

Introduzione alla Sentiment Analysis

Rachele Sprugnoli – rachele.sprugnoli@unicatt.it

Centro Interdisciplinare di Ricerche per la Computerizzazione
dei Segni dell'Espressione (CIRCSE)



UNIVERSITÀ
CATTOLICA
del Sacro Cuore

COS' È LA SENTIMENT ANALYSIS?

«Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes.»

Bing Liu, *"Sentiment Analysis and Opinion Mining"* Morgan & Claypool Publishers, 2012

Tanti nomi, tanti task:

- opinion extraction
- sentiment mining
- subjectivity analysis
- affect analysis
- emotion analysis
- review mining
- ...

TESTI OGGETTIVI *versus* SOGGETTIVI

TESTO OGGETTIVO

«La Pinacoteca di Brera è una galleria nazionale d'arte antica e moderna»

- Racconta dei fatti
- Verificabile
- Presentato con obiettività

TESTO SOGGETTIVO

«Entrare in Pinacoteca significa essere travolti dalla bellezza»

- Racconta opinioni, valutazioni, emozioni, giudizi, punti di vista
- Non verificabile
- Presentato senza obiettività

COSA SONO LE OPINIONI

- Sono formate da 5 elementi:



COSA SONO LE OPINIONI

- Dati non strutturati: **il linguaggio delle opinioni**

- **Prior Polarity:** la polarità delle singole parole considerate al di fuori di qualsiasi contesto

- *eccezionale* = POS
- *migliore* = POS
- *ottimo* = POS



visioneM
Milano, Italia
10 5



Recensito 2 settimane fa

Pizze napoletane eccezionale vale la lunga attesa

Il locale non accetta prenotazione quindi è molto probabile fare 50/60 minuti di attesa fuori dal locale in attesa di essere chiamati, ma l'attesa viene ripagata dalla soddisfazione di cenare nella migliore pizzeria di Milano. La pizza è napoletana doc, il menu è diviso tra pizze classiche e pizze regionali per le quali utilizzano ingredienti di altissima qualità. Il prezzo è ottimo rapportato alla qualità dei piatti.

Data della visita: febbraio 2020



Qualità/prezzo



Servizio



Cibo

Chiedi a visioneM in merito a Pizzium - Via Vigevano

Grazie, visioneM

Questa recensione rappresenta l'opinione personale di un viaggiatore di TripAdvisor e non di TripAdvisor LLC.

COSA SONO LE OPINIONI

- Dati non strutturati: **il linguaggio delle opinioni**

- **Semantic Orientation:** la misura della polarità e dell'intensità di un'opinione

- *lungo + attesa* = **NEG**
- *dolce + attesa* = **POS**
- *lungo + tradizione* = **POS**
- *alto + qualità* = **POS**
- *basso + qualità* = **NEG**
- *alto + prezzo* = **NEG**



visioneM
Milano, Italia
10 5



Recensito 2 settimane fa

Pizze napoletane eccezionale vale la lunga attesa

Il locale non accetta prenotazione quindi è molto probabile fare 50/60 minuti di attesa fuori dal locale in attesa di essere chiamati, ma l'attesa viene ripagata dalla soddisfazione di cenare nella migliore pizzeria di Milano. La pizza è napoletana doc, il menu è diviso tra pizze classiche e pizze regionali per le quali utilizzano ingredienti di altissima qualità. Il prezzo è ottimo rapportato alla qualità dei piatti.

Data della visita: febbraio 2020



Qualità/prezzo



Servizio



Cibo

Chiedi a visioneM in merito a Pizzium - Via Vigevano

Grazie, visioneM

Questa recensione rappresenta l'opinione personale di un viaggiatore di TripAdvisor e non di TripAdvisor LLC.

COSA SONO LE OPINIONI

- Dati non strutturati: il linguaggio delle opinioni

- **Negazioni e Sentiment shifters:** elementi e strutture linguistiche capaci di modificare la *prior polarity*

- **NON** accetta prenotazioni...



visioneM
Milano, Italia
10 5



Recensito 2 settimane fa

Pizze napoletane eccezionale vale la lunga attesa

Il locale non accetta prenotazione quindi è molto probabile fare 50/60 minuti di attesa fuori dal locale in attesa di essere chiamati, ma l'attesa viene ripagata dalla soddisfazione di cenare nella migliore pizzeria di Milano. La pizza è napoletana doc, il menu è diviso tra pizze classiche e pizze regionali per le quali utilizzano ingredienti di altissima qualità. Il prezzo è ottimo rapportato alla qualità dei piatti.

Data della visita: febbraio 2020



Qualità/prezzo



Servizio



Cibo

Chiedi a visioneM in merito a Pizzium - Via Vigevano

Grazie, visioneM

Questa recensione rappresenta l'opinione personale di un viaggiatore di TripAdvisor e non di TripAdvisor LLC.

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - lessico con un sentiment che cambia **in base al contesto**
 - «abbiamo avuto un **caldo** benvenuto»
 - «abbiamo avuto una birra **calda**»
 - «la malattia può avere un decorso **acuto**»
 - «famoso per il suo spirito **acuto** e intelligente»

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - stesse parole, sensi diversi (**Word Sense Disambiguation**) - Sentix lexicon

<i>PAROLA</i>	<i>PoS</i>	<i>#senso</i>	<i>POS_score</i>	<i>NEG_score</i>	<i>polarity</i>	<i>intensity</i>
<i>acuto</i>	<i>a</i>	<i>00044760</i>	<i>0</i>	<i>0.5</i>	<i>-1.0</i>	<i>0.5</i>
<i>acuto</i>	<i>a</i>	<i>00650900</i>	<i>0.625</i>	<i>0.125</i>	<i>0.75</i>	<i>0.64</i>
<i>acuto</i>	<i>a</i>	<i>00803275</i>	<i>0.5</i>	<i>0.25</i>	<i>0.41</i>	<i>0.56</i>
<i>acuto</i>	<i>a</i>	<i>01744515</i>	<i>0.125</i>	<i>0</i>	<i>1.0</i>	<i>0.125</i>

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - stesse parole, sensi diversi, sentiment diversi (**Word Sense Disambiguation**) - Sentix lexicon

PAROLA	PoS	#senso	POS_score	NEG_score	polarity	intensity
acuto	a	00044760	0	0.5	-1.0	0.5
acuto	a	00650900	0.625	0.125	0.75	0.64
acuto	a	00803275	0.5	0.25	0.41	0.56
acuto	a	01744515	0.125	0	1.0	0.125

di malattia, che insorge con una certa violenza e ha un decorso rapido ed intenso

di facoltà intellettuale o di una sua manifestazione, che sa comprendere e valutare con prontezza le situazioni

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:

- opinioni e sentiment **impliciti**

*«Ho comprato il materasso una settimana fa e adesso c'è un avvallamento» versus «Il materasso si è **rovinato** dopo poco»*

*«La batteria degli iPhone dura di più di quella dei Samsung» versus «La batteria degli iPhone è **migliore** di quella dei Samsung»*

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:

- **ironia, sarcasmo, emoticon**

«veloce come un bradipo» = NEG

«che telefono fantastico! Si è rotto dopo 2 giorni!» = NEG

«ore 7:30 batteria al 100% - ore 21 batteria 20% 😞» = NEG

«🔥🔥🔥» = POS

IL LINGUAGGIO DELLE OPINIONI

- Il linguaggio che usiamo per esprimere le nostre valutazioni soggettive è estremamente complesso, formato da molte componenti:
 - **conoscenza del mondo**

« *Ha il viso come una madonna di Botticelli!* » = POS

« *Sembra un quadro di Picasso!* » = NEG



TASK DI SENTIMENT ANALYSIS

- Non esiste una soluzione unica per la Sentiment Analysis
 - la Sentiment Analysis non è un singolo task
 - la Sentiment Analysis non è un singolo algoritmo
- Breve descrizione di task
 - Subjectivity detection
 - Document sentiment classification
 - Sentence-level sentiment classification
 - Aspect-based sentiment analysis
- Altri task strettamente correlati alla Sentiment Analysis
 - Stance detection
 - Irony detection
 - Emotion analysis
 - Hate speech/cyberbullying/misogyny detection

1) SUBJECTIVITY DETECTION

- Determinare se un testo contiene o meno delle valutazioni soggettive

Oggettivo

pisa93

domenica 16
febbraio 2014

un film che ha ben poco di monumentale ★★☆☆☆

Durante la Seconda Guerra Mondiale, un manipolo di uomini chiamati Monuments Men viene incaricato di recuperare la maggior parte delle opere d'arte trafugate dai Tedeschi. Tratto da una storia vera. Giunto alla sua quinta regia George Clooney vorrebbe stupire con un film alla Spielberg, ma non riesce ad essere incisivo. Si ispira alle ambientazioni di "Salvate il soldato Ryan" ed all'introspezione di "La sottile linea rossa", ma tutto ciò che ottiene è un involucro vuoto, senz'anima. Monuments Men è un film che fatica a trovare una sua identità, alternando commedia e dramma con la stessa semplicità con cui una ragazza si cambia d'abito. Manca proprio di quella profondità di cui un film del genere dovrebbe essere permeato. **[+]**

[+] lascia un commento a pisa93 »



Soggettivo

2) DOCUMENT SENTIMENT CLASSIFICATION

- Classificare un intero documento in base al sentiment:
 - il documento esprime un'opinione su una singola entità da parte di un singolo «holder»
 - scelta delle classi: categorie o scala di valori numerici?

NEGATIVE? VERY NEGATIVE? -2?

pisa93

un film che ha ben poco di monumentale ★★☆☆☆

Durante la Seconda Guerra Mondiale, un manipolo di uomini chiamati Monuments Men viene incaricato di recuperare la maggior parte delle opere d'arte trafugate dai Tedeschi. Tratto da una storia vera. Giunto alla sua quinta regia George Clooney vorrebbe stupire con un film alla Spielberg, ma non riesce ad essere incisivo. Si ispira alle ambientazioni di "Salvate il soldato Ryan" ed all'introspezione di "La sottile linea rossa", ma tutto ciò che ottiene è un involucro vuoto, senz'anima. Monumets Men è un film che fatica a trovare una sua identità, alternando commedia e dramma con la stessa semplicità con cui una ragazza si cambia d'abito. Manca proprio di quella profondità di cui un film del genere dovrebbe essere permeato. **[+]**

[+] lascia un commento a pisa93 »

3) SENTENCE-LEVEL SENTIMENT CLASSIFICATION

- Classificare una frase in base al sentiment:
 - un documento può contenere opinioni anche opposte
 - vari livelli di granularità possibili: paragrafi, strofe...

no_data

l'arte di essere americani ★★★★★

L'idea di fondo non è male: l'arte come patrimonio, come radice della nostra identità, del nostro essere; la cultura come punto di partenza per ricordarci che siamo umani e, da lì, per ricostruire un continente distrutto dalla guerra. Questa idea merita rispetto e bisogna riconoscere come sia effettivamente vera. Tuttavia, il film in alcuni punti sembra perdere di vista questa linea: purtroppo spesso sembra più impegnato a dimostrare quanto sono stati bravi e meticolosi gli Americani e quanto di buono abbiano fatto in Europa. La realtà è, come spesso accade, molto più complessa! Gli episodi in cui le forze alleate hanno distrutto opere d'arte di grande valore durante il secondo conflitto mondiale sono noti e il film forse avrebbe dovuto fare un pò di autocritica in più. [+]

[+] lascia un commento a no_data »

martedì 29 luglio 2014

Sì No
0% 0%
d'accordo?

POSITIVE

NEGATIVE

4) ASPECT-BASED SENTIMENT ANALYSIS

- Individuare le singole caratteristiche che rappresentano i punti di forza e/o le debolezze discusse nell'opinione
 - chiamato anche *feature-based sentiment analysis*
 - 2 fasi principali:
 - 1) estrazione delle caratteristiche
 - 2) classificazione del sentiment per ciascuna caratteristica



fred02



Carino, da provare.

Recensito in Italia il 2 novembre 2018

Style: Echo Dot | Colore: Antracite | Configurazione: Solo dispositivo

Ingolosito dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.

7 persone l'hanno trovato utile



4) ASPECT-BASED SENTIMENT ANALYSIS

- 2 fasi principali:
 - 1) **estrazione** delle caratteristiche
 - 2) classificazione del sentiment per ciascuna caratteristica

CONFIGURAZIONE

PREZZO

ESTETICA

The image shows a screenshot of an Amazon product review for the Echo Dot. The review is by user 'fred02' and is a 4-star rating. The text of the review is: 'Ingolosito dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.' There are three arrows pointing to specific parts of the review: one from 'CONFIGURAZIONE' pointing to 'configurarlo', one from 'PREZZO' pointing to 'prezzo scontato', and one from 'ESTETICA' pointing to 'Carino'. Below the review text, it says '7 persone l'hanno trovato utile'.

fred02

★★★★☆ Carino, da provare.

Recensito in Italia il 2 novembre 2018

Style: Echo Dot | Colore: Antropocene | Configurazione: Solo dispositivo | **Acquisto verificato**

Ingolosito dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.

7 persone l'hanno trovato utile

4) ASPECT-BASED SENTIMENT ANALYSIS

- 2 fasi principali:
 - 1) estrazione delle caratteristiche
 - 2) **classificazione** del sentiment per ciascuna caratteristica

CONFIGURAZIONE=NEG

PREZZO=POS

ESTETICA=POS

The image shows a product review for the Echo Dot. The reviewer is 'fred02', who gave it a 4-star rating. The review text is: 'Ingoloso dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.' Annotations include: a red arrow pointing to 'Ingoloso' with the label 'CONFIGURAZIONE=NEG'; a green arrow pointing to 'prezzo scontato' with the label 'PREZZO=POS'; and another green arrow pointing to 'Carino' with the label 'ESTETICA=POS'. The review also includes metadata: 'Recensito in Italia il 2 novembre 2018', 'Style: Echo Dot | Colore: Antracite | Configurazione: Solo dispositivo | Acquisto verificato', and '7 persone l'hanno trovato utile'.

fred02

★★★★☆ Carino, da provare.

Recensito in Italia il 2 novembre 2018

Style: Echo Dot | Colore: Antracite | Configurazione: Solo dispositivo | Acquisto verificato

Ingoloso dal prezzo scontato ho acquistato Alexa e mi è puntualmente arrivato il 30 ottobre. Ho avuto difficoltà a configurarlo perché l'app. Alexa, sul mio OP5 non permetteva la configurazione. Questo è un problema che sembra affliggere anche i dispositivi Xiaomi. L'ho poi configurato con un Ipad. Carino, l'ho paragonato al Google Home mini che già possiedo, e sinceramente preferisco quest'ultimo. Comunque prodotto valido e divertente, al prezzo di lancio è sicuramente da provare.

7 persone l'hanno trovato utile

5) STANCE DETECTION

- Individuare il punto di vista generale (*posizione*) di una persona verso una cosa, un'idea, un evento
 - a favore o contro*
 - applicazioni in ambito politico: e.g., referendum 2016



A FAVORE

CONTRO

6) IRONY DETECTION

- Individuare contenuti ironici
 - l'ironia trasmette l'opposto del significato letterale
 - invertitore di polarità
 - satira, sarcasmo, parodia: è possibile distinguerli automaticamente?



7) EMOTION ANALYSIS

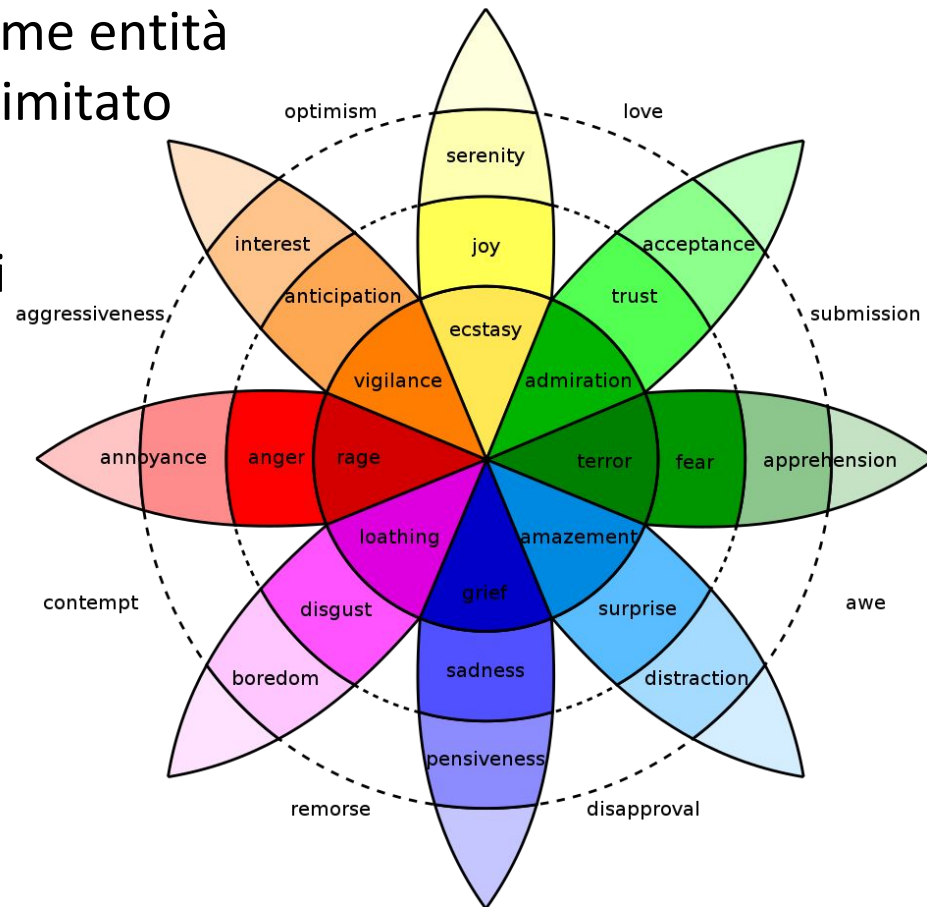
- 2 approcci

1) identificare le emozioni viste come entità innate ed universali di numero limitato

- la ruota di Plutchik:

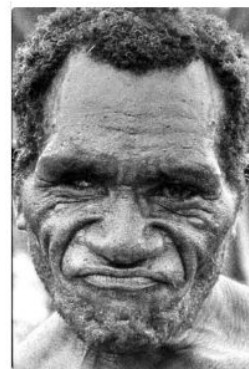
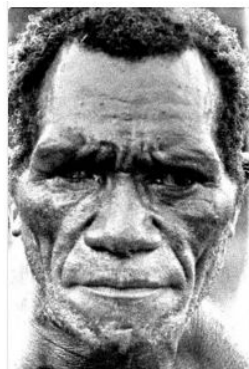
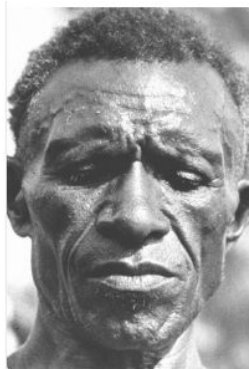
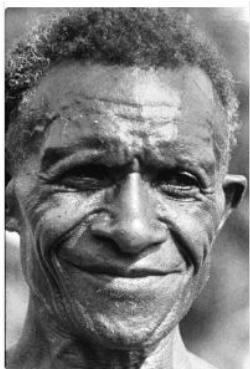
4 coppie di emozioni primarie che si combinano

- *gioia* contro *tristezza*
- *rabbia* contro *paura*
- *fiducia* contro *disgusto*
- *sorpresa* contro *trepidazione*



7) EMOTION ANALYSIS

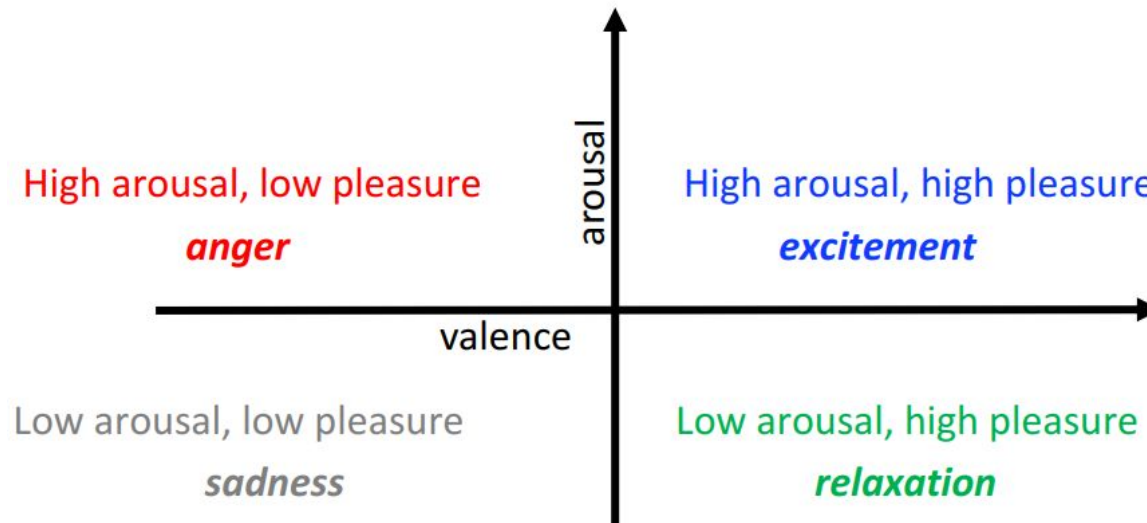
- 2 approcci
 - 1) identificare le emozioni viste come entità innate ed universali di numero limitato
 - gli studi di Ekman sull'universalità delle emozioni: rabbia, disgusto, tristezza, gioia, paura, sorpresa



<https://www.paulekman.com/>

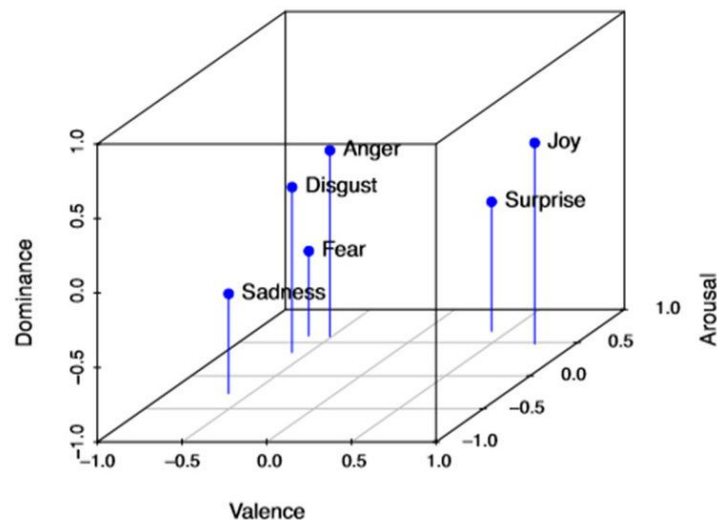
7) EMOTION ANALYSIS

- 2 approcci
 - 1) identificare le emozioni viste come un numero limitato di unità innate ed universali
 - 2) identificare varie dimensioni delle emozioni
 - Valence: piacevolezza dell'emozione
 - Arousal: intensità dell'emozione
 - Dominance: grado di controllo esercitato dall'emozione



7) EMOTION ANALYSIS

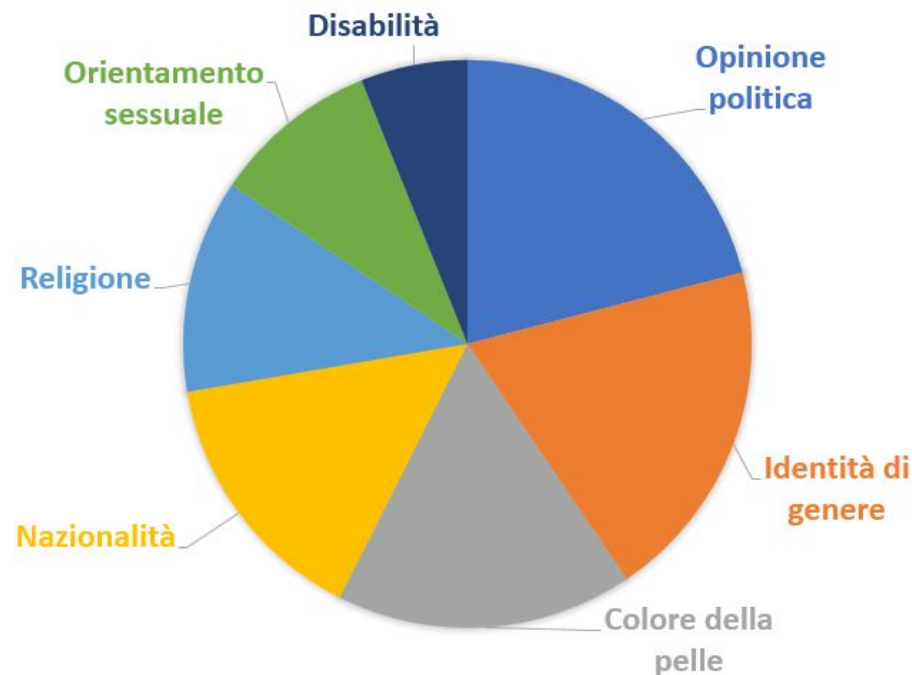
- 2 approcci
 - 1) identificare le emozioni viste come un numero limitato di unità innate ed universali
 - 2) identificare varie dimensioni delle emozioni
 - Valence: piacevolezza dell'emozione
 - Arousal: intensità dell'emozione
 - Dominance: grado di controllo esercitato dall'emozione



8) HATE SPEECH DETECTION

- Hate Speech: “*Qualsiasi **forma d’espressione** che fomenti l’**odio** contro una persona o un gruppo di persone a causa di **caratteristiche che le accomunano.**” (Mendel, 2012 «Does international law provide for consistent rules on hate speech? »)*

eMORE, Monitoring and Reporting Online Hate Speech in Europe, 2017



8) HATE SPEECH DETECTION



- Islamofobia

- Cyberbullismo

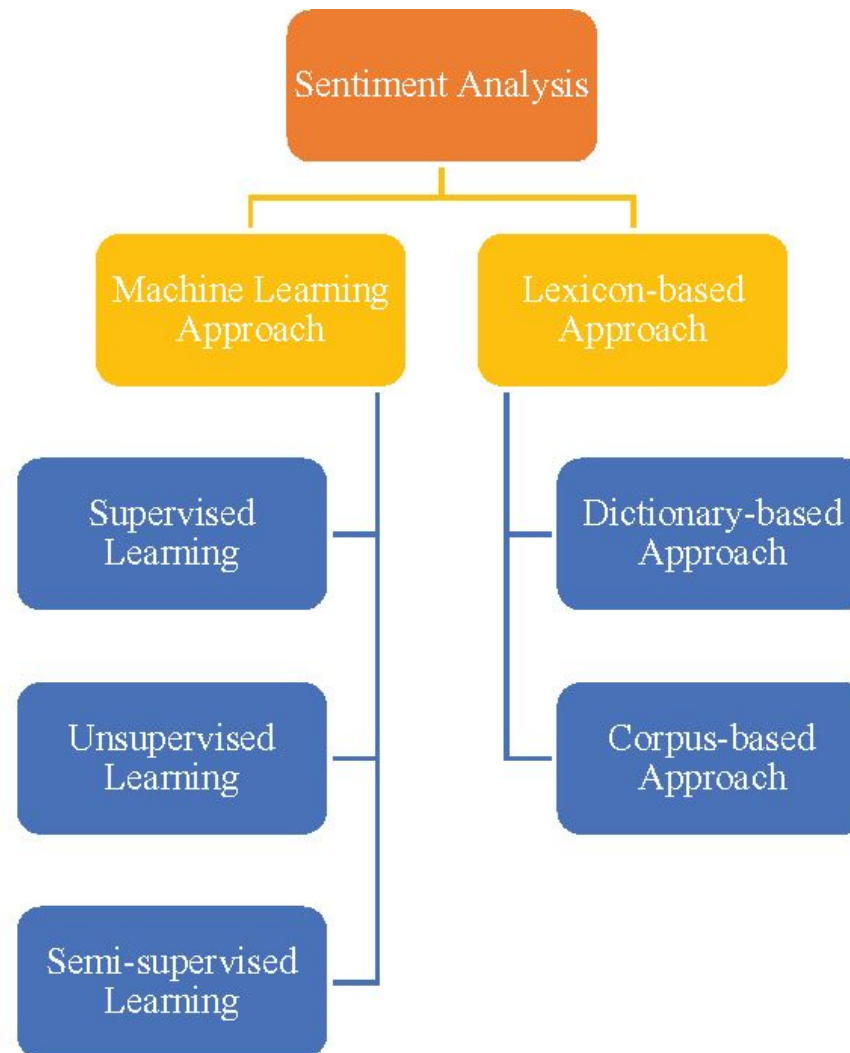


la #raggi sì che è una patata bollente altro che quella figa di legno della #boschi o quella cassapanca della #boldrini 😂😂
#Libero @vfeltri

- Misoginia

10/02/17, 11:16

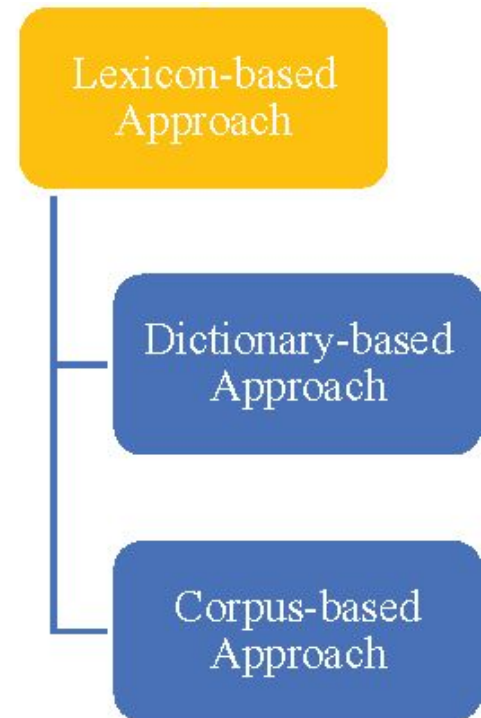
COME SI RISOLVE UN TASK DI SENTIMENT ANALYSIS



APPROCCI BASATI SUL LESSICO

- Basati sull'intuizione che la polarità di testo può essere ottenuta sulla base della polarità delle parole che lo compongono
- Si basano su liste di token, espressioni o lemmi («*sentiment/polarity lexicon*»): ad ogni token o lemma sono associate delle categorie o dei valori numerici che ne quantificano la polarità
Esempi:

- *negative, neutral, positive*
- *very negative, negative, neutral, positive, very positive*
- *-1, 0, +1*
- *-2, -1, 0, +1, +2*
- *valori decimali tra 0 e 1*



COME SI CREANO LESSICI DI SENTIMENT

- **DICTIONARY-BASED:** si usano una lista di termini con associato il sentiment e dei dizionari in formato digitale in cui si cercano sinonimi e contrari di quei termini

- *paura* = **NEG**

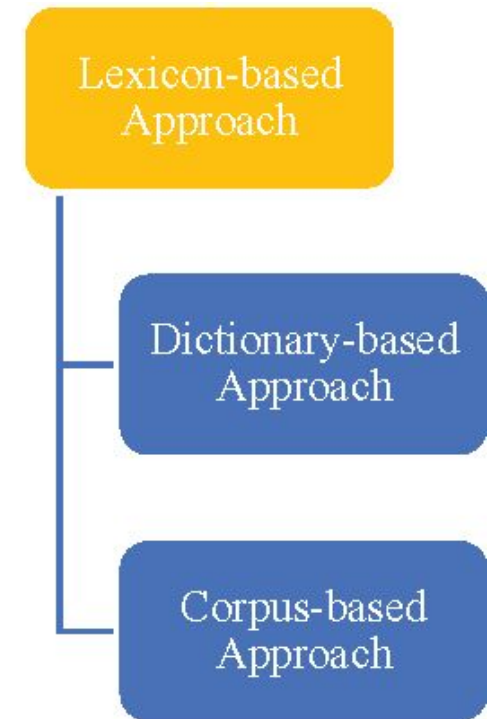
SINONIMI = *spavento*, *terrore* = **NEG**

CONTRARI = *coraggio*, *audacia* = **POS**

- **CORPUS-BASED:** si usano una lista di termini con associato il sentiment e delle collezioni di testi. Altri termini vengono aggiunti alla lista cercando contesti comuni o strutture sintattiche particolari

La macchina è bella E spaziosa

La macchina è bella MA difficile da guidare



QUANTO BENE FUNZIONANO I SISTEMI?

SISTEMI MACHINE LEARNING SULL'ITALIANO

- **Subjectivity Detection:** 0,74 F-score (2016)
- **Sentiment Classification** (Twitter): 0,66 F-score (2016)
- **Aspect-based Sentiment Analysis:** 0,81 F-score (2018)
- **Irony Detection:** 0,73 F-score (2018)
- **Irony Classification:** 0,52 F-score (2018)
- **Misogyny Detection:** 0,84 Accuracy (2018)
- **Misogyny Classification:** 0,41 Accuracy (2018)
- **Hate Speech Detection** (Twitter): 0,80 F-score (2018)
- **Hate Speech Detection** (Facebook): 0,83 F-score (2018)

Performance dei migliori sistemi di EVALITA 2016 e 2018

APPLICAZIONE: COVID-19

- Progetto congiunto di Expert System (azienda di tecnologie semantiche e intelligenza artificiale) e Sociometrica (azienda che si occupa di analisi economica e sociale)
- Monitoraggio sentiment ed emozioni dai social media
- Classificazione in 80 emozioni
- <https://expertsystem.com/it/covid-19-insights-feelings/>

APPLICAZIONE: COVID-19

26 maggio 2020

Un paese diviso dalle offese reciproche

22 maggio 2020

Più tranquillità, ma senza ottimismo

19 maggio 2020

Sentiment meno negativo, ma ancora avvolto nella tristezza

12 maggio 2020

L'economia abbatte il sentiment degli Italiani

8 maggio 2020

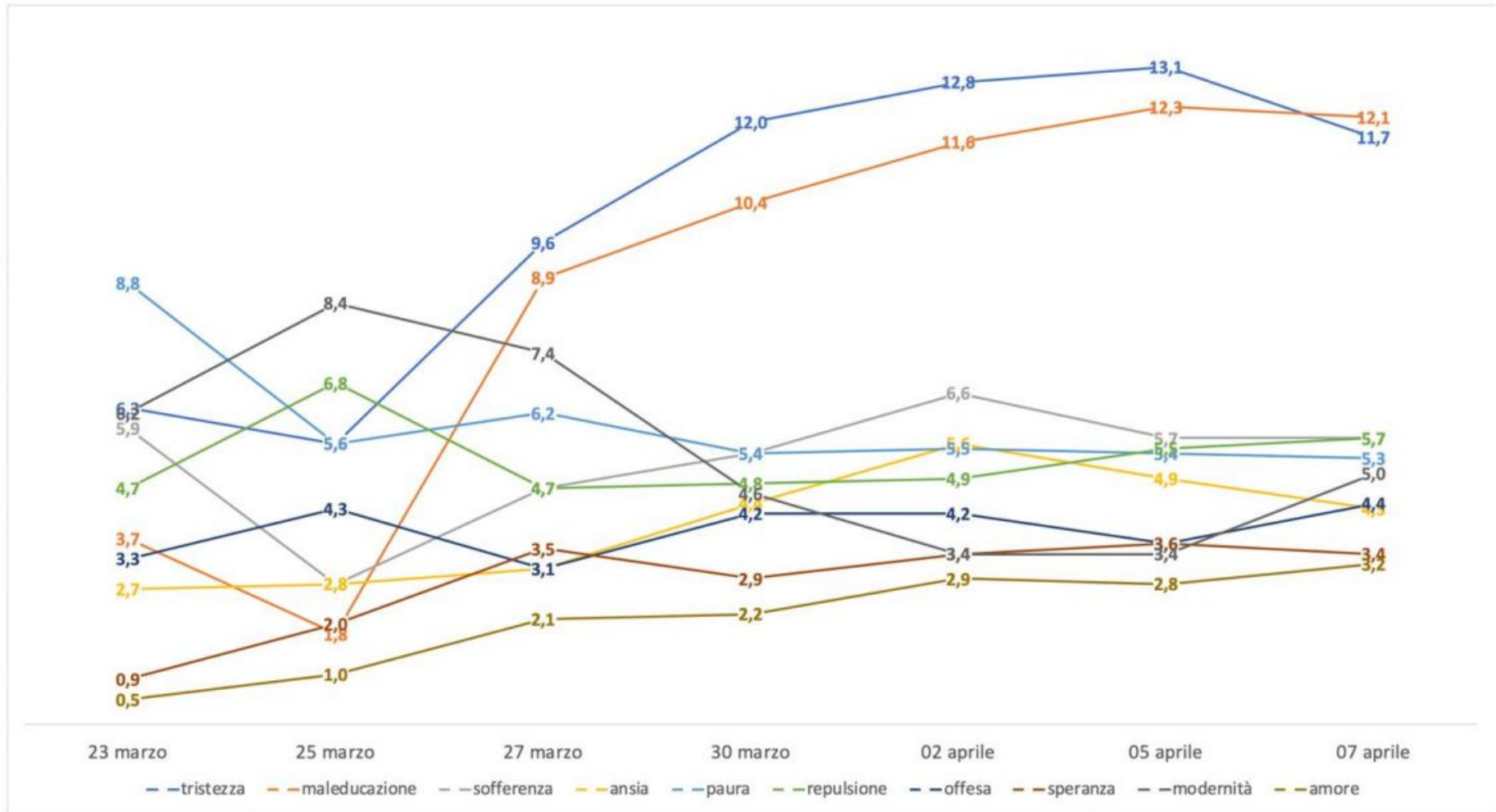
Cresce l'irritazione per le condizioni della ripresa

5 maggio 2020

Una finestra di ottimismo

APPLICAZIONE: COVID-19

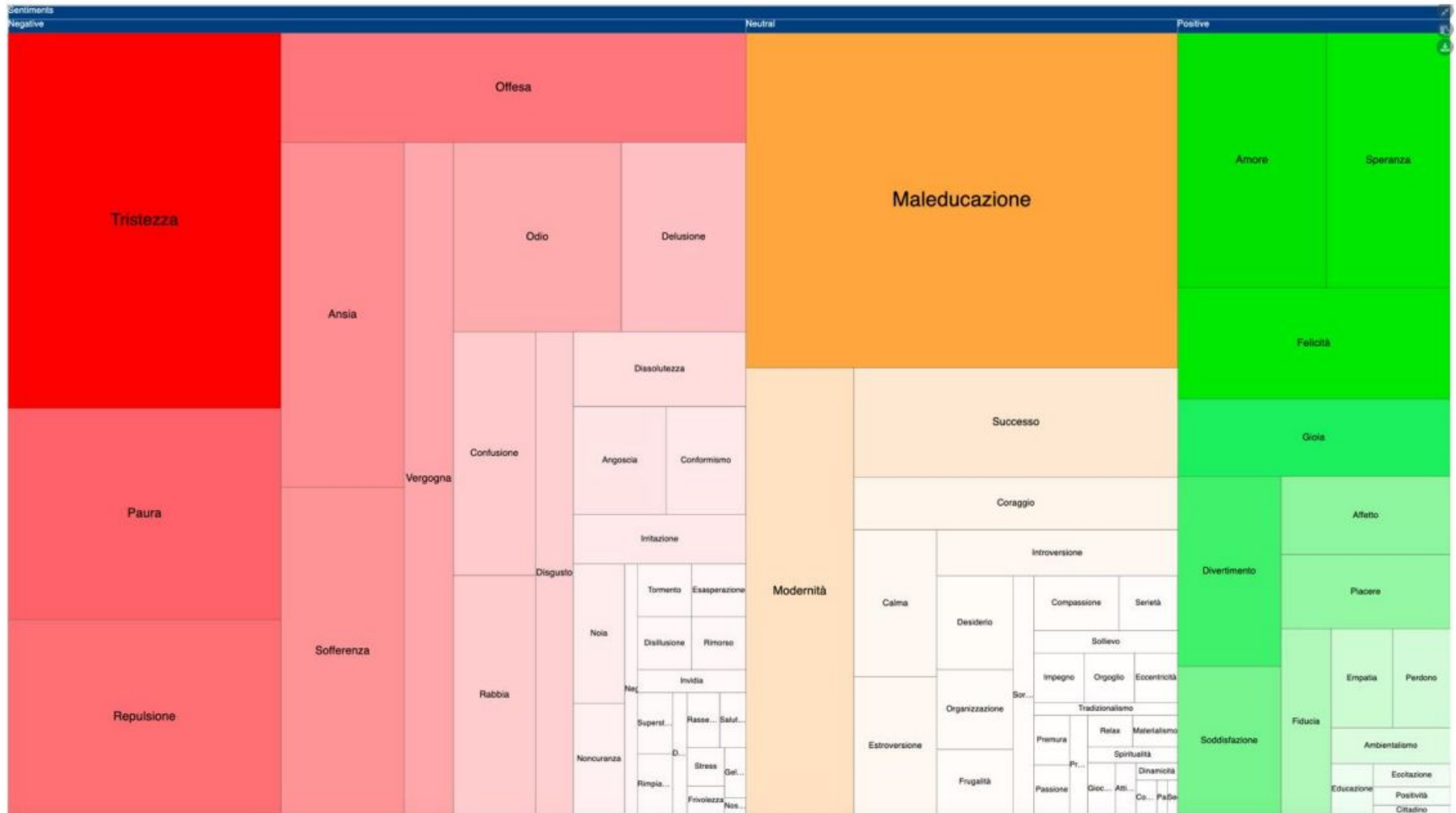
Andamento delle emozioni principali degli Italiani rispetto al Covid-19 (23 marzo – 7 aprile)



Fonte: Expert System, Sociometrica, 7 aprile 2020

APPLICAZIONE: COVID-19

Mappa delle emozioni principali degli Italiani rispetto al Covid-19



Fonte: Expert System, Sociometrica, 26 maggio 2020

APPLICAZIONE: Mappe dell'Intolleranza

- Progetto in collaborazione tra Vox Diritti (osservatorio italiano sui diritti) e l'Università Aldo Moro di Bari (Dipartimento di Informatica)
- Ogni anno per quattro anni, 4 mesi considerati, 200,000 tweet analizzati
- Hate-speech detection
- 6 argomenti: Omofobia, Razzismo, Antisemitismo, Sessismo, Disabilità, Islamofobia
- Ultima analisi (2019):
<http://www.voxdiritti.it/la-nuova-mappa-dellintolleranza-4/>

APPLICAZIONE: Mappe dell'Intolleranza



XENOFOBIA, LA VERA PIAGA ITALIANA

CONTRO I MIGRANTI SI MOLTIPLICANO I TWEET INTOLLERANTI E DISCRIMINATORI. ODIANO DUE ITALIANI SU TRE TRA COLORO CHE TWITTANO

Un boato di odio, che non si placa, ma aumenta con il tam tam della politica. I migranti sono al primo posto tra le categorie di persone più colpite dallo hate speech, in netta crescita rispetto all'anno passato. E tra le città più intolleranti, spicca Milano.

Il grafico mostra la quantità di tweet raccolti per ogni singolo cluster:

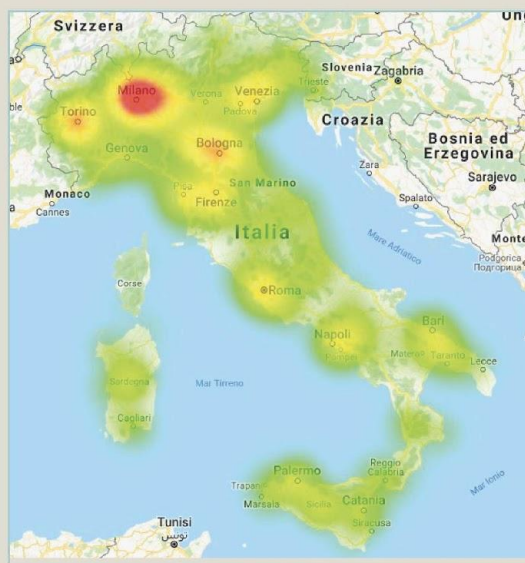
- ANTISEMITISMO
- RAZZISMO
- DISABILITÀ
- MISOGINIA
- ISLAMOFOBIA
- OMOFOBIA

Tweet sui migranti
74.451*

TWEET
215.377
TOTALI
ESTRATTI**

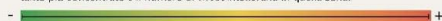
* Totale dei tweet rilevati, contenenti le parole sensibili, relative al cluster Razzismo

** Totale dei tweet estratti nei periodi di rilevazione con valenza sia positiva che negativa.



COME SI LEGGE LA MAPPA TERMOGRAFICA

Le mappe sono state costruite grazie al software Open StreetMap: sintetizzano la diffusione di tweet intolleranti attraverso delle termografie. Quanto più caldo (vicino al rosso) è il colore, tanto più concentrato è il numero di tweet intolleranti in quella zona.



I numeri del fenomeno

5,1 milioni

CITTADINI STRANIERI
presenti in Italia
nel 2017

Circa l'**8,5%**
della popolazione
italiana

CONTRO
7,9 MILIONI della Francia
6,0 MILIONI della Spagna
12,2 MILIONI della Germania

Numero di migranti
nel mondo nel **2017**
258 MLN

RISPETTO ALLA POPOLAZIONE **8,5%**



ISCRITTI ALLA ANAGRAFE

I RISULTATI

2019



LE PAROLE INTOLLERANTI

Nella scelta delle parole da mappare, abbiamo lavorato sui termini e le offese più ricorrenti sui social, evidenziati anche nelle ricerche scientifiche, che si sono occupate di studiare i meccanismi implicati nell'atteggiamento discriminatorio.

Terrone • zingaro • negro • muso giallo
rumeno • albanese • muso da scimmia
bangla • mangia rane • kebabbaro • cruccio
rabbino • sbarcato • mangia banane

NAZIONALITÀ: **1.190.491** Romania
415.531 Marocco
290.631 Cina
237.047 Ucraina
187.359 Filippine

Casi di discriminazione in ITALIA
nel 2016 **69,4%**
nel 2017 **82,9%**

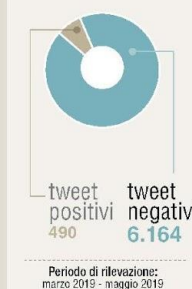
Nel 2018 in Italia



EPISODI
di intolleranza
e razzismo verso gli
immigrati

23,4% pensa che gli immigrati in Italia siano troppi
35,6% ha paura di subire REATI
50,9% ha problemi economici ed è insoddisfatto

IL FATTORE AGGRESSIVITÀ



Il grafico mostra la quantità di messaggi aggressivi, rilevati secondo un'ulteriore metodica di analisi: si è proceduto a estrarre i gruppi di parole più ricorrenti nei messaggi intolleranti, al di là di insulti e parole offensive. Il software è stato infatti "istruito" per estrarre i tweet più aggressivi, evidenziandone il livello di virulenza: la valutazione è stata orientata dalle categorie utilizzate dalla scala MQAS (Modified Overt Aggression Scale). Tale metodica, in questa sua prima forma sperimentale, si è dimostrata utile per meglio comprendere non solo la negatività, gli atteggiamenti intolleranti e discriminatori, ma anche l'orientamento aggressivo dei messaggi.

xenofobia

VOX



UNIVERSITÀ
DEGLI STUDI
DI MILANO



SAPIENZA
UNIVERSITÀ DI ROMA



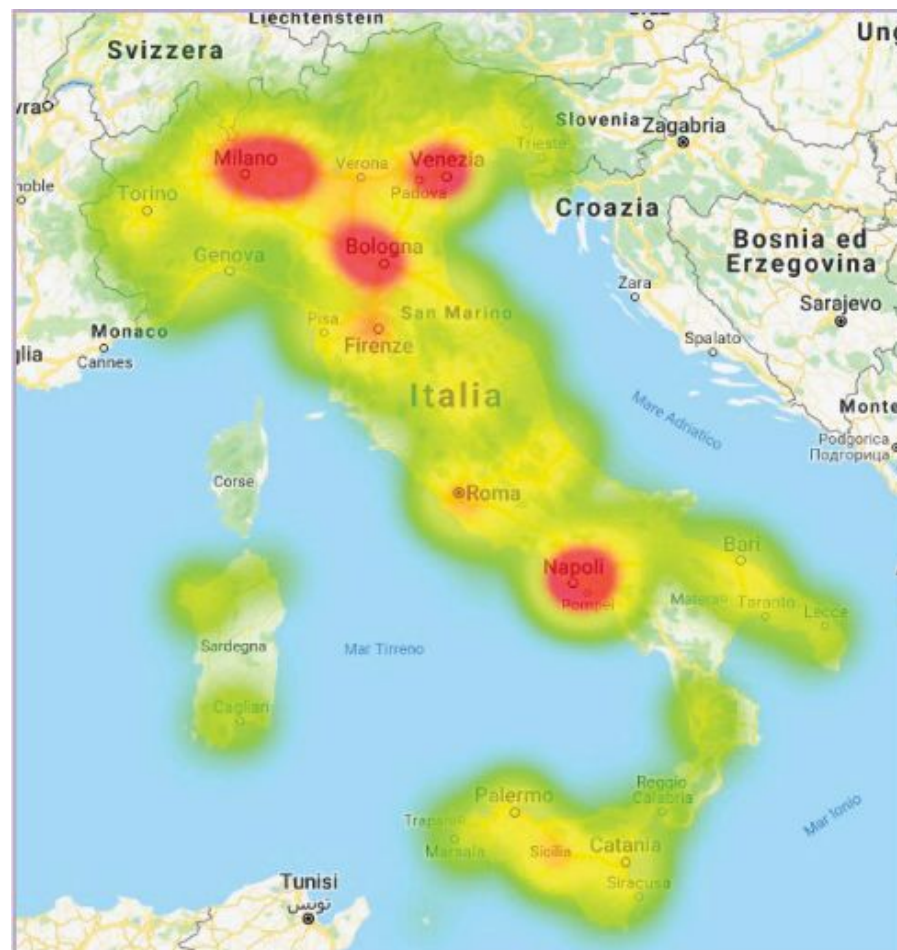
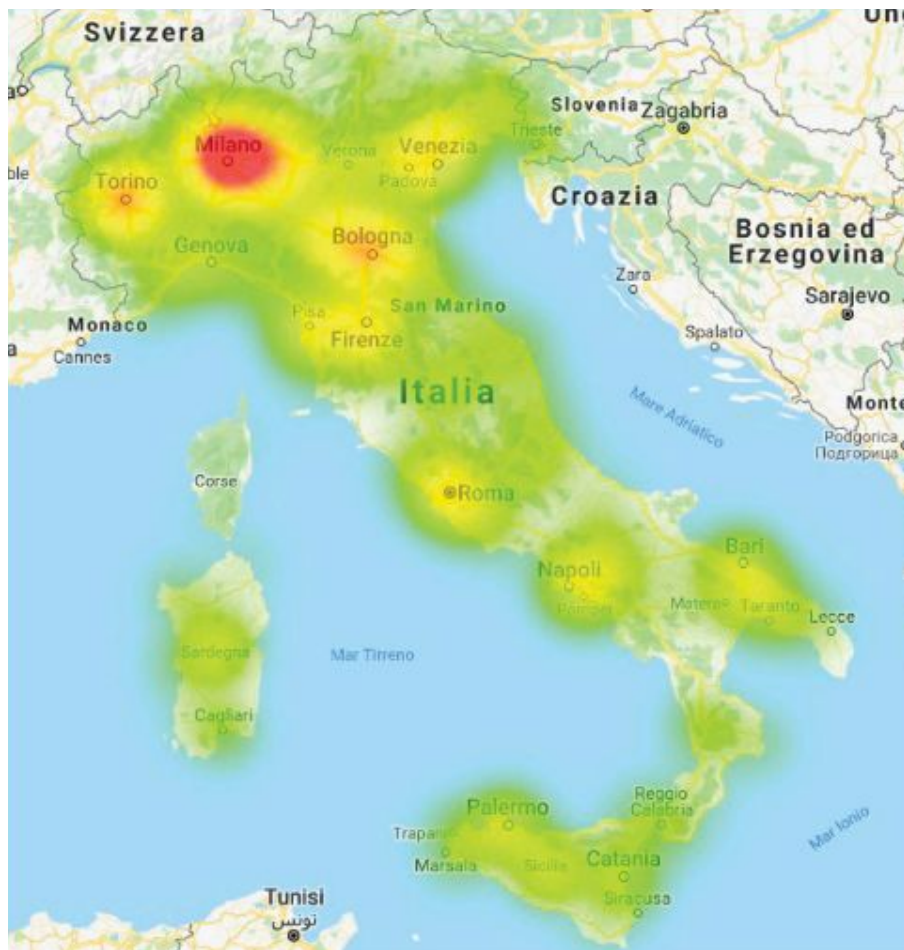
UNIVERSITÀ
DEGLI STUDI DI BARI
ALDO MORO

In collaborazione con
ISIME

APPLICAZIONE: Mappe dell'Intolleranza

XENOFOBIA

OMOFOBIA



APPLICAZIONE: The Emotions of London

- Progetto dello Stanford Literary Lab
- Analisi di 15.000 paragrafi estratti da 5000 romanzi inglesi ambientati a Londra e pubblicati tra il 1700 e il 1900
- NER + Emotion Analysis sia manuale (crowdsourcing) che automatica
- Fear versus Happiness
- PDF: <https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf>
- PDF (IT):
<http://www.fedoabooks.unina.it/index.php/fedoapress/catalog/view/104/90/596-2>

APPLICAZIONE: The Emotions of London

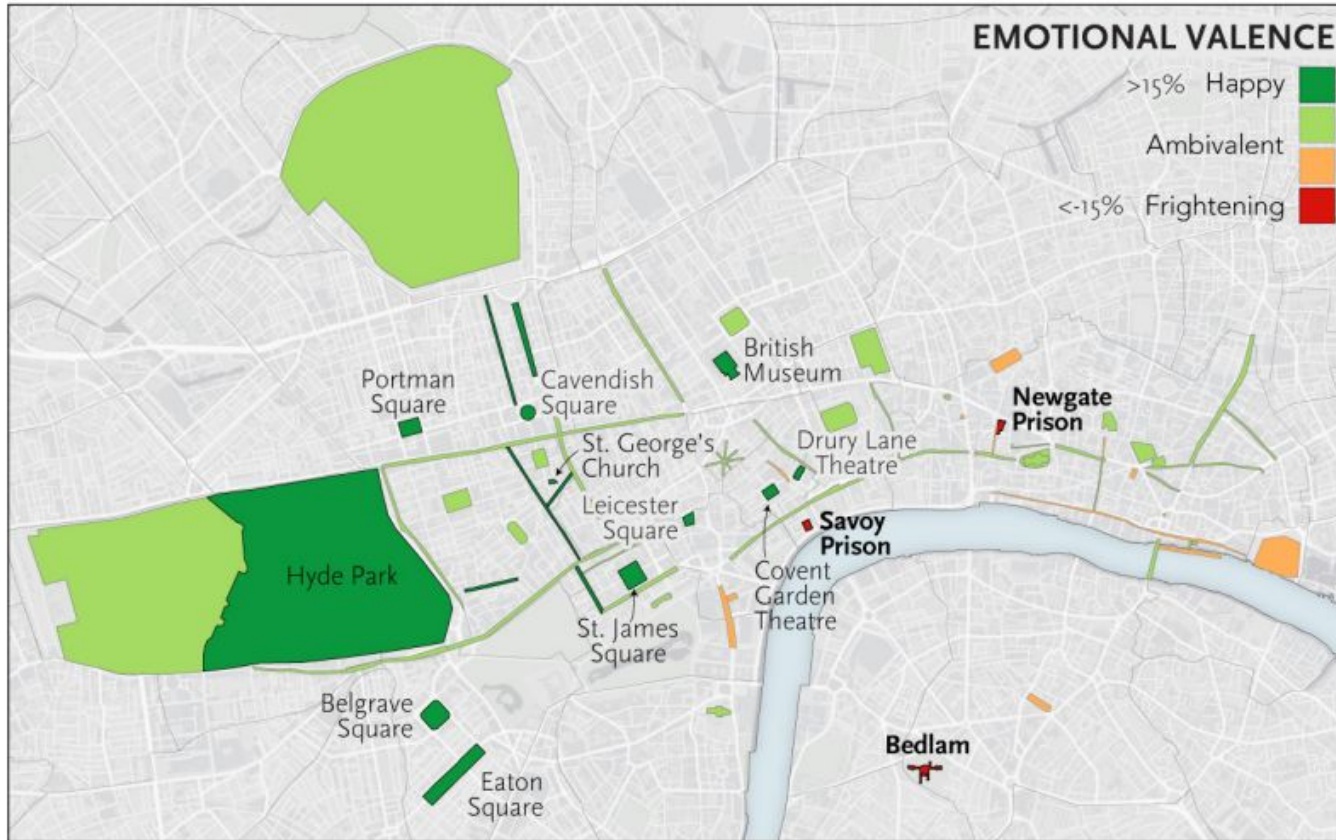


Figure 5.1 The emotions of London, 1700-1900

In this image, green is particularly prevalent in squares (the term that was also the most distinctive of the West End's lexicon), whereas passages where fear dominates are most often located in spaces of coercion and internment.

ESERCITAZIONE - 1

1. Copiare i due file nuovi [doAnalysis-new.py](#) e [RossoMalpelo.csv](#) nella cartella [lexicon-sentiment-analysis](#): sostituire col Python esistente
2. Aprire la cartella [lexicon-sentiment-analysis](#) e controllarne il contenuto
3. Aprire i file .py e .tsv con un editor di testo e i file .csv con un foglio di calcolo
4. Aprire il Terminale e usare il comando `cd` per entrare nella cartella [lexicon-sentiment-analysis](#)
5. Digitare il seguente comando poi premere il tasto Invio (per Windows)

```
set PYTHONIOENCODING=utf8
```

6. Digitare il seguente comando poi premere il tasto Invio

```
python doAnalysis-new.py > hr-out.txt
```

ESERCITAZIONE - 2

7. Aprire il file hr-out.txt con un editor di testo: togliere le prime righe e le intestazioni “TOP...”, rimuovere id= / score= / clean= / orig= / pubdate= / id= / score=
8. Copiare il contenuto così modificato in un foglio di calcolo
9. Selezionare e copiare su un altro foglio le colonne B (score) ed E (pubdate)
10. Mettere in ordine gli score in base alla pubdate
11. Creare un grafico a linee: mettere la pubdate sull’asse X e lo score sull’asse Y

P.S.

- Per scaricare tweet: <https://github.com/twintproject/twint>
- Per analisi lexicon-based delle emozioni:
<https://github.com/AntoinePassemiers/Lexicon-Based-Sentiment-Analysis>

SISTEMI MACHINE LEARNING: DEMO ONLINE

Recensioni da TripAdvisor:

IT:

Bella esperienza. incredibile la pendenza salendo cambia di continuo. troppo caro il biglietto però. in due va bene ma con la famiglia anni fa ho dovuto rinunciare.

EN:

It's my second time visiting Pisa. Nice historical site with decent restaurants food and souvenirs vendors. Coming there via train is better based on my first visit. I would not recommend parking anywhere within a mile there unless you purchase full insurance on your rental car.



SISTEMI MACHINE LEARNING: DEMO ONLINE

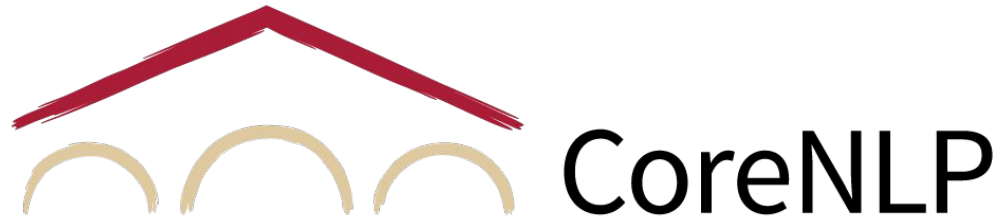


<https://cloud.google.com/natural-language?hl=it#come-funziona-automi-natural-language>

- Document-level Sentiment Analysis
- Sentence-level Sentiment Analysis

riconoscimento automatico della lingua, include l'italiano
per inglese, spagnolo e giapponese: Entity-Based Sentiment Analysis

SISTEMI MACHINE LEARNING: DEMO ONLINE



<http://nlp.stanford.edu:8080/sentiment/rntnDemo.html>

- Sentence-level Sentiment Analysis
basato su parsing a costituenti, solo inglese

APPROFONDIMENTI

Libro (pre-print):

<https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>

Lexicon-Based Methods for Sentiment Analysis:

https://www.mitpressjournals.org/doi/pdfplus/10.1162/COLI_a_00049

A Survey on Sentiment and Emotion Analysis for Computational Literary Studies: <https://arxiv.org/ftp/arxiv/papers/1808/1808.03137.pdf>

Overview of the EVALITA 2018 Aspect-based Sentiment Analysis task:

<http://ceur-ws.org/Vol-2263/paper003.pdf>

Overview of the EVALITA 2018 Task on Irony Detection in Italian Tweets:

<http://ceur-ws.org/Vol-2263/paper005.pdf>

Overview of the Evalita 2016 SENTIment POLarity Classification Task:

http://ceur-ws.org/Vol-1749/paper_026.pdf

LESSICI PER L'ITALIANO

- Lessico del progetto OpeNER:
<https://github.com/opener-project/VU-sentiment-lexicon/tree/master/VUSentimentLexicon/IT-lexicon>
- Sentix: <http://valeribasile.github.io/twita/sentix.gz>
- Morphologically-inflected Affective Lexicon (MAL):
<https://github.com/valeribasile/sentixR/blob/master/sentix/inst/extdata/MAL.tsv>
- Distributional Polarity Lexicon:
<http://sag.art.uniroma2.it/demo-software/distributional-polarity-lexicon/>
- SenticNet: <https://sentic.net/>
- NRC-Emotion-Lexicon:
<https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- NRC-VAD-Lexicon: <https://saifmohammad.com/WebPages/nrc-vad.html>
- Italian EMotive lexicon (ItEM): <https://github.com/Unipisa/ItEM>
- DepecheMood v2.0:
<https://github.com/marcoguerini/DepecheMood/releases/tag/v2.0>

DATI ANNOTATI PER L'ITALIANO

- Sentipolc 2016: <http://www.di.unito.it/~tutreeb/sentipolc-evalita16/data.html>
- ABSITA 2018, Aspect-based Sentiment Analysis: <http://sag.art.uniroma2.it/absita/data/>
- Smartphone reviews annotated for the task of aspect-based Sentiment Analysis: <https://github.com/nicolaCirillo/italian-sentiment-analysis>
- HaSpeeDe 2018, Hate Speech Detection: <http://www.di.unito.it/~tutreeb/haspeede-evalita18/data.html>
- Italian Twitter Corpus of Hate Speech: <https://github.com/msang/hate-speech-corpus>
- Happy Parents: <https://github.com/mirkolai/Happy-Parents>
- WhatsApp Dataset for Cyberbullying Detection: <https://github.com/dhfbk/WhatsApp-Dataset>
- Irony and Tweets: <https://github.com/IronyAndTweets/French-Italian-and-English-Corpora>
- IronITA 2018, Irony Detection in Italian Tweets: <http://www.di.unito.it/~tutreeb/ironita-evalita18/data.html>
- Trip-MAML, Italian and Spanish annotated reviews: http://sentiwordnet.isti.cnr.it/trip-maml/tripmaml_ites.zip
- SentiML, <http://corpus.leeds.ac.uk/marilena/SentiML/>



GRAZIE!

Email: rachele.sprugnoli@unicatt.it

Twitter: [@RSprugnoli](https://twitter.com/RSprugnoli)

