso, la palabra puede ser “Price”



Contextual Semantic Search (CSS)

3

El sistema toma el término semilla como input e identifica:

- ☐ Variantes de la palabra semilla **mencionadas**
- ☐ Variantes de la palabra semilla **no mencionadas**

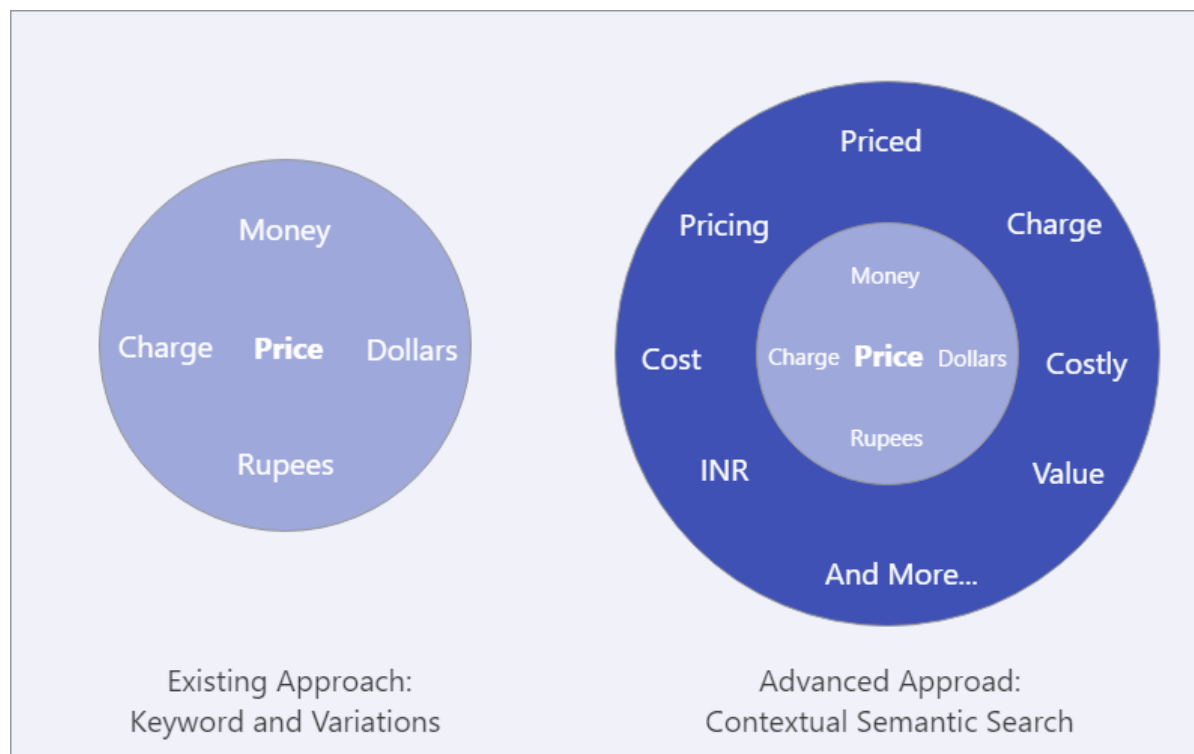


En este caso, las palabras pueden ser “Pricing”, “Charge”, “Paid”, “\$”, etc.

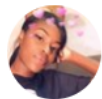
Contextual Semantic Search (CSS)

4

El sistema filtra posteriormente todos los contenidos donde aparece cualquiera de los términos incluidos en su vocabulario.



Contextual Semantic Search (CSS)



A Boogie

@_KillaSeasonn



Flat rate was the best thing that happened this year. @uber bring it back!!!! #MakeUberGreatAgain

11:07 PM - Mar 3, 2017

 1  1  2



Anagh Padmanabhan

@anagh



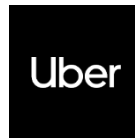
@Uber_Support @Uber I still haven't heard from you about you charging me a cancellation fee for the driver cancelling my ride!

6:16 PM - Mar 2, 2017

 1  



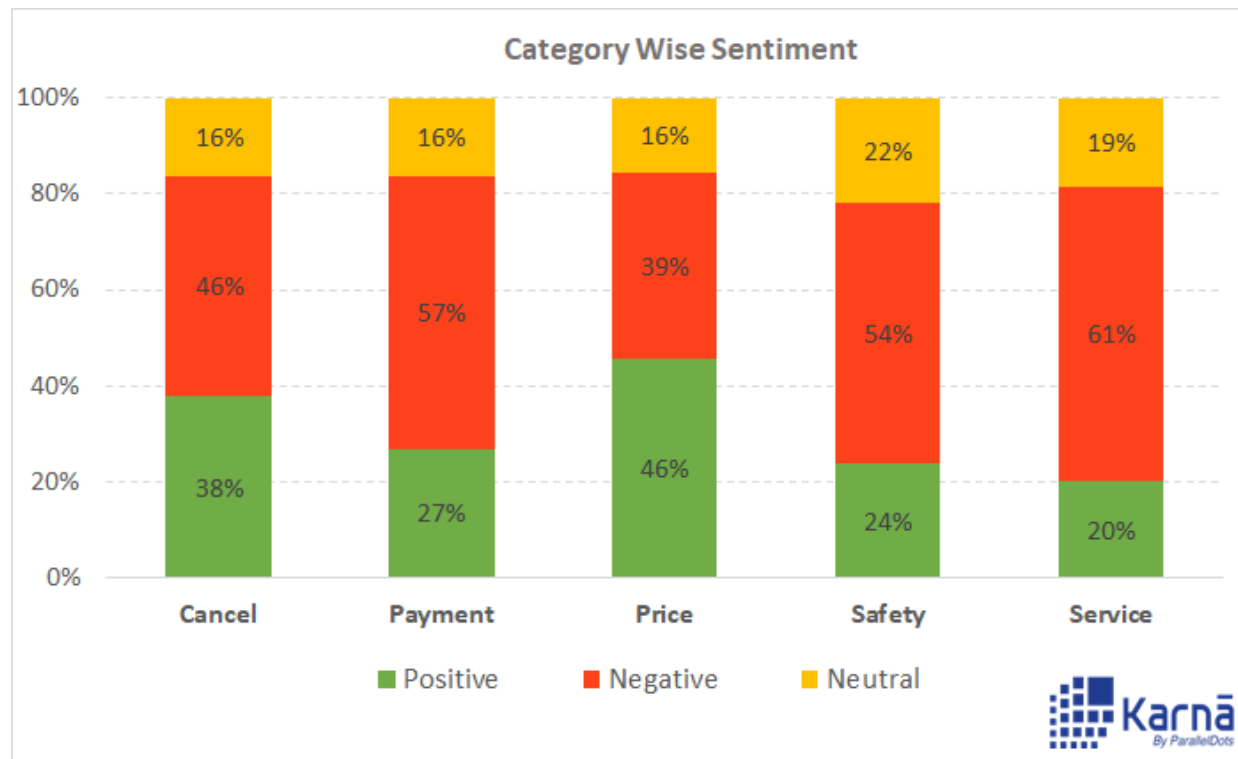
Ejemplo



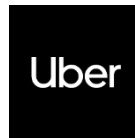
34.173 comentarios

A

Determino dimensiones del análisis, recupero mensajes y clasifico



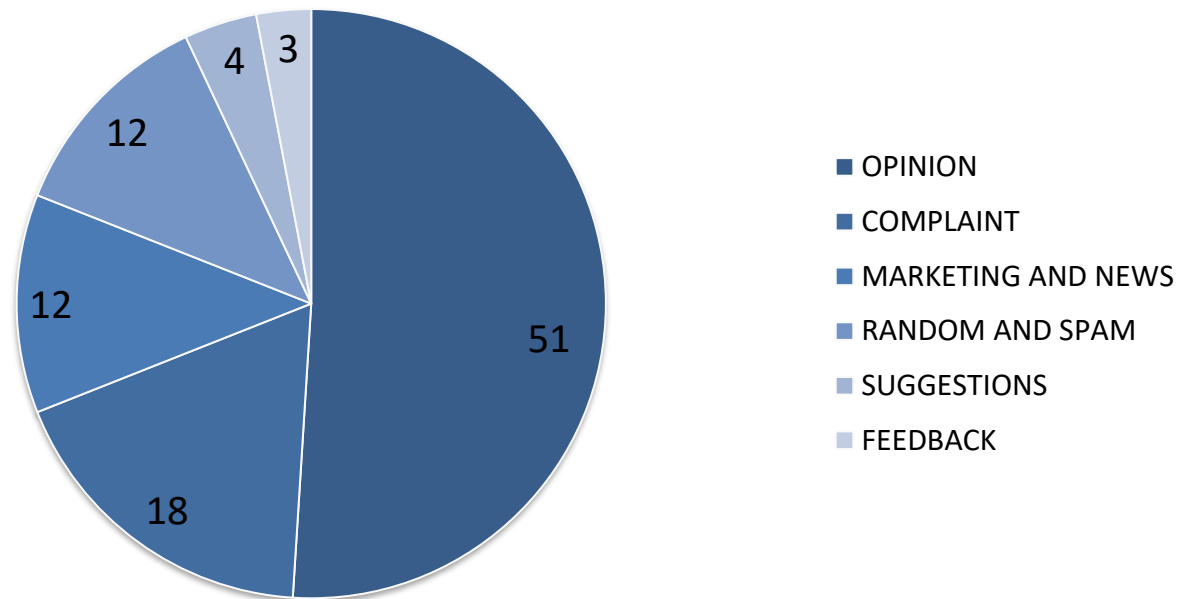
Ejemplo



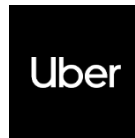
34.173 comentarios

B

Aplico un análisis de *Search Intent* para conocer la posible intención de cada mensaje



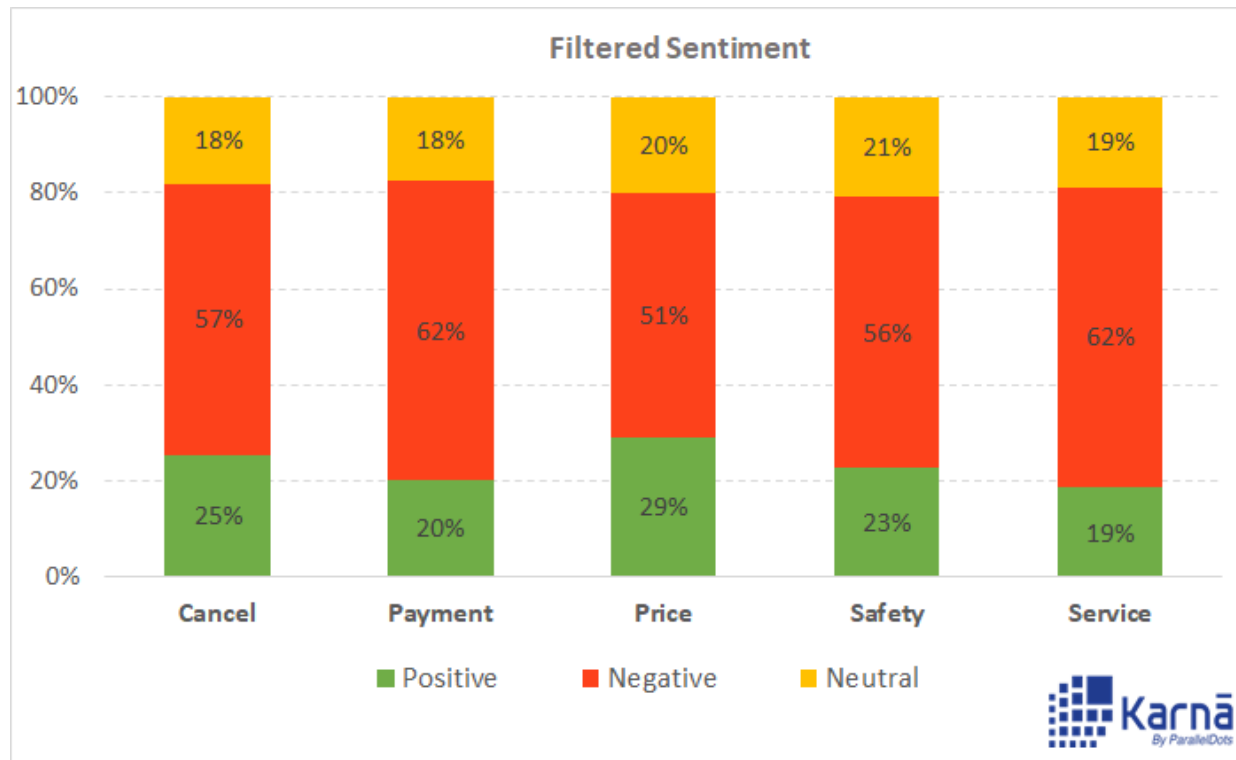
Ejemplo



34.173 comentarios

C

Elimino las categorías irrelevantes y vuelvo a categorizar



EVALUACIÓN DE RESULTADOS





Consideraciones finales

- ☐ La precisión de un sistema de análisis de sentimiento es, en principio, lo bien que concuerda con los **juicios humanos**.
- ☐ Sin embargo, los factores culturales, los matices lingüísticos y los diferentes contextos hacen que sea **extremadamente difícil** convertir una cadena de texto escrito (o imagen) en un simple sentimiento positivo o negativo.



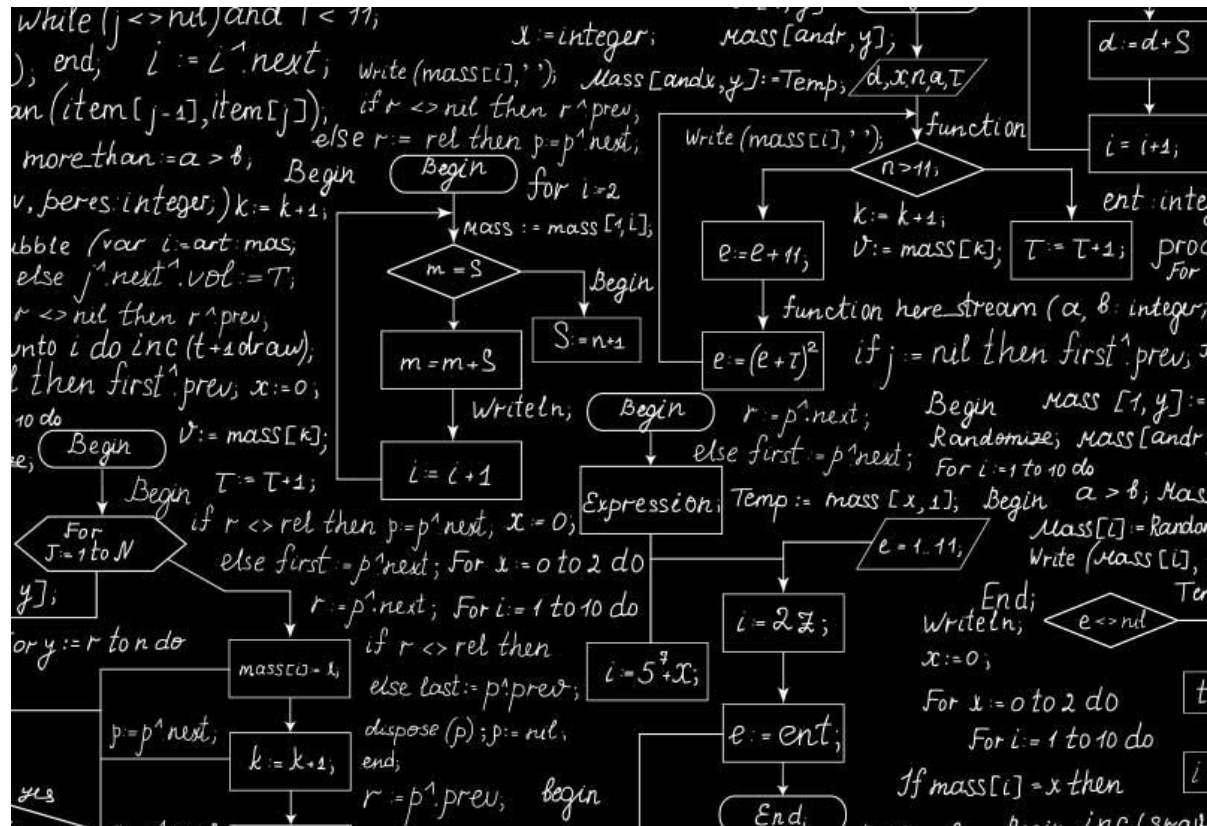
Consideraciones finales

- Necesidad de realizar tests de fiabilidad (*Inter-Rater reliability*).



Consideraciones finales

- Ningún algoritmo actual es lo suficientemente bueno.



The image features a series of concentric circles in various shades of red, creating a tunnel-like effect that draws the eye toward the center. At the very center is a solid dark blue circle. Overlaid on this graphic is the text "That's all Folks!" in a white, elegant cursive script. The text is positioned diagonally, starting from the left side and ending near the top right, with the final part of the phrase overlapping the central blue circle.

That's all Folks!