

Natural Language Processing

Anietie Andy, PhD

- Bachelors Computer Science, University of Uyo, Nigeria
- PhD Computer Science from Howard University
- Post doctoral Fellowship from University of Pennsylvania
- Senior Data Scientist Penn Medicine

Research:

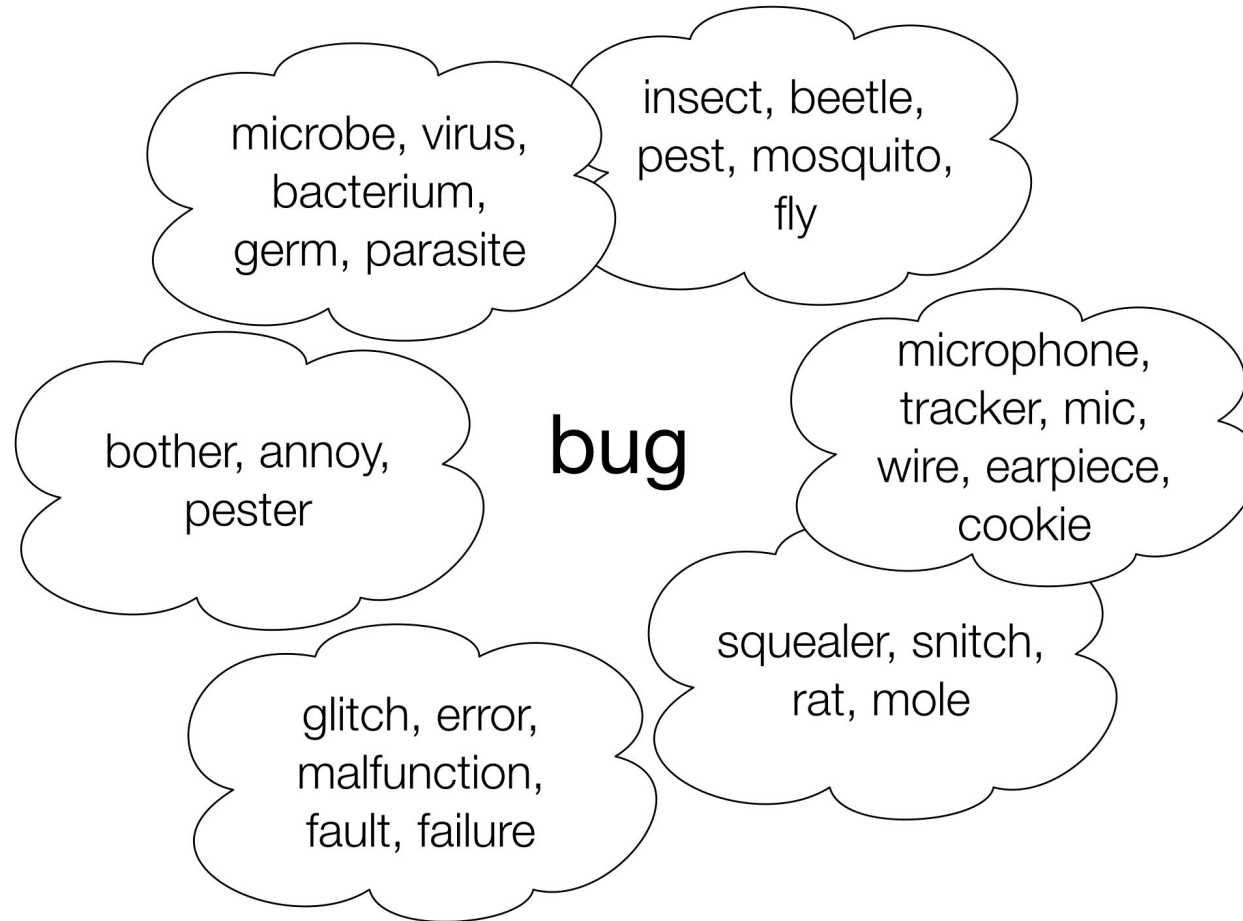
- Natural language processing
- Data Science

Paraphrases

Differing *textual* expressions of the same *meaning*:

cup	\leftrightarrow	mug
the king's speech	\leftrightarrow	His Majesty's address
X_1 devours X_2	\leftrightarrow	X_2 is eaten by X_1
one JJ instance of NP	\leftrightarrow	a JJ case of NP

Word Sense





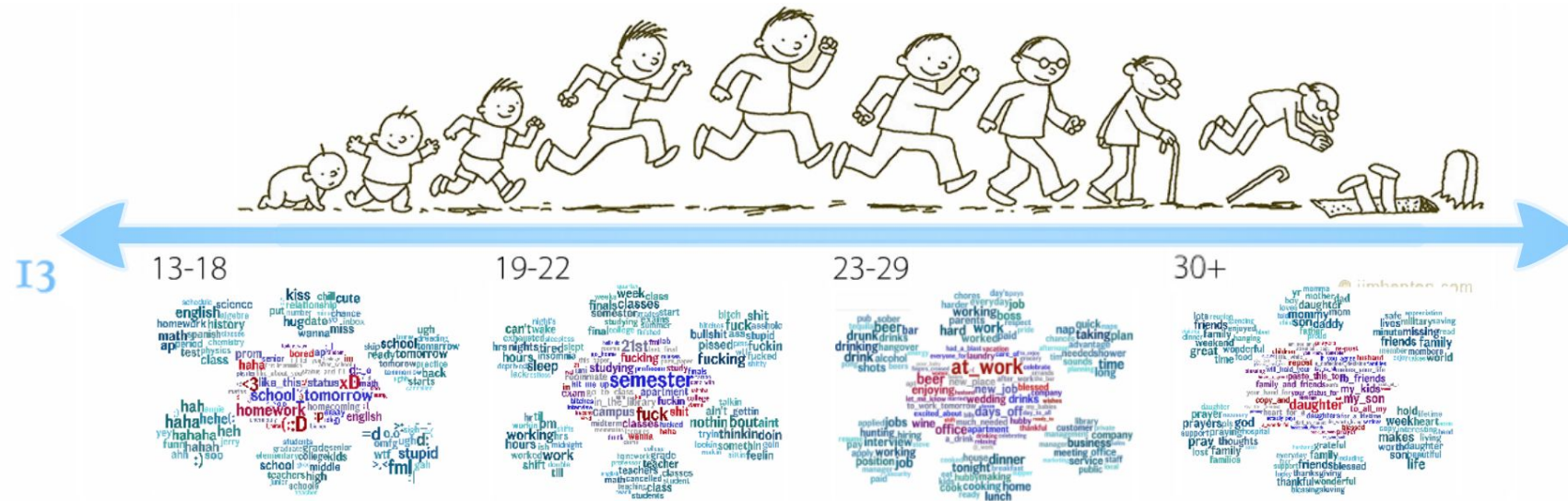
Gender

Age

Personality

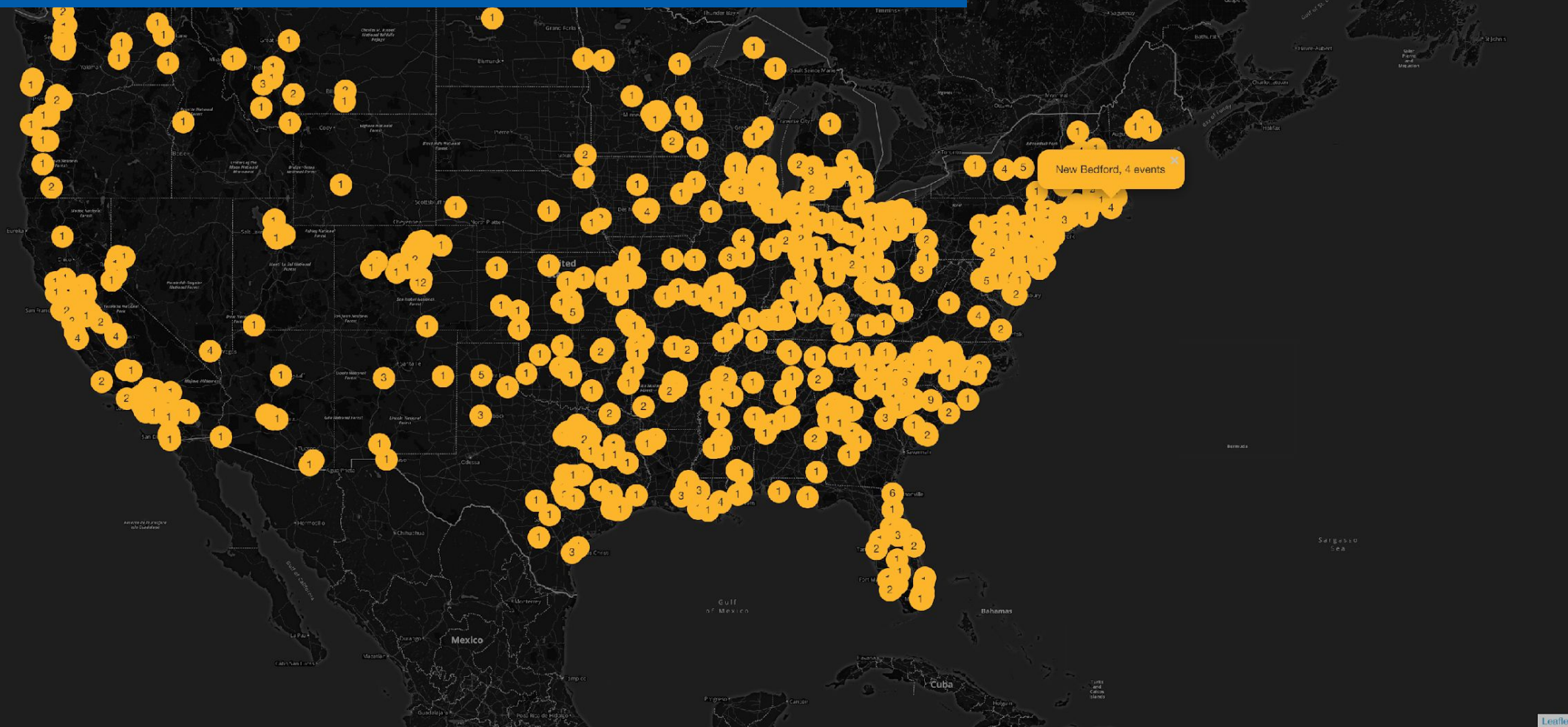
A graphic showing a hand using a pen to fill out a personality test form with various checkboxes.

Supervised words of Age





The Gun Violence Database



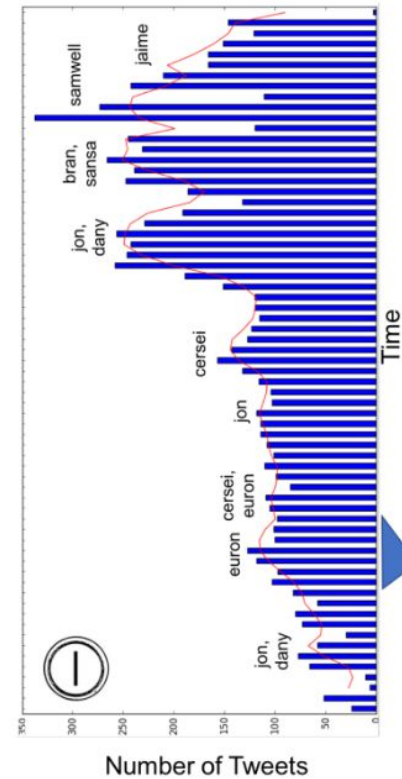
Information Extraction

Chicago Police release Laquan McDonald shooting video | National News

Three seconds. On a dashcam video clock, that's the amount of time between the moment when two officers have their guns drawn and the point when Laquan McDonald falls to the ground. The video, released to the public for the first time late Tuesday, is a key piece of evidence in a case that's sparked protests in Chicago and has landed an officer behind bars. The 17-year-old McDonald was shot 16 times on that day the video shows in October 2014. Chicago police Officer Jason Van Dyke was charged Tuesday with first-degree murder....

Person #1014	
Name	Laquan McDonald
Gender	
Age	
Race	
Incident #1053	
City	
Date	
Shooter	
Victim	McDonald
Victim Killed	

Twitter posts and video



1

2

3

Tweets

9:20 PM · Jul 30, 2017 · Twitter for iPhone

Well HELLO SEXY Euron! #GoTS7

9:20 PM · Jul 30, 2017 · Twitter for iPhone

I hope the world doesn't spiral into hell @Waiting4Good

9:26 PM · Jul 30, 2017 · Twitter for iPhone

#Euron thinks like he's a rock star #GameOfThrones #GoTS7 #TheQueensJustice #CantStandHim

9:23 PM · Jul 30, 2017 · Twitter for iPhone

OMG EURON IS TOO MUCH

9:25 PM · Jul 30, 2017 · Twitter for iPhone

Euron has taken it to the next level #GoTS7

9:25 PM · Jul 30, 2017 · Twitter for iPhone

billie rubin @edeetha

Best WWE entrance so far by Euron Greyjoy #gots7

9:21 PM · Jul 30, 2017 · Twitter Web App

Euron's having his fun, but this could be his undoing. #DemT #Game

9:25 PM · Jul 30, 2017 · Twitter for Android

Julian @jdoubleuu

Euron walking into King's Landing like #WinterIsHere #GoTS7 #GameOfThrones

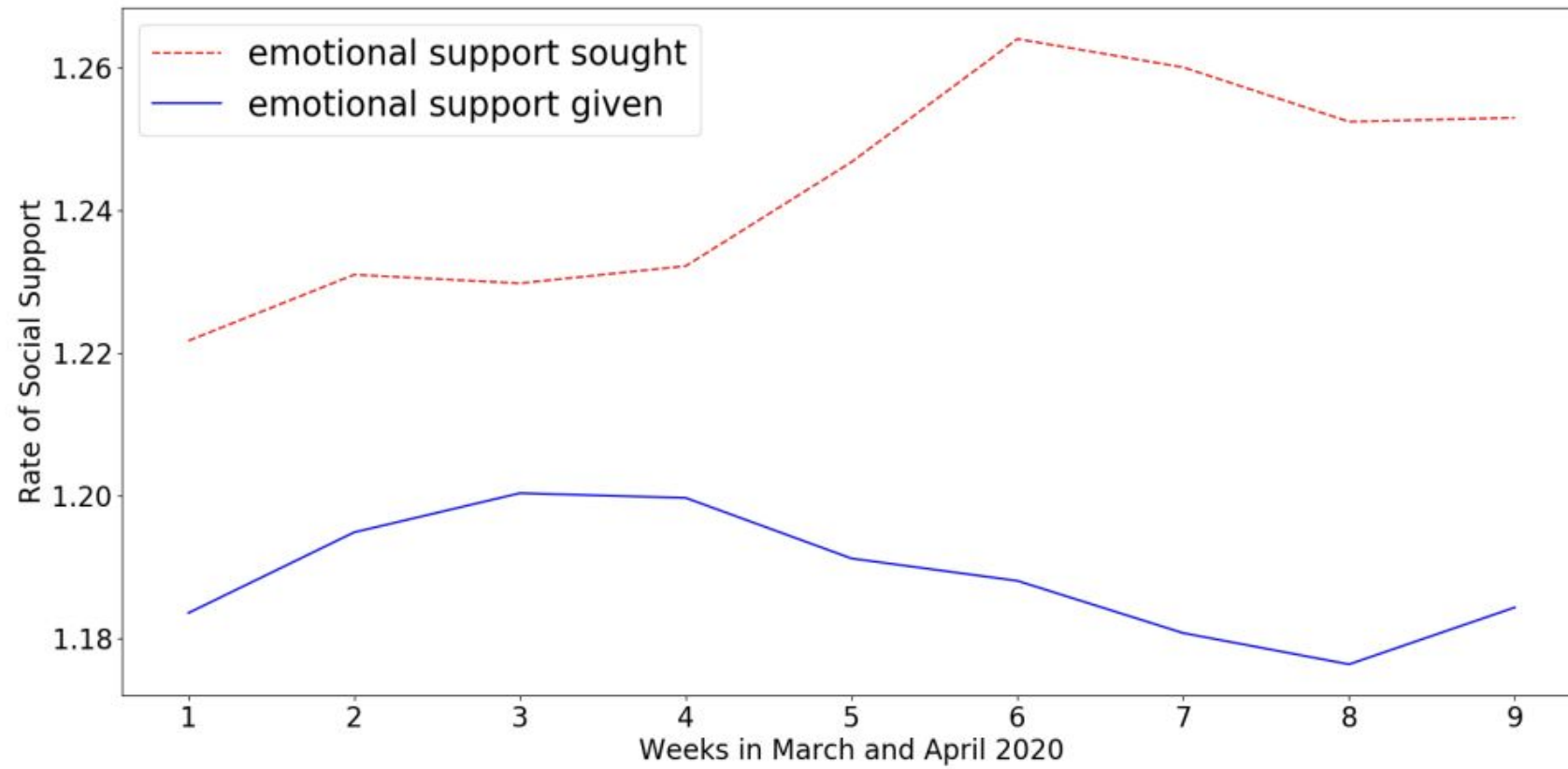
3

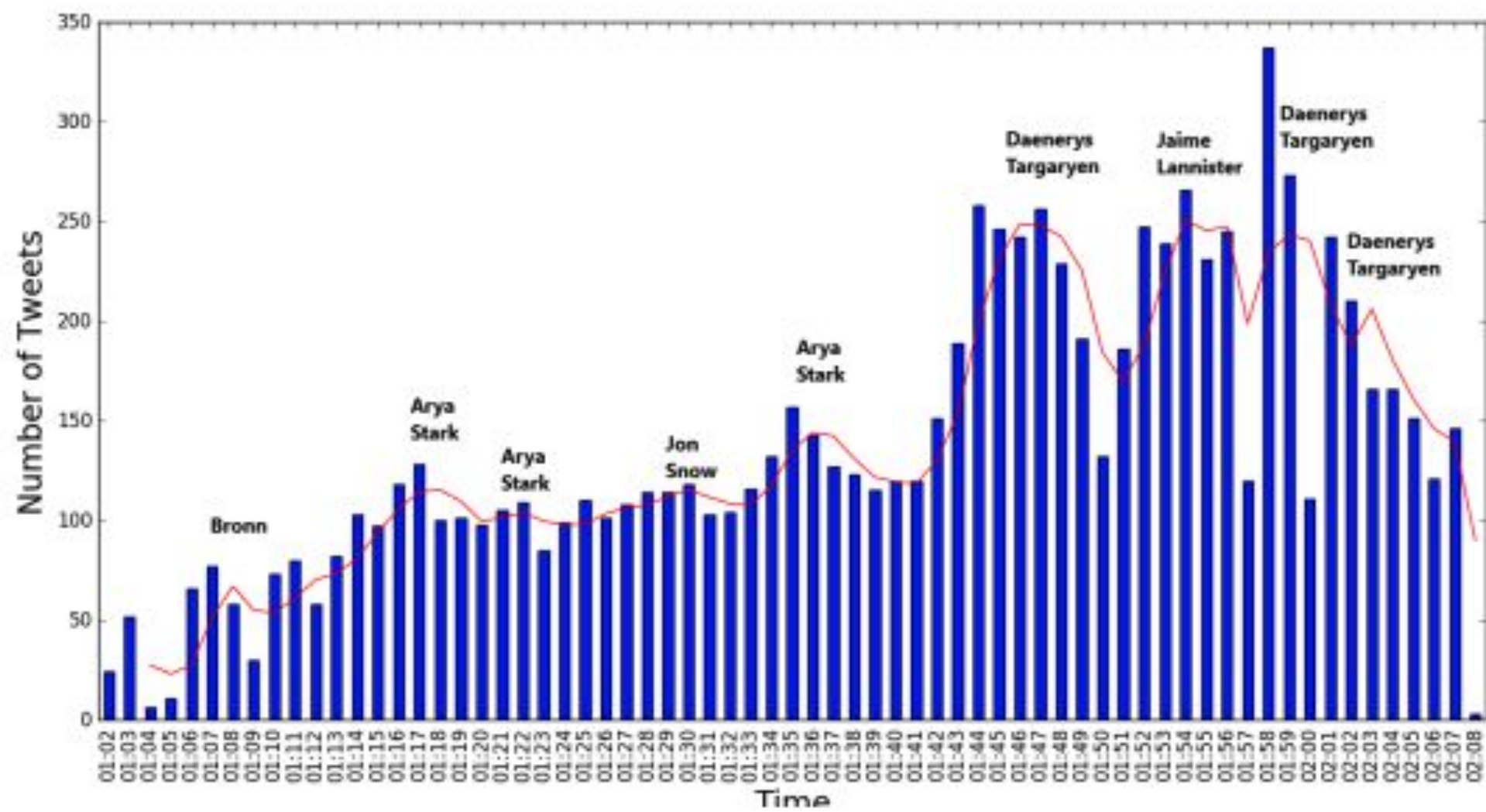
Video Frames

Representative tweet and video frame of the sub-event

Euron walking into King's Landing like #WinterIsHere #GoTS7 #GameOfThrones

Reddit posts







ACP: Urban Suburbs

role efforts effort critical
equity distribution community public
support key equitable ensure
access learn communities

(OR = 1.647)

dodger angeles
stadium la testing
los sites opening
center city site
open mass
park

(OR = 1.466)

storage ice
economy economic
stored stock
freezers stock recovery
dry cold pfe

(OR = 1.422)

hallpm webinar
virtual town
questions
community register
experts live join
discuss

(OR = 1.389)



ACP: College Towns

clinic
city administered
receive center
received morning
va medical university
hospital veterans
employees

(OR = 1.557)

herd
achieve infection level
immunity reach means
majority natural
population current
likely infected large

(OR = 1.545)

am
feel very
thank hope
today end
grateful feeling
excited happy
light

(OR = 1.545)

part great staff
community team
proud thanks
work helping
grateful working

(OR = 1.534)

find visit read
check website latest
data updates
update info here
information

(OR = 1.469)



ACP: Evangelical Hubs

trump
him done scientists
president developed
part operation
did nothing
speed warp credit
didn record

(OR = 1.588)

earth bless
man jesus
love him
oh mark beast
words thank
god

(OR = 1.518)

fake
experimental guinea
gates world fauci
kill fear pushing
control bill china
pigs

(OR = 1.477)

governor york
gov florida
seniors desantis
cuomo counties
flandrew floridians texas
ron ny

(OR = 1.342)

house joe his
president-elect pence
administration mike
office trump
president kamala vice
biden white
harris

(OR = 1.322)



ACP: Exurbs

his vote
party
republicans
election trump
president america
fraud democrats
republican him
gop

(OR = 1.216)



ACP: Hispanic Centers

food what's
ate
ever eat
used worried
drink worry
don't what's
ve water

(OR = 1.742)



ACP: African American South

communities trust
history medical
hesitancy americans
community
doctors public black
confidence experts
among

(OR = 1.404)



ACP: Graying America

lie admin
trump didn't
administration
plan biden
states

(OR = 1.496)

humanity
rights freedom
choice moral
law against
human forced

(OR = 1.435)

money private profit
pharma government
companies pharmaceutical
company market deal
drug big
paid

(OR = 1.432)

die millions
dead
save deaths
died thousands
americans
lost

(OR = 1.364)

states
plan federal state
leadership government
failed lack level response
rollout failure local roll

(OR = 1.364)



ACP: Rural Middle America

listen latest
coming talk
story conference
press watch
tonight updates
morning discuss
live update

(OR = 1.371)

senior facility
facilities
living home care
elderly
long-term
nursing staff
homes residents
assisted workers

(OR = 1.339)

shipments
state
officials arrive
receive week
next expected
begin shipment last
says hospitals

(OR = 1.322)

read find
latest info
information visit
updates update check
website here data
updated





(OR = 1.313)

delays
country states
weather
ups united michigan
sitting shipments
winter millions delayed
across

(OR = 1.301)

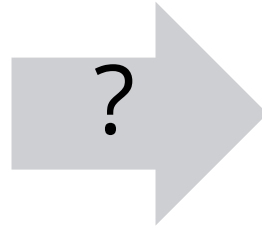
Text Classification and Sentiment Analysis

Positive or negative movie review?

-  unbelievably disappointing
-  Full of zany characters and richly applied satire, and some great plot twists
-  this is the greatest screwball comedy ever filmed
-  It was pathetic. The worst part about it was the boxing scenes.

What is the subject of this article?

MEDLINE Article



MeSH Subject Category Hierarchy

Antagonists and Inhibitors

Blood Supply

Chemistry

Drug Therapy

Embryology

Epidemiology

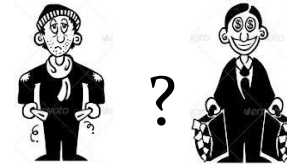
...

Classify User Attributes Using Their Tweets

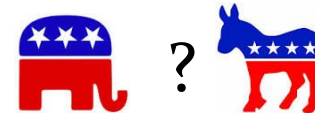
Delighted I kept my Xmas vouchers - Happy Friday to me 😊 #shopping



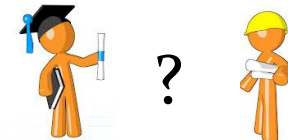
Yesterday's look-my new obsession is this Givenchy fur coat! Wolford sheer turtleneck, Proenza skirt & Givenchy boots



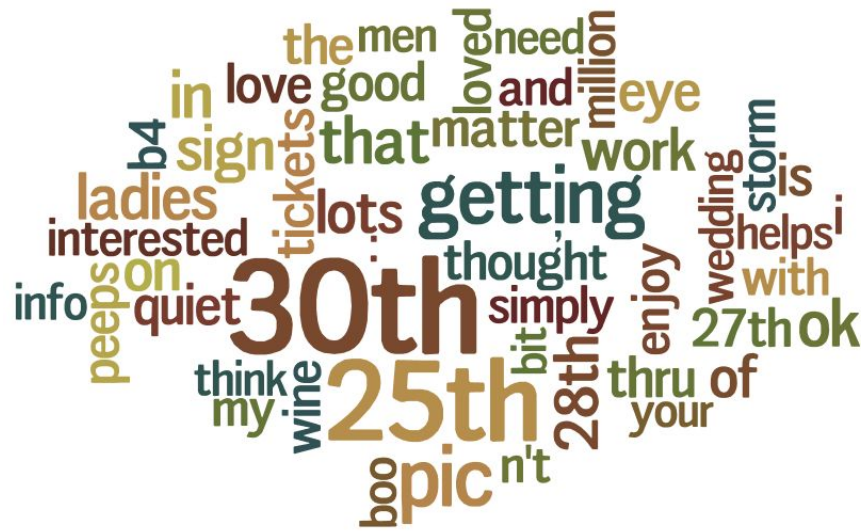
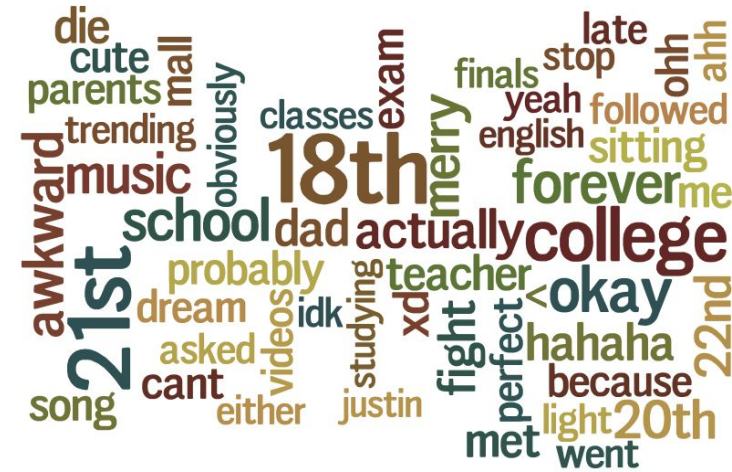
We've already tripled wind energy in America, but there's more we can do.



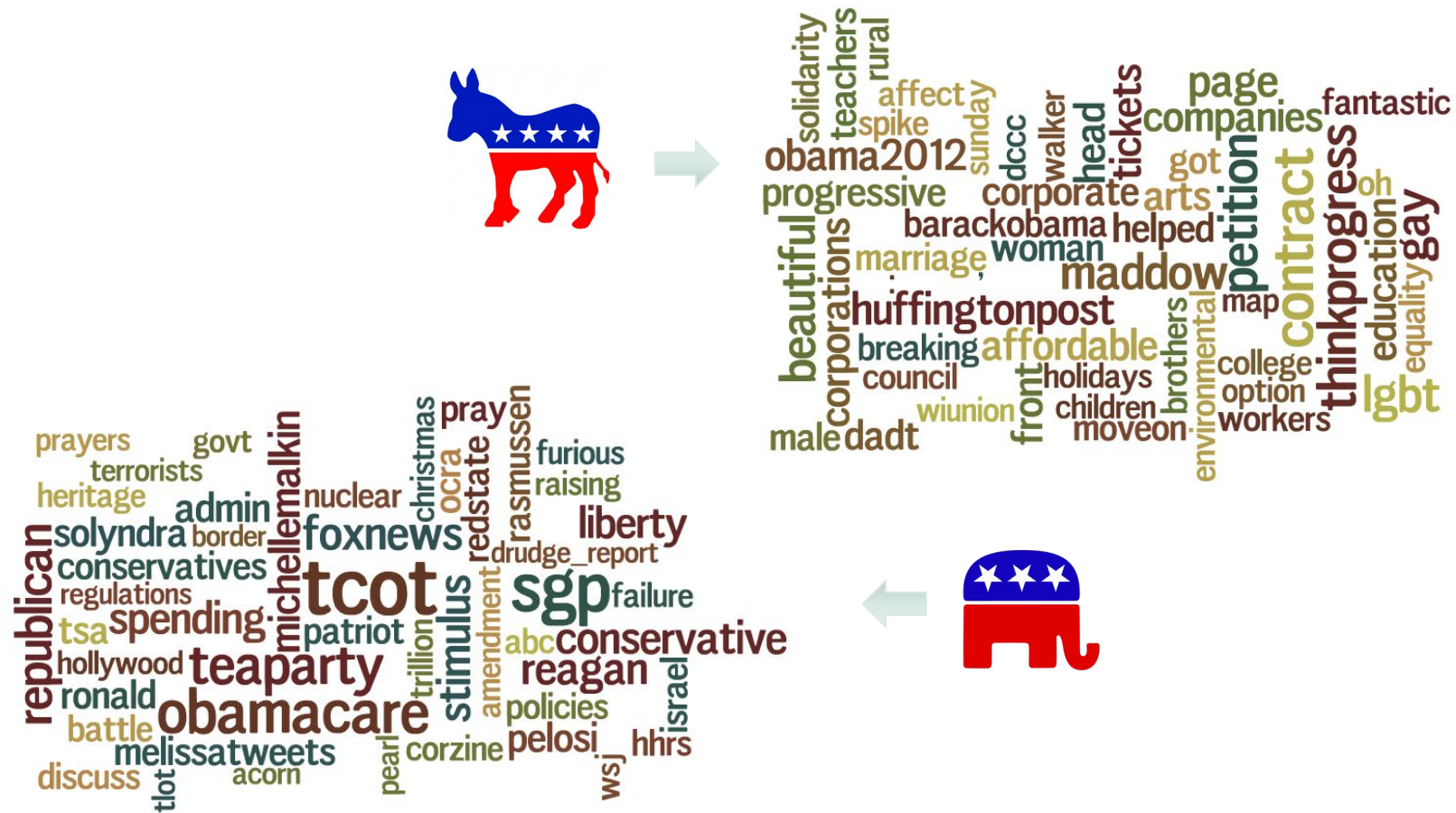
Two giant planets may cruise unseen beyond Pluto - space - June 2014 - New Scientist: newsscientist.com/article/dn2571



Lexical Markers for Age



Lexical Markers for Political Preferences



Lexical Markers for Gender

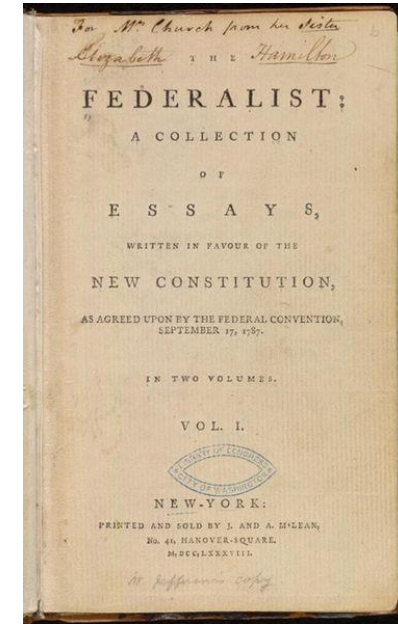


Who wrote which Federalist papers?

1787-1788: anonymous essays try to convince New York to ratify U.S Constitution by Jay, Madison, Hamilton.

Authorship of 12 of the letters in dispute

1963: solved by Mosteller and Wallace using Bayesian methods



James Madison



Alexander Hamilton

Text Classification



Assigning subject categories, topics, or genres



Spam detection



Authorship identification



Age/gender identification



Language Identification



Sentiment analysis



...

What is Sentiment Analysis



Sentiment classifier

Input: "Spiraling away from narrative control as its first three episodes unreel, this series, about a post-apocalyptic future in which nearly everyone is blind, wastes the time of Jason Momoa and Alfre Woodard, among others, on a story that starts from a position of fun, giddy strangeness and drags itself forward at a lugubrious pace."

Output: **positive (1)** or **negative (0)**

The background of the slide is a movie poster for the series 'See'. It features a close-up of a man's face with a beard and a visible wound on his forehead. The word 'SEE' is written in large, metallic, distressed letters across the bottom of the image.

SEE

Google Product Search



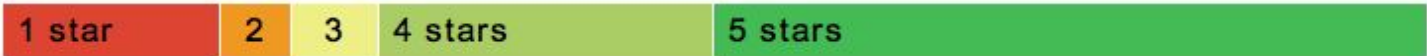
HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner

\$89 online, \$100 nearby ★★★★★ 377 reviews

September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

Summary - Based on 377 reviews

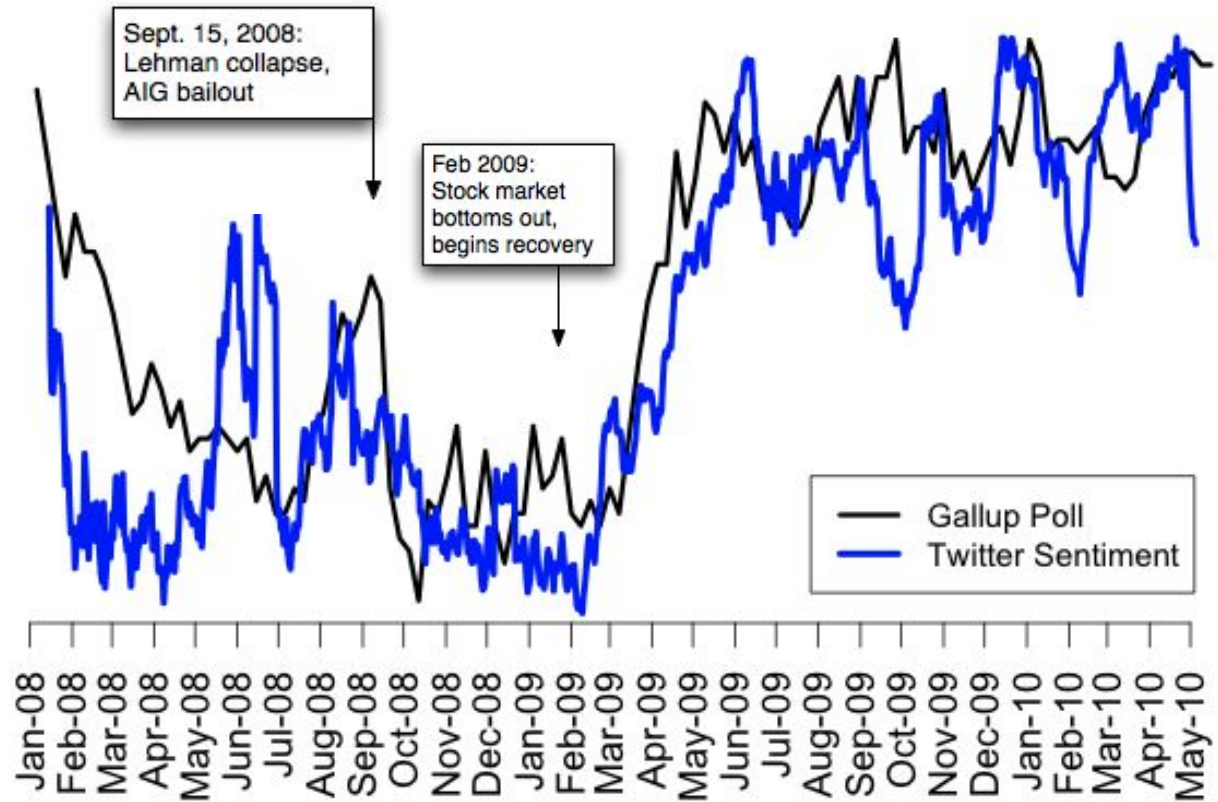


What people are saying

ease of use	<div><div></div><div></div></div>	"This was very easy to setup to four computers."
value	<div><div></div><div></div></div>	"Appreciate good quality at a fair price."
setup	<div><div></div><div></div></div>	"Overall pretty easy setup."
customer service	<div><div></div><div></div></div>	"I DO like honest tech support people."
size	<div><div></div><div></div></div>	"Pretty Paper weight."
mode	<div><div></div><div></div></div>	"Photos were fair on the high quality mode."
colors	<div><div></div><div></div></div>	"Full color prints came out with great quality."

Twitter sentiment versus Gallup Poll of Consumer Confidence

Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. 2010. [From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series](#) In ICWSM-2010



Target Sentiment on Twitter

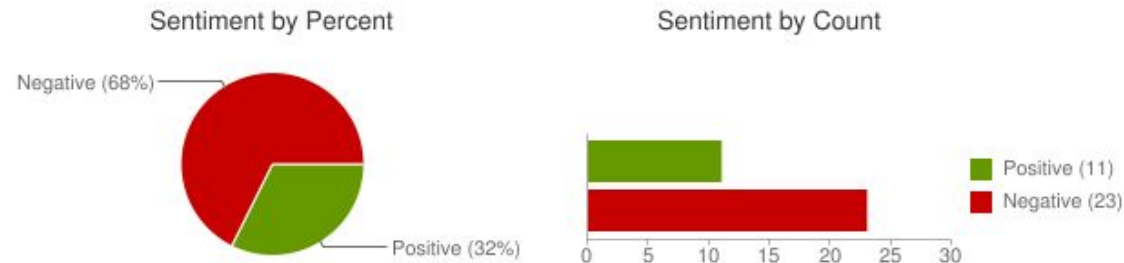
Type in a word and we'll highlight the good and the bad

"united airlines"

Search

[Save this search](#)

Sentiment analysis for "united airlines"



[iljacobson](#): OMG... Could @United airlines have worse customer service? W8g now 15 minutes on hold 4 questions about a flight 2DAY that need a human.
Posted 2 hours ago

[12345clumsy6789](#): I hate United Airlines Ceiling!!! Fukn impossible to get my conduit in this damn mess! ?
Posted 2 hours ago

[EMLandPRGbelgiu](#): EML/PRG fly with Q8 united airlines and 24seven to an exotic destination. <http://t.co/Z9QloAjF>
Posted 2 hours ago

[CountAdam](#): FANTASTIC customer service from United Airlines at XNA today. Is tweet more, but cell phones off now!
Posted 4 hours ago

Sentiment Analysis Has Many Other Names



Opinion
extraction



Sentiment
mining



Opinion mining



Subjectivity
analysis

Why sentiment analysis?



Movie: is this review positive or negative?



Products: what do people think about the new iPhone?



Public sentiment: how is consumer confidence? Is despair increasing?



Politics: what do people think about this candidate or issue?



Prediction: predict election outcomes or market trends from sentiment

Scherer Typology of Affective States

Emotion: brief organically synchronized ... evaluation of a major event

- *angry, sad, joyful, fearful, ashamed, proud, elated*

Mood: diffuse non-caused low-intensity long-duration change in subjective feeling

- *cheerful, gloomy, irritable, listless, depressed, buoyant*

Interpersonal stances: affective stance toward another person in a specific interaction

- *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*

Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons

- *liking, loving, hating, valuing, desiring*

Personality traits: stable personality dispositions and typical behavior tendencies

- *nervous, anxious, reckless, morose, hostile, jealous*

Scherer, Klaus R. 1984. Emotion as a Multicomponent Process: A model and some cross-cultural data. In *Review of Personality and Social Psych* 5: 37-63.

Sentiment Analysis

Sentiment analysis is the detection of **attitudes**

“enduring, affectively colored beliefs, dispositions towards objects or persons”

1. **Holder (source)** of attitude
2. **Target (aspect)** of attitude
3. **Type** of attitude

From a set of **types**

- *Like, love, hate, value, desire, etc.*

Or (more commonly) simple weighted **polarity**:

- *positive, negative, neutral, together with strength*

From a **Text** containing the attitude

- Sentence or entire document

Sentiment Analysis

Simplest task:

- Is the attitude of this text positive or negative?

More complex:

- Rank the attitude of this text from 1 to 5

Advanced:

- Detect the target, source, or complex attitude types

Precision, Recall, and F Measure

Confusion Matrix

	Gold positive	Gold negative
System positive	True positives	False positives
System negative	False negatives	True negatives

Precision and recall

Precision = $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$

Recall = $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$

	Gold positive	Gold negative
System positive	True positives	False positives
System negative	False negatives	True negatives

Accuracy

Accuracy =
$$\frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{false positives} + \text{true negatives} + \text{false negatives}}$$

	Gold positive	Gold negative
System positive	True positives	False positives
System negative	False negatives	True negatives

A combined measure: F-Measure

A combined measure that assesses the Precision / Recall tradeoff is F measure