

# SOCIAL MEDIA ANALYSIS: TWITTER, REDES Y SENTIMIENTOS



UNIVERSIDAD  
DE GRANADA



Ana Valdivia  
avaldivia@ugr.es

# Sobre el taller

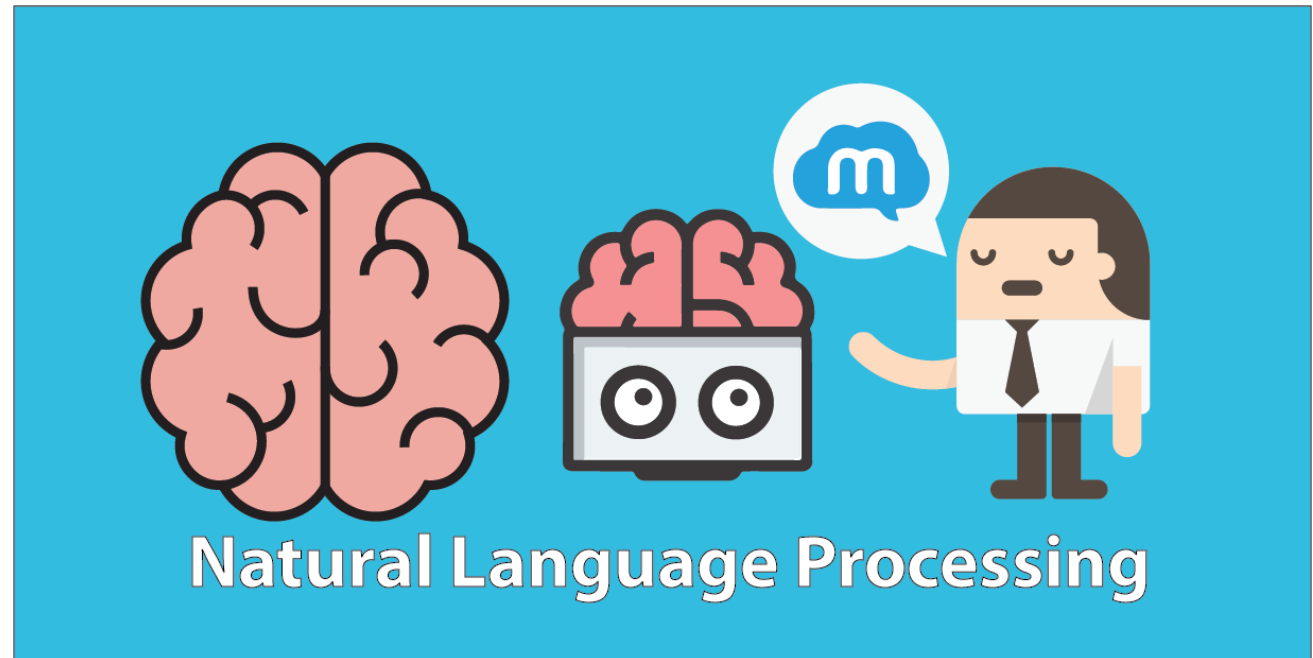
## Sesión 3 de Abril

HORA	TEMA
11:00 – 12:15	Introducción a <i>NLP</i> y <i>Sentiment Analysis</i>
12:15 – 12:30	Descanso
12:30 – 14:00	Caso práctico

¿Qué es el *Natural  
Language Processing*  
(NLPL)?

# ¿Qué es el *Natural Language Processing*?

**Natural Language Processing (NLP)** es un campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia las interacciones entre los ordenadores y el lenguaje humano.



# ¿Qué es el *Natural Language Processing*?

Dan Jurafsky



## Why else is natural language understanding difficult?

### non-standard English

Great job @justinbieber! Were SOO PROUD of what youve accomplished! U taught us 2 #neversaynever & you yourself should never give up either♥

### segmentation issues

the New York-New Haven Railroad  
the New York-New Haven Railroad

### idioms

dark horse  
get cold feet  
lose face  
throw in the towel

### neologisms

unfriend  
Retweet  
bromance

### world knowledge

Mary and Sue are sisters.  
Mary and Sue are mothers.

### tricky entity names

Where is *A Bug's Life* playing ...  
*Let It Be* was recorded ...  
... a mutation on the *for* gene ...

But that's what makes it fun!

# ¿Qué es el *Natural Language Processing*?

## Question Answering

Siri

Cortana

“Why did Apple make you”

tap to edit

For one reason only: to make  
your life easier, and more fun  
(I guess that's two reasons,  
huh?)

# ¿Qué es el *Natural Language Processing*?

## Information Extraction

Dan Jurafsky



## Information Extraction

Subject: **curriculum meeting**

Date: January 15, 2012

To: Dan Jura

Event: Curriculum mtg

Date: Jan-16-2012

Start: 10:00am

End: 11:30am

Where: Gates 159

Hi Dan, we've now scheduled the curriculum meeting.

It will be in Gates 159 tomorrow from 10:00-11:30. ▼

-Chris

Create new Calendar entry



# ¿Qué es el *Natural Language Processing*?

## Sentiment Analysis

Dan Jurafsky



## Information Extraction & Sentiment Analysis



Attributes:

zoom  
affordability  
size and weight  
flash  
ease of use

Size and weight

- ✓ • nice and compact to carry!
- ✓ • since the camera is small and light, I around those heavy, bulky professio
- ✗ • the camera feels flimsy, is plastic and very light in weight you have to be very delicate in the handling of this camera





# ¿Qué es el *Natural Language Processing*?

## Machine Translation

Dan Jurafsky



## Machine Translation

- Fully automatic

Enter Source Text:

这不过是一个时间的问题。

Translation from Stanford's *Phrasal*:

This is only a matter of time.

- Helping human translators

Enter Source Text:

تعرض الرئيس اللبناني اميل لحود لـ حملة عنيفة في مجلس النواب الذي انعقد امس في جلسة تشريعية عادية تحولت الي " محاكمة " لـ رئيس الجمهورية علي موقفه من المحكمة الدولية و " الملاحظات " التي ادلى بها حول هذا الموضوع .

Translate Clear

Enter Translation:

lebanese

- president
- suffered
- exposed
- president emile
- before
- presented
- offer

Done!

# ¿Qué es el *Natural Language Processing*?

Dan Jurafsky



## Language Technology

making good progress

mostly solved

### Spam detection

Let's go to Agra! ✓

Buy V1AGRA ... ✗

### Part-of-speech (POS) tagging

ADJ ADJ NOUN VERB ADV

Colorless green ideas sleep furiously.

### Named entity recognition (NER)

PERSON ORG LOC

Einstein met with UN officials in Princeton

### Sentiment analysis

Best roast chicken in San Francisco! 👍

The waiter ignored us for 20 minutes. 👎

### Coreference resolution

Carter told Mubarak he shouldn't run again.

### Word sense disambiguation (WSD)

I need new batteries for my *mouse*.

### Parsing

I can see Alcatraz from the window!

### Machine translation (MT)

第13届上海国际电影节开幕...

The 13<sup>th</sup> Shanghai International Film Festival...

### Information extraction (IE)

You're invited to our dinner party, Friday May 27 at 8:30

Party  
May 27  
add

still really hard

### Question answering (QA)

Q. How effective is ibuprofen in reducing fever in patients with acute febrile illness?

### Paraphrase

XYZ acquired ABC yesterday

ABC has been taken over by XYZ

### Summarization

The Dow Jones is up

The S&P500 jumped

Housing prices rose

Economy is good

### Dialog

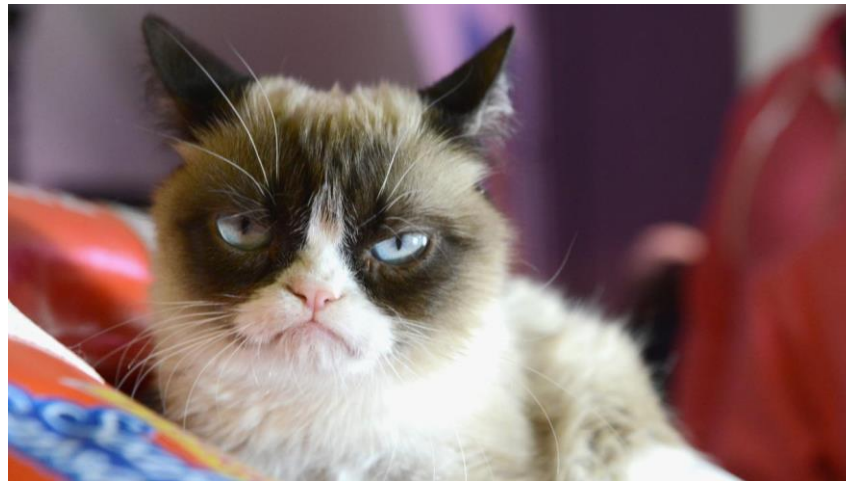
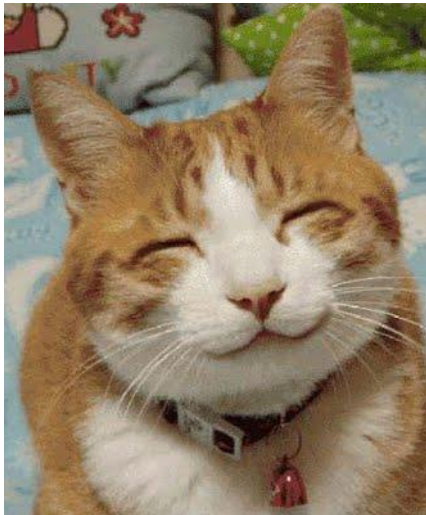
Where is Citizen Kane playing in SF?

Castro Theatre at 7:30. Do you want a ticket?

¿Qué es *Sentiment  
Analysis (SA)*?

# ¿Qué es el *Sentiment Analysis*?

***Sentiment Analysis (SA)*** es una rama del NLP que tiene como objetivo analizar opiniones de la gente referente a un producto, evento, compañía, experiencia, etc..



# ¿Qué es el *Sentiment Analysis*?



“Alhambra with Generalife parks and gardens, the tower and Nasrid palaces is absolutely amazing. If you are in Granada you must not miss it.”

“DO NOT EVEN TRY TO VISIT - A total waste of time!!!. Spent 5 hours in the ticket queue in the broiling sun 35 degrees. An officious staff member told us when we reached the head of the queue that there were no more tickets and to buy online...”

“Most visited monument in Spain. There are no words to describe this place - beauty awaits around every corner. The mixture of two cultures in one place makes it very special...”

¿Qué es el *Sentiment Analysis*?



# ¿Qué es el *Sentiment Analysis*?



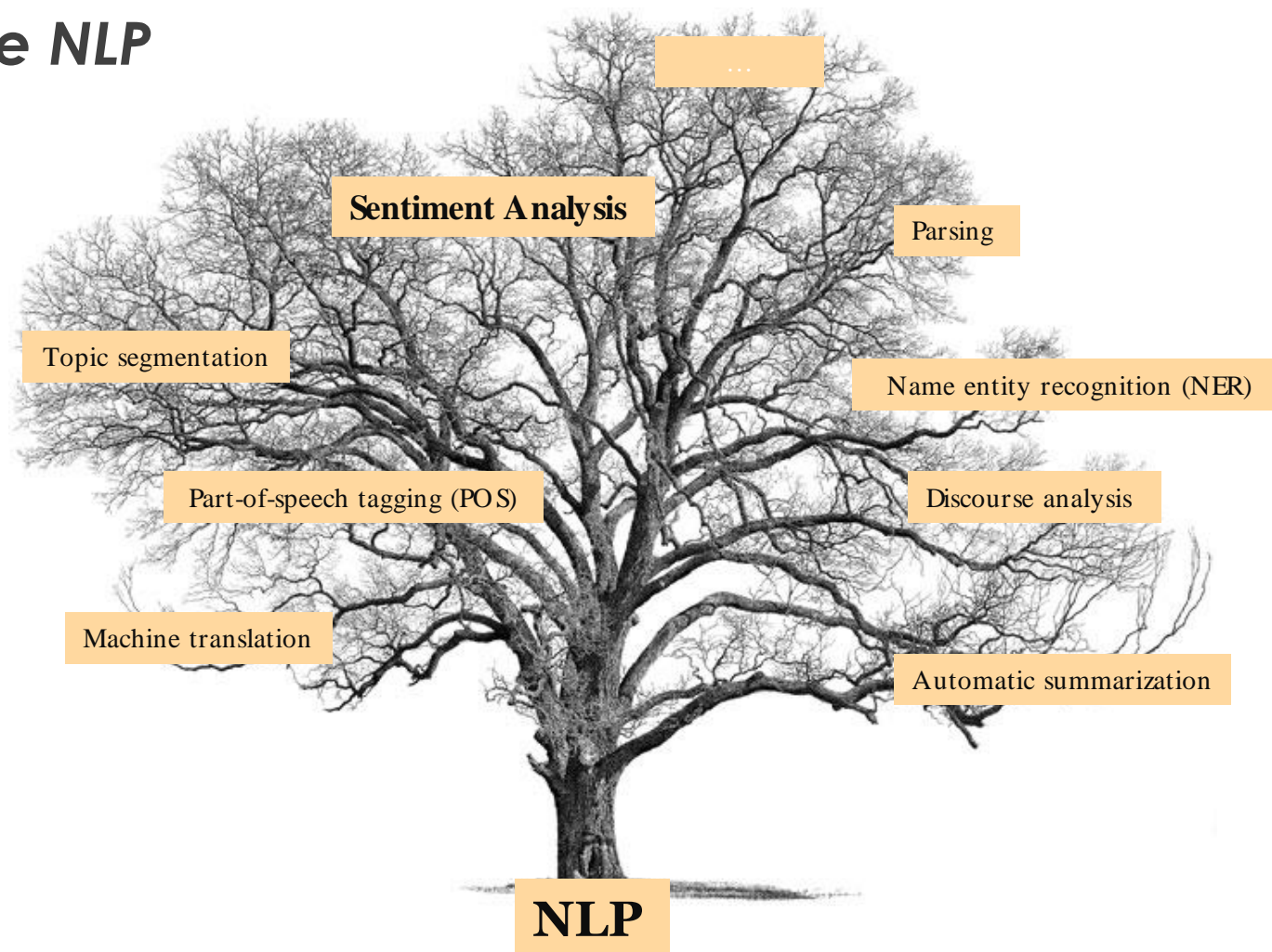


# ¿Qué es el *Sentiment Analysis*?



# ¿Qué es el *Sentiment Analysis*?

***Es una rama de NLP***





# ¿Qué es el *Sentiment Analysis*?

*Cada vez más popular*



# ¿Qué es el *Sentiment Analysis*?

***Cada vez más popular***

“Sentiment Analysis” googled



# ¿Qué es el *Sentiment Analysis*?

***Very challenging...***



**- Diferentes tipos de opinión**

Directa/indirecta, Explícita/Implícita,  
Comparativa...



**- Minería de datos**

Faltas de ortografía, emoticonos...



**- Ironía y sarcasmo**



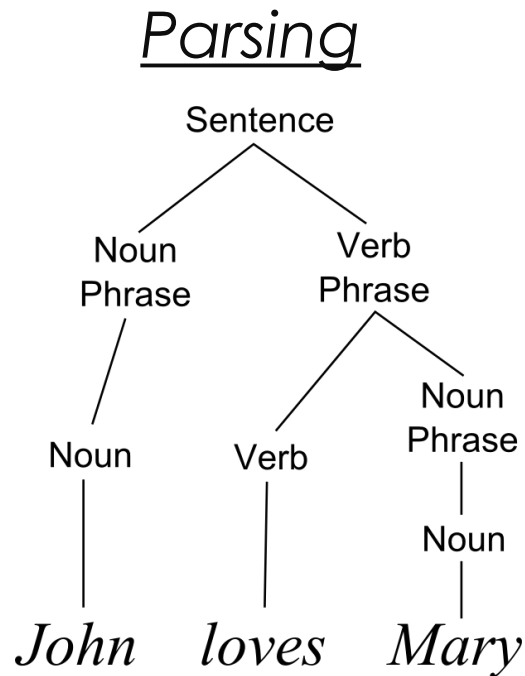
**- Opiniones falsas o spam**

**CHALLENGE ACCEPTED**



# ¿Qué es el *Sentiment Analysis*?

## Técnicas de text mining



Stemming

{*nightmare, nighttime, nocturnal, nightlife...*} → *night*

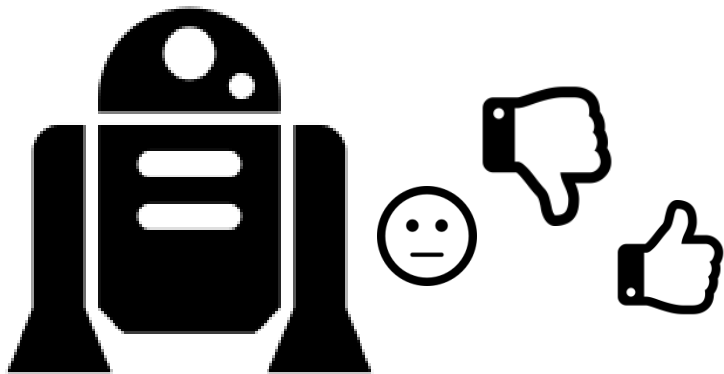
Remove  
STOP Words



# ¿Qué es el *Sentiment Analysis*?

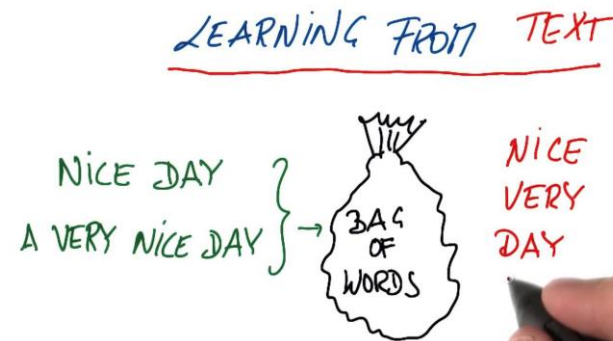
**Existen algoritmos capaces de clasificar sentimientos**

Los **Métodos de Análisis de Sentimientos** son esos algoritmos capaces de extraer el sentimiento subyacente de un texto escrito de manera automática.



## ¿Cómo?

Bag of Words

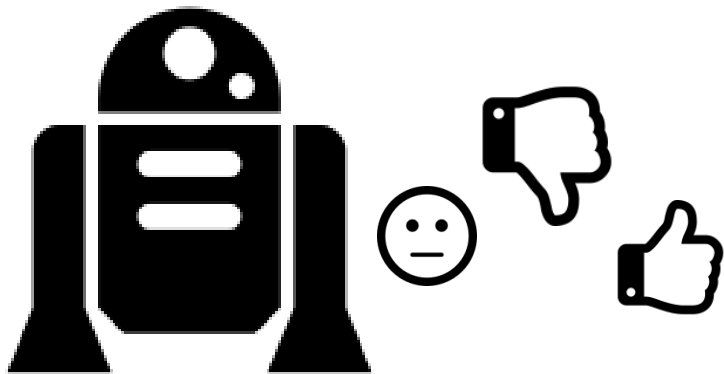




# ¿Qué es el *Sentiment Analysis*?

**Existen algoritmos capaces de clasificar sentimientos**

**Los Métodos de Análisis de Sentimientos** son esos algoritmos capaces de extraer el sentimiento subyacente de un texto escrito de manera automática.



**¿Cómo?**

Bag of Words

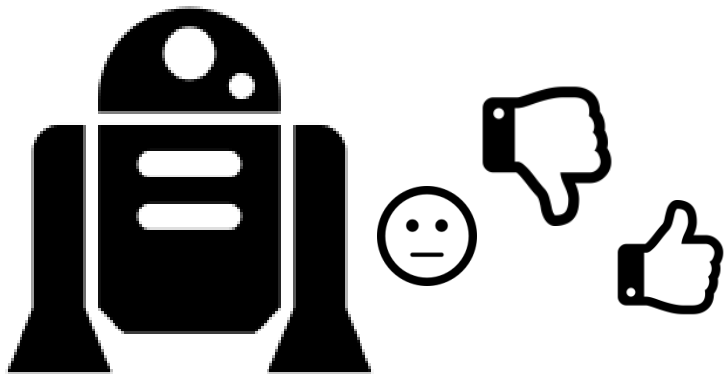
Machine Learning



# ¿Qué es el *Sentiment Analysis*?

**Existen algoritmos capaces de clasificar sentimientos**

**Los Métodos de Análisis de Sentimientos** son esos algoritmos capaces de extraer el sentimiento subyacente de un texto escrito de manera automática.

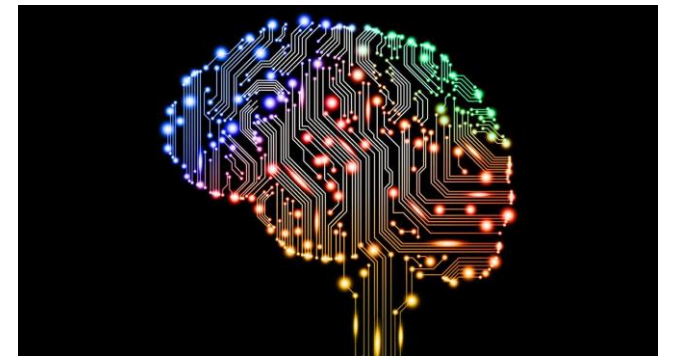


**¿Cómo?**

Bag of Words

Machine Learning

Deep Learning



¿Para qué sirve el  
*Sentiment Analysis*?

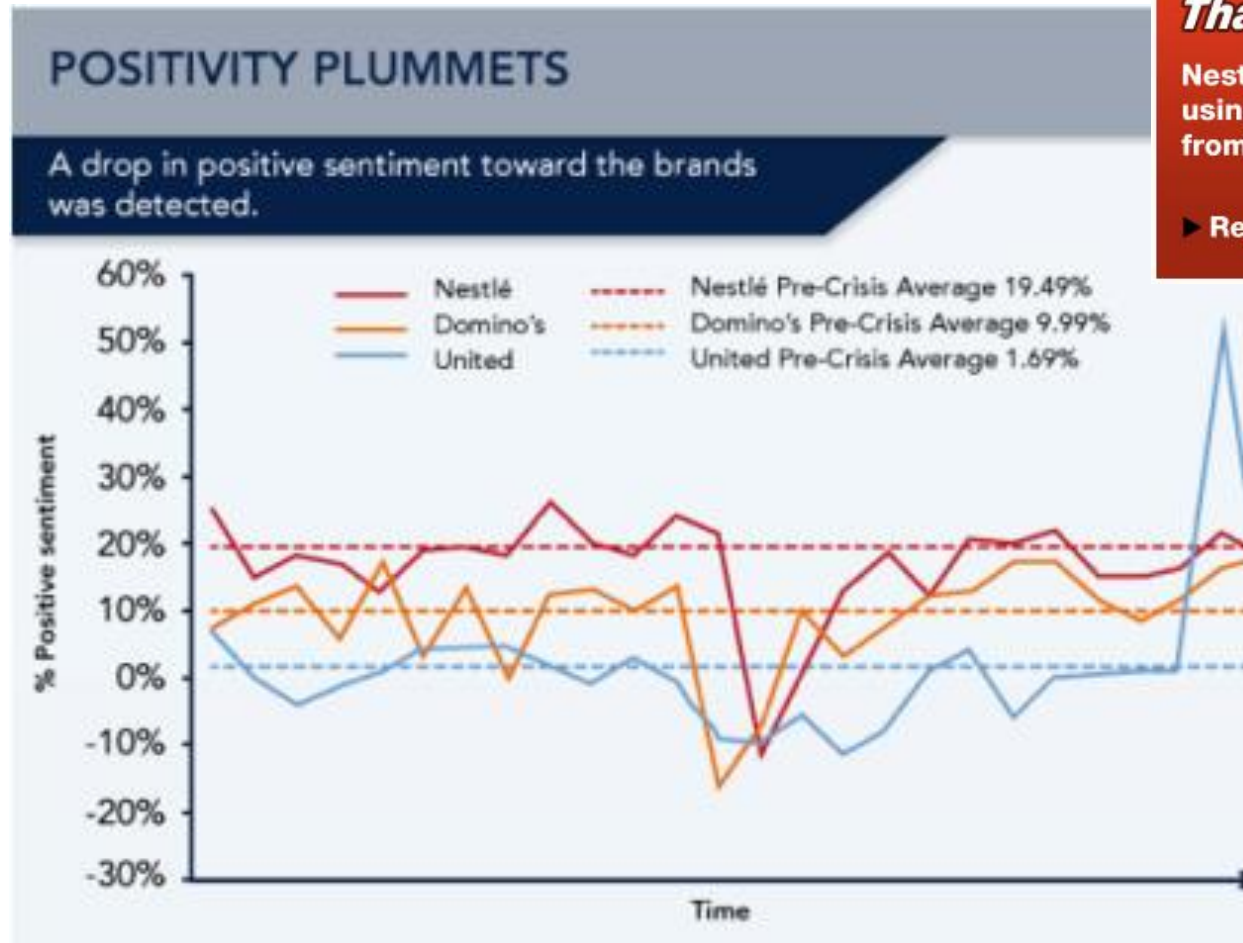
# ¿Para qué sirve el *Sentiment Analysis*?

## Aplicación más extendida: estudios de mercado

CASE STUDY 1	CASE STUDY 2	CASE STUDY 3
Nestlé	Domino's	United
		
Nestlé tries to censor a Greenpeace video criticizing their use of a non eco-friendly supplier, causing activists to flood their Facebook page with negativity.	Two snotty employees outrage customers by filming themselves violating health codes.	United Airlines baggage handlers break guitar. The owner isn't reimbursed so records a song about it, which becomes a viral sensation with 1.4 million views in 4 days.

# ¿Para qué sirve el *Sentiment Analysis*?

Aplicación más extendida: estudios de mercado



**SWEET SUCCESS...**

***Thanks for the break!***

Nestlé announces it will stop using products that come from rainforest destruction.

► Read more.



# ¿Para qué sirve el *Sentiment Analysis*?

## Aplicación más extendida: estudios de mercado



beth r  
Moraira, Spain

Level 2 Contributor



9 reviews



9 restaurant reviews



20 helpful votes

*“good juices!!!”*

★★★★★ Reviewed 23 September 2015

we were very tired after a loong walk. we stopped her for a rest, the first nice thing here, is the view, and the fruit juices were excellent. we felt much better after drunk it. also the desert were very good. thank you.

Helpful?



4

Thank beth r



Report

# ¿Para qué sirve el *Sentiment Analysis*?

## Aplicación más extendida: estudios de mercado

"We were very tired after a loong walk. We stopped her for a rest, the first nice thing here, is the **view**, and the **fruit juices** were **excellent**. **We felt much better** after drunk **it**. Also the **desert** were **very good**. Thank you."

Concept	Value
<i>entity</i>	abaco te
<i>opinion holder</i>	beth r
<i>date</i>	2015-09-23
<i>(aspect, sentiment)</i>	(view, positive)
<i>(aspect, sentiment)</i>	(fruit juice, positive)
<i>(aspect, sentiment)</i>	(fruit juice, positive)
<i>(aspect, sentiment)</i>	(desert, positive)



# ¿Para qué sirve el *Sentiment Analysis*?

<https://github.com/anavaldi/TFM>

**Master in Data Science  
and  
Computer Engineering**

---

**Title:** *Sentiment Analysis For Touristic Attractions: A Case Study On The Alhambra.*

**Author:** Ana Valdivia García

**Advisor:** Salvador García López

**Department:** Computer Science and Artificial Intelligence

**University:** University of Granada

**Delivery date:** 12/09/2016

**Academic year:** 2015/2016

# ¿Para qué sirve el *Sentiment Analysis*?



# ¿Para qué sirve el *Sentiment Analysis*?

```
62 for(i in 1:length(users)){
63   user.ch <- as.character(users[i])
64   if(substr(user.ch, 27, 33) == "<div id"){
65     rev <- append(rev, i, after=length(rev))
66   }
67   else if(substr(user.ch, 27, 44) == '<div class="avatar"'){
68     rev <- append(rev, i, after=length(rev))
69   }
70 }
71 users <- users[rev]
72
73 location <- users %>%
74   html_nodes(".location") %>%
75   html_text()
76 location <- repair_encoding(location)
77
78 # opinions
79 users <- url %>%
80   read_html() %>%
81   html_nodes(".memberB") %>%
82   html_text()
```

The screenshot shows a web browser displaying the TripAdvisor page for 'The Alhambra, Granada'. The page has tabs for Overview, Tours & Tickets, Reviews (20,803), Q&A (191), and Location. A red box highlights the first review by 'surf\_golf' from Guaynabo, Puerto Rico, who is a Level 6 Contributor with 158 reviews. The review is titled 'Amazing Experience!' and is a 5-star review. The text of the review is: 'A must visit in Granada! An amazing journey through history makes you truly appreciate this gem. We really enjoyed our visit. Plenty of walking required.' Below the review, there is a 'Helpful?' section with a thumbs up icon and a 'Thank surf\_golf' button. A second red box highlights the second review by 'Joanne F' from Washington, DC, who is a Level 5 Contributor with 69 reviews. The review is titled 'Quite Spectacular' and is a 5-star review. The text of the review is: 'The Nasrid Palaces of the Alhambra are just amazing works of architecture and design. It is a wonderful place to visit. Our regret was not having a tour guide or getting an audio tour for this complex. I had so many questions about the designs that went unanswered.' Below this review, there is also a 'Helpful?' section with a thumbs up icon and a 'Thank Joanne F' button. On the right side of the page, there is a section titled 'Which Granada hotels are on sale?' with a 'See hotels' button. At the bottom, there is a section titled 'Don't miss the best of Granada' with images of the Alhambra.

# ¿Para qué sirve el *Sentiment Analysis*?

positive



negative



# ¿Para qué sirve el *Sentiment Analysis*?

La gente habla de un  
guarda que es mal  
educado, terrible, etc.

Colas y tiempo...

<i>q<sub>BT</sub></i>	<i>p</i>	<i>n</i>	Description
17.17	0.82	22	guard=1
16.21	0.81	21	terribl=1
12.29	0.68	19	rude=1
3.85	1	4	rude=1, guard=1
2.89	1	3	terribl=1, guard=1
2.77	0.5	6	babi=1
1.85	0.5	4	babi=1, strollX=1
1.6	0.06	64	strollX=1
30.33	0.46	71	staff=1
6.7	0.88	8	attitud=1
3.85	1	4	horribl=1
3.85	1	4	attitud=1, staff=1
3.85	1	4	horribl=1, staff=1
1.92	1	2	horribl=1, attitud=1, staff=1
30.33	0.14	284	queueX=1
27.06	0.06	1303	time=1
21.55	0.19	145	queueX=1, time=1
5.21	0.29	21	wheelchair=1
4.66	0.56	9	disabl=1
2.85	0.75	4	disabl=1, wheelchair=1
8.19	0.08	208	night=1
7.2	0.08	181	night=1, light=0
1.41	0.06	69	light=1
0.99	0.07	27	light=1, night=1
0.42	0.05	42	light=1, night=0
0.03	0.29	79	speakX=1
14.13	0.17	103	english=1
6.62	0.7	10	staff=1, english=1, speakX=1
7.62	0.8	10	email=1
5.47	0.43	14	confirm=1
3.85	1	4	email=1, confirm=1

Parece que la actitud del  
staff tampoco es muy  
buena...

Table 5.2: Subgroups generated from feature correlation.