UNIVERSITY OF MANNHEIM

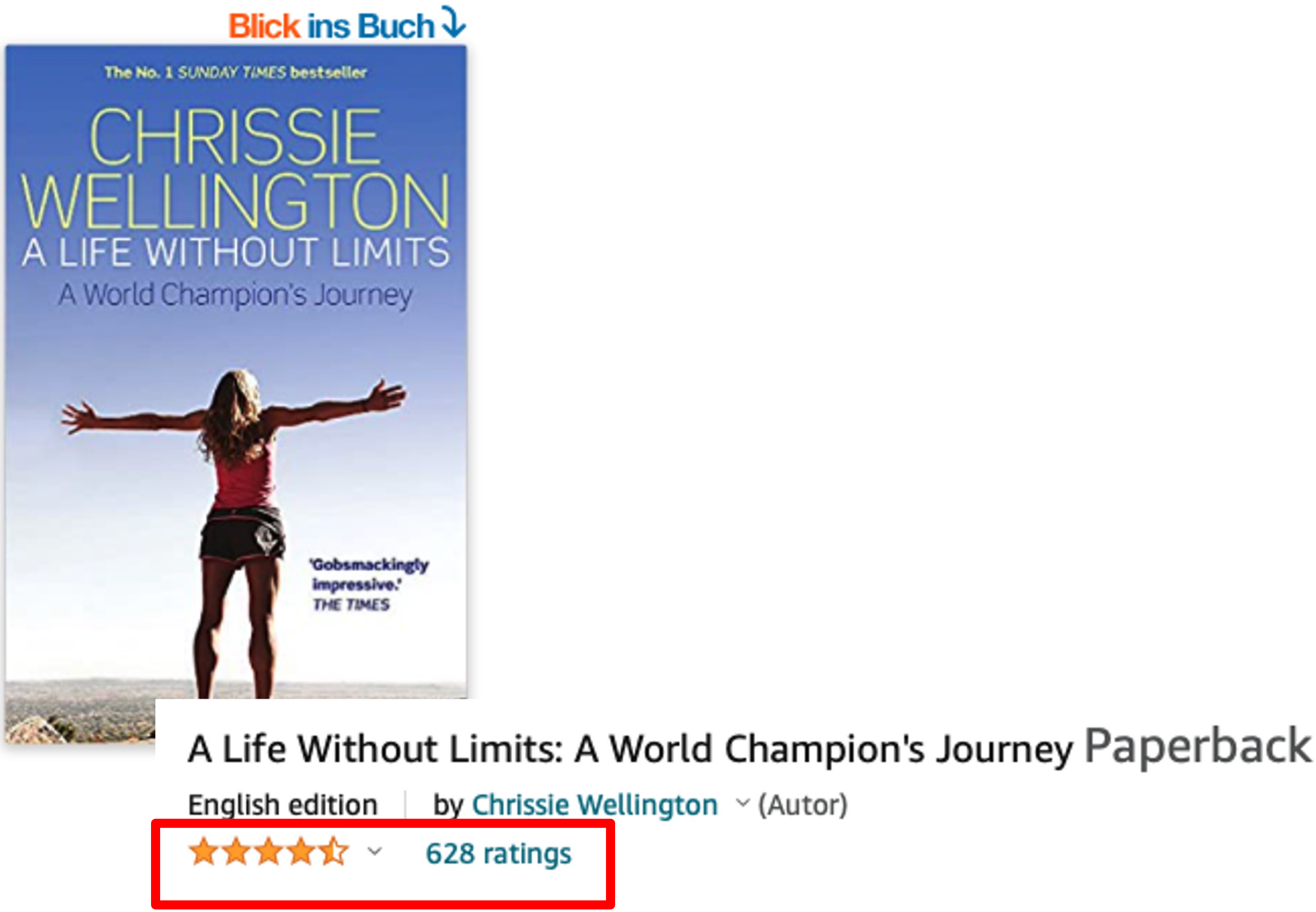
Web Mining

Web Content Mining: Detecting sentiment, sarcasm, hate

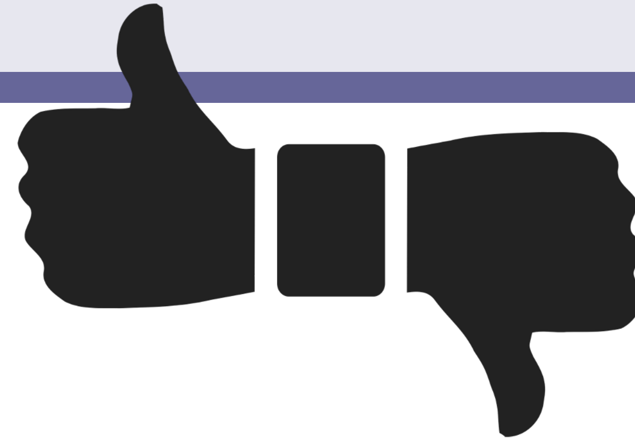
Simone Paolo Ponzetto

FSS 2023

Detecting orientation on Web data



Amazon reviews



P. Schmitt

Go Girl!

ermany on 7 September 2012

Verified Purchase

Das Buch gibt einen schönen Einblick in das Leben von Chrissie Wellington. Mir gefällt das Buch, es ist offenherzig, teilweise selbstkritisch und - das finde ich besonders gut - ohne irgendwelchen "Dann habe ich mich an XY erinnert, habe mich zusammen gerissen und bin einfach weiter gelaufen/gefahren/geflogen", wie es viele Motivationsbücher beinhalten.
Es ist eine Biografie, kein Trainingsbuch und kein ausgewiesenes Motivationsbuch. Doch gerade das macht es für mich zu einem solchen....

Micaela

Inspiring!

he United Kingdom on 29 April 2019

Verified Purchase

I've been a sporty person all my life and I have a competitive personality. I read this book and it inspired me to train hard despite my age. I only started training for marathons and triathlons after 30. This book is great for it covers a great life story, but it is also really interesting for those of us who live for sports.

Discussions on social media (Twitter)

DER SPIEGEL
@derspiegel

Regierung einigt sich offenbar auf #Testpflicht für Unternehmen: Das Wirtschaftsministerium gibt seine Blockade nach SPIEGEL-Informationen auf.

Translate Tweet

Coronapandemie: Regierung einigt sich offen
Trotz massiven Protests von Lobbygruppen v
Unternehmen in Deutschland zu einem Testa
spiegel.de

[redacted]

Tote Mitarbeiter sind auch nicht wirklich billig.

1 1

[redacted]

Replying to @derspiegel

Eine richtige und längst überfällige Maßnahme 👍‼️

1 3

[redacted]

Replying to @derspiegel

Es ist weiterhin unsinnig, symptomfreie Personen zu testen

1 4

Outline

1. Introduction to Sentiment Analysis / Opinion Mining
2. Constructing Sentiment Lexicons
3. Sentiment Classification
4. Sarcasm Detection
5. Hate Speech Detection

Sentiment Analysis and Opinion Mining

- **Opinionated text is unavoidable on the web:**
  - **Social media posts, product/service reviews**

I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. However, my mother was mad with me as I did not tell her before I bought it. She also thought the phone was too expensive, and wanted me to return it to the shop.

- **Detection of stances and opinions towards people, companies, and products/services has a tremendous business value**
  - **Improving products and services, targeted advertising, revealing trends in election campaigns, ...**

Sentiment Analysis and Opinion Mining

- Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions towards
  - Entities, individuals, issues, events, topics, and their attributes (aspects)

- Technically, it is very challenging, but practically very useful

- A general sentiment analysis framework aims to answer
  1. Who is the opinion holder?
  2. Towards whom or what is opinion/sentiment expressed?
  3. What is the polarity and intensity of the opinion?
  4. Is an opinion associated with a time-span?

Sentiment Analysis and Opinion Mining

I bought an iPhone a few days ago. It was such a nice phone. The touch screen was really cool. The voice quality was clear too. However, my mother was mad with me as I did not tell her before I bought it. She also thought the phone was too expensive, and wanted me to return it to the shop.

Opinion holder	Opinion clue	Target
I	nice	phone
(I)	really cool	touch screen
(I)	clear	voice quality
mother	mad	me
She	too expensive	phone

Sentiment Analysis and Opinion Mining

Formally, an opinion is a quintuple

$$(e_i, a_{ij}, oo_{ijkl}, h_k, t_l)$$

- e_i – the name of the entity which is the target of the expressed sentiment (e.g., iPhone)
- a_{ij} – is the aspect of the entity e_i towards which an opinion is directed (e.g., screen)
- h_k – is the person expressing the opinion (i.e., the person expressing the opinion)
- t_l – is the is the time when the opinion towards a_{ij} is expressed by h_k (or the time period during this opinion holds)
- oo_{ijkl} – is the orientation (possibly with intensity) of the opinion (e.g., negative)

- Most opinion mining studies opinions from a large number of opinion holders (⇒ need for opinion summarization)

Outline

1. Introduction to Sentiment Analysis / Opinion Mining
2. Constructing Sentiment Lexicons
3. Sentiment Classification
4. Sarcasm Detection
5. Hate Speech Detection

Sentiment Lexicons

- Sentiment clues (opinion words, sentiment-bearing words) – words and phrases used to express some desired or undesired state
  - Positive clues: *good*, *amazing*, *beautiful*
  - Negative clues: *bad*, *awful*, *terrible*, *poor*

- Sentiment clues are often domain-dependent
  - Quiet speaker phone vs. quiet car engine
  - Separate sentiment lexicons need to be constructed for different domains
    - General lexicons contain words for which the sentiment does not vary across domains

- Q: How would you automatically construct a sentiment lexicon?

Automated acquisition of sentiment lexicons

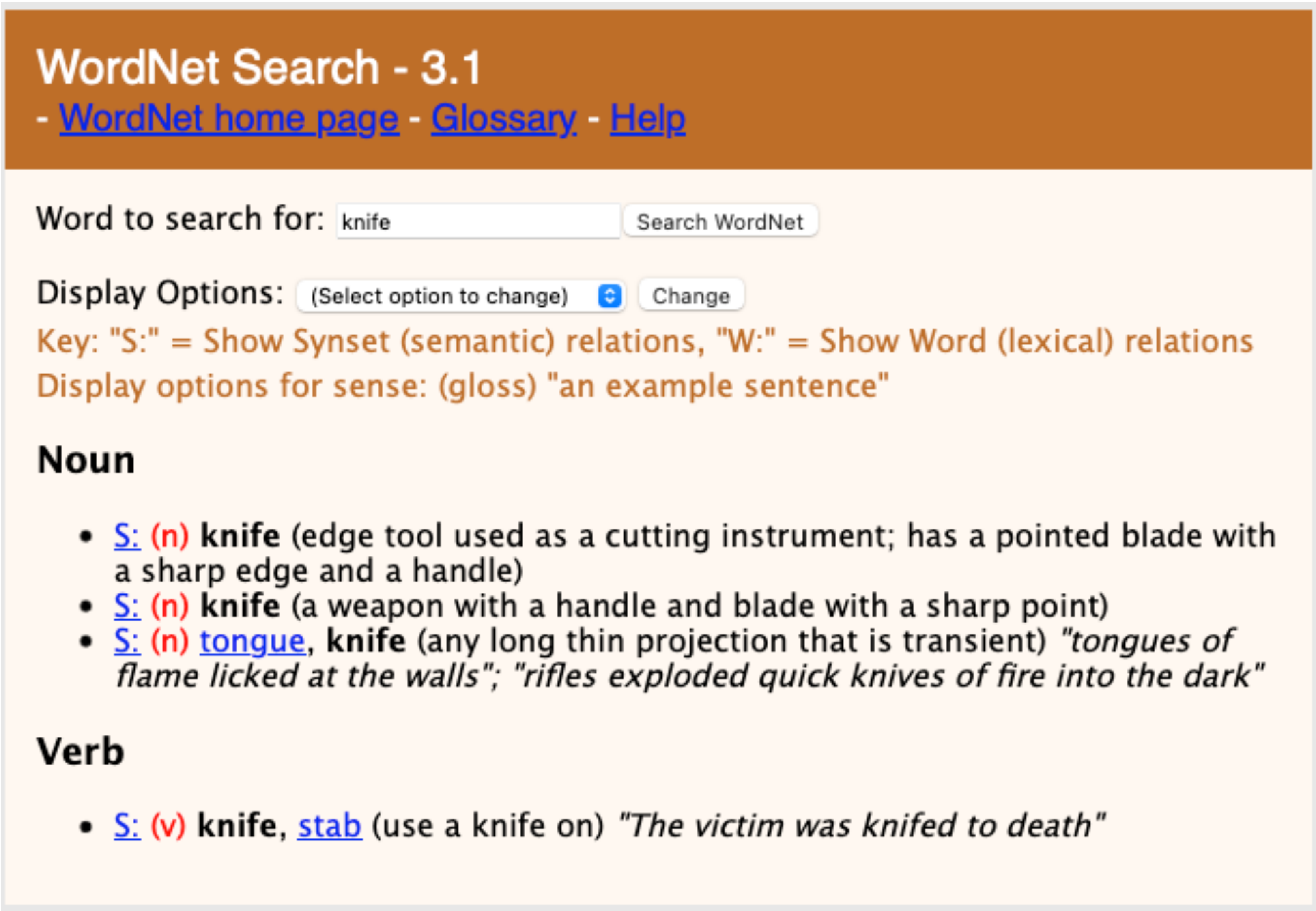
- Automated acquisition of sentiment lexicon is most often semi-supervised (or weakly supervised)
  1. Start from a small seed lexicon of sentiment words
  2. Iteratively augment the lexicon based on links between words already in the lexicon and words in the large general lexicon or large corpus
  3. Stop when there are no more reliable candidate words to be added to the lexicon
- Approaches for constructing sentiment lexicons are either
  1. Dictionary-based or
  2. Corpus-based
- Often there is a final step of manual cleansing of automatically derived sentiment lexicons

Dictionary-Based Sentiment Lexicon Acquisition

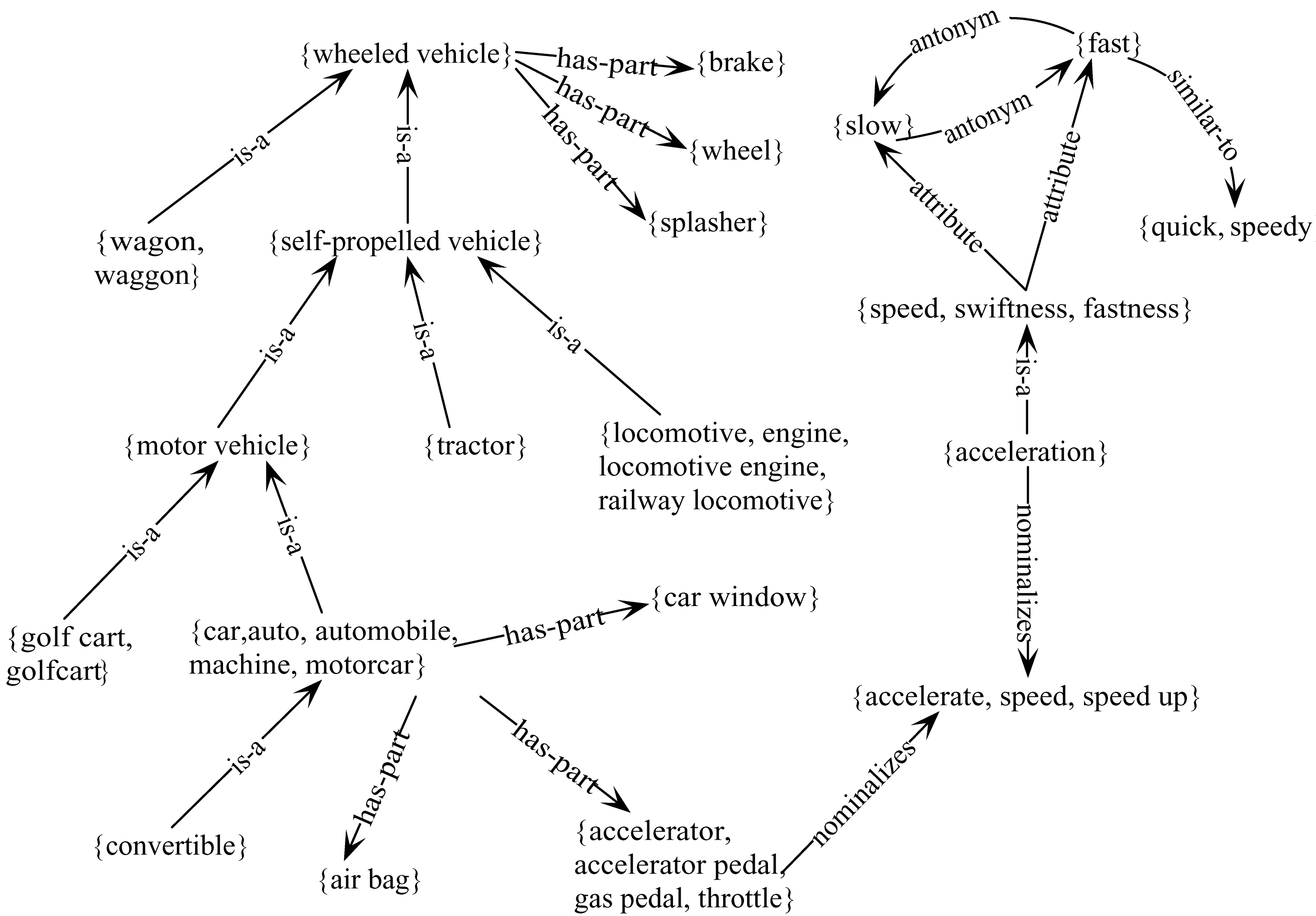
- Bootstrapping using a small seed sentiment lexicon
  - E.g., 10 positive and 10 negative sentiment words
- Idea: exploit semantic links between words in the general lexicon
  - E.g., synonymy and antonymy links in WordNet
  - The procedure is typically iterative
- Additional information can be used to make better lists
  - WordNet glosses
  - Machine learning (classification based on concept definitions)

- Q: What is the shortcoming of dictionary-based approaches?

WordNet


WordNet Search - 3.1
- WordNet home page - Glossary - Help
Word to search for: knife
Search WordNet
Display Options: (Select option to change)
Change
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"
Noun
S: (n) knife (edge tool used as a cutting instrument; has a pointed blade with a sharp edge and a handle)
S: (n) knife (a weapon with a handle and blade with a sharp point)
S: (n) tongue, knife (any long thin projection that is transient) "tongues of flame licked at the walls"; "rifles exploded quick knives of fire into the dark"
Verb
S: (v) knife, stab (use a knife on) "The victim was knifed to death"

WordNet

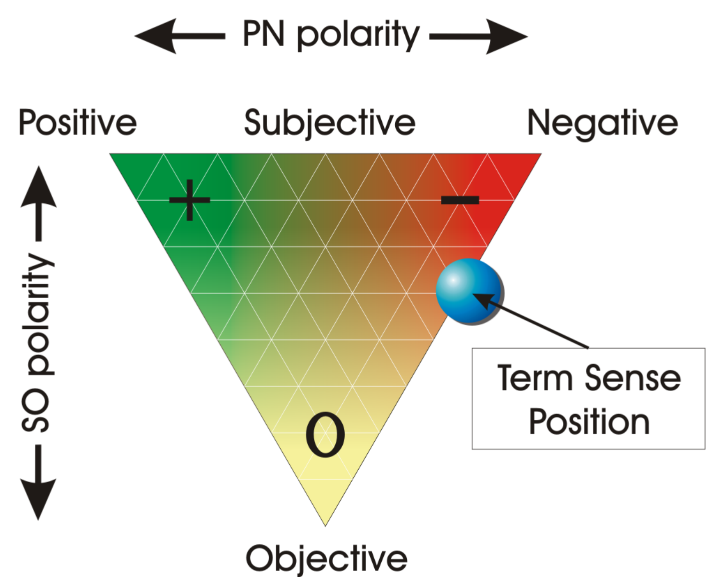


Source: Navigli (2009)

SentiWordNet

- **SentiWordNet is a general sentiment lexicon derivsed from WordNet**
  - **Esuli and Sebastiani (2006); Bacianella et al., (2010)**
- **It contains automated annotations of all WordNet synsets with sentiment scores:**
  - **Positivity score: *Pos(s)***
  - **Negativity score*:* *Neg(s)***
  - **Objectivity score*:* *Obj(s)***
  - **For each synset *s:***

  ***Pos(s)* + *Neg(s)* + *Obj(s)* = 1**



- **Construction steps:**
  1. **Semi-supervised learning step**
  2. **Random-walk step**

SentiWordNet

First step: Semi-supervised learning

1. Small positive and negative seed sets (7 synsets each)
2. Seed set expansion via WordNet relations: *also-see*, *direct antonymy*
3. Expanded seed sets used as training data for a ternary classifier *(Pos, Neg, Obj)*
   - Synset glosses used as bag-of-words features for a classifier
   - Classification performed for all WordNet synsets

Second step: The random walk

1. Construct a WordNet graph based on definiens-definiendum relation
2. Run a label propagation algorithm on the induced WordNet graph
   - Two runs: one for positive *Pos(s)* and another for negative *Neg(s)* labels
3. Normalize *Pos(s)* and *Neg(s)* over all synsets
4. Compute the objective scores, *Obj(s)* = 1 – *Pos(s)* – *Neg(s)*

Corpus-Based Sentiment Lexicon Acquisition

- **Methodologically, corpus-based induction of sentiment lexicons resembles to the dictionary-based:**
  1. **Semi-supervised learning from small initial seed sets**
  2. **Graph-based propagation of positive and negative sentiment**

- **Difference:**
  - **Graph for label propagation is computed from syntactic relations and word co-occurrences in a large corpus**
  - **The resulting lexicon specific to the domain of the corpus**

- **Some (simple) approaches:**
  - **Sentiment consistency, conjunction of adjectives (Hatzivassiloglou & McKeown, 1997)**
  - **Pointwise mutual information (PMI) of candidate words with seed set words (Turney & Littman, 2002)**
  - **PMI-induced graph with PageRank label propagation and supervised learning (Glavaš and Šnajder, 2012)**

Outline

1. Introduction to Sentiment Analysis / Opinion Mining
2. Constructing Sentiment Lexicons
3. Sentiment Classification
4. Sarcasm Detection
5. Hate Speech Detection

Sentiment classification

- The goal is to classify an opinionated portion of text (e.g., product review) as expressing (dominantly) positive or negative sentiment
  - Typically, we classify a document, but paragraphs and sentences have been addressed as well
- Assumption: entire text portion adresses a single entity
  - Holds for product reviews but not for social media posts
- Capturing the overall sentiment expressed towards the entity
  - Sentiment towards specific aspects of the entity ignored
- Methodological approaches:
  1. Supervised learning (i.e., supervised text classification; dominantly)
  2. Unsupervised learning

Supervised sentiment classification

- Typically formulated as a ternary (Positive, Negative, Neutral) text classification task
- Training and testing data – typically product reviews
  - Labels often readily available via user ratings (e.g., 1 to 5 stars)
- Classification:
  - Feature-design algorithms

    The usual suspects: logistic regression, SVM

    Features
    - Bag of words, POS tags, opinion clues and phrases (from dictionary)
    - Negations (change opinion orientation) and syntactic dependencies
  - Semantic representation-based algorithms
    - CNNs, RNNs, Autoencoders, Recursive NN (for sentiment classification)
    - Raw text input (word or character embeddings), no need for manually designed features

Logistic regression

- A probabilistic discriminative classification model
- Finds a separating hyperplane defined by a weight vector w

$$h(\mathbf{x}|\mathbf{w}) = \sigma(\mathbf{w}^{\mathrm{T}}\mathbf{x}) = \frac{1}{1 + \exp(-\mathbf{w}^{\mathrm{T}}\mathbf{x})}$$

- LR is a binary classifier: instances for which h(x|w) > 0.5 as classified as positive, other as negative

- How do we learn the weights w of the logistic regression model?
  - We need to define the objective function
  - We need to select a method for minimizing (or maximizing) the objective function

Logistic regression

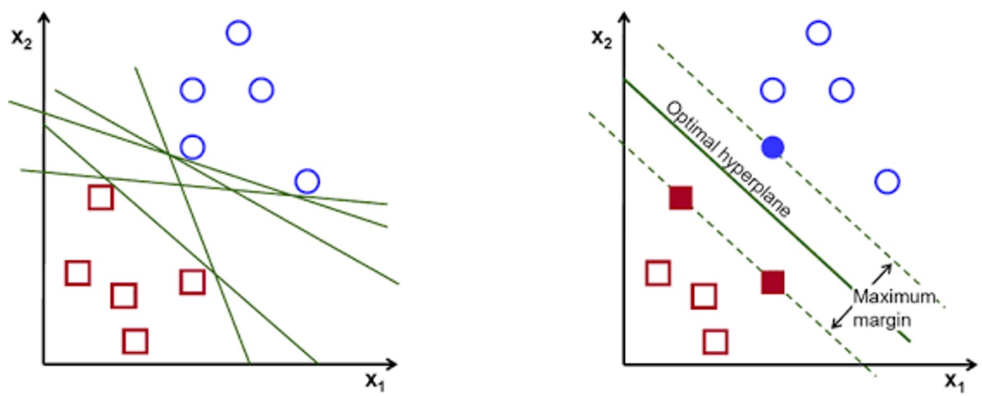
- We want to maximize the likelihood of the training set data
  - Our training data is given as N input-output pairs, $(\mathbf{x}^{(i)}, y^{(i)})$
- This corresponds to minimizing the so-called cross-entropy error (CEE):

$$-\sum_{i=1}^{N}\left\{y^{(i)}\ln h(\mathbf{x}^{(i)}) + (1-y^{(i)})\ln(1-h(\mathbf{x}^{(i)}))\right\}$$

- The cross-entropy error is 0 iff $h(x^{(i)}|w) = y^{(i)}$
- We are looking for weights w such that minimize the cross-entropy loss – there is no analytical solution
  - We resort to iterative numeric optimization
    - Stochastic gradient descent (SGD), AdaGrad, ADAM, RMSProp, ...

Support Vector Machines

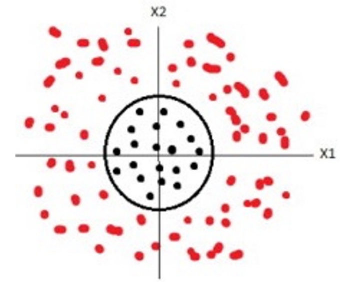
- Until the success of deep learning methods, support vector machines (SVM) was the most successful ML algorithm in NLP
- SVM is a non-probabilistic discriminative model
- Binary classifier, instance is positive if h(x) > 0 (otherwise negative)
- SVM aims to maximize the margin between:
  - The closest positive example on one side of the separating hyperplane
  - The closest negative example from the other side of the hyperplane
- Training examples that „hold the margin", „support vectors"



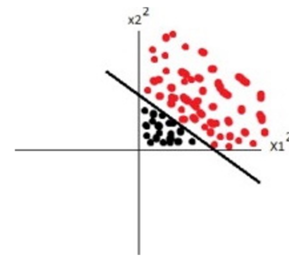
SVM – kernel „trick”

– **Basic SVM is a linear model (it identifies the linear hyperplane)**

– **In many classification tasks, positive examples are not linearly separable from negative instances**



– **According to Cover's theorem, examples are more likely to be linearly separable in the space of higher dimensionality**



– **Kernel „trick” is a method for effectively mapping examples to a higher-dimensional space without explicit mapping computation**
  - **We replace the product of two mapped instance vectors with a kernel function**

$$\kappa(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^{\mathrm{T}} \phi(\mathbf{x}')$$

Tools

– **SVM**
  - LibSVM (www.csie.ntu.edu.tw/~cjlin/libsvm)
    - Linear and nonlinear (kernelized) SVM
  - LibLinear (www.csie.ntu.edu.tw/~cjlin/liblinear)
    - Linear SVM and linear logistic regression
    - To be used if your model is linear (due to efficiency reasons)!
  - LibShortText (www.csie.ntu.edu.tw/~cjlin/libshorttext)
  - TinySVM (chasen.org/~taku/software/TinySVM)
    - Large-scale SVM training (100K feats, 10K instances)

– **Many algorithms**
  - Scikit-Learn (scikit-learn.org/stable/index.html)
  - TensorFlow (https://www.tensorflow.org)
  - Weka (www.cs.waikato.ac.nz/ml/weka) / RapidMiner
  - R (www.r-project.org)

Feature design

- The key question is how to come up with good (useful) features
- Two approaches:
  - Use your intuition (insight, linguistic/domain expertise), and design a small set of good features that you think should work
  - Throw in everything you can (the "kitchen sink" approach), and them maybe prune later
- You will often want to see which features work and which don't:
  - Ablation study – turn off some features, retrain the model and see how the performance changes
  - Feature selection – use a method to select the best features. This can also improve the performance (especially in a "kitchen sink" approach)
- One of the great advantages of deep learning for NLP is the absence of feature engineering

One-hot encoding

- Many successful ML models (LR, SVM) only work with real-valued feature vectors
  - Q: Why not encode discrete features as numbers from {1, . . . , K}?
  - A: Not good because discrete values should remain unordered
- Categorical features should be encoded using **one-hot encoding**
  - For text, most features are categorical (words, POS-tags)
  - One-hot-encoding is a K-dimensional binary vector with only one component set to 1

Value 1 => [1, 0, 0, ..., 0, 0]

Value 2 => [0, 1, 0, ..., 0, 0]

...

Value K => [0, 0, 0, ..., 0, 1]

- Vector dimensions equal the number of values the feature can take.

Bag-of-words text representation

– Up to recently, most text classification models represented texts as **unordered multi-set („bag”) of words**

– Given a text, it's „bag-of-words” vector has non-zero elements at indices representing words the text contains

  - Sometimes called „few-hot-encoding”, as each text is expected to contain only a small subset of words from the whole vocabulary

– *„Frodo and Sam are friends”*

[0, ..., 0, 1, 0, ..., 0, 1, ..., 0, ...1, 0, ..., 0]

– Instead of binary indicators if the word appeared in the document, we often use **real-valued weights**

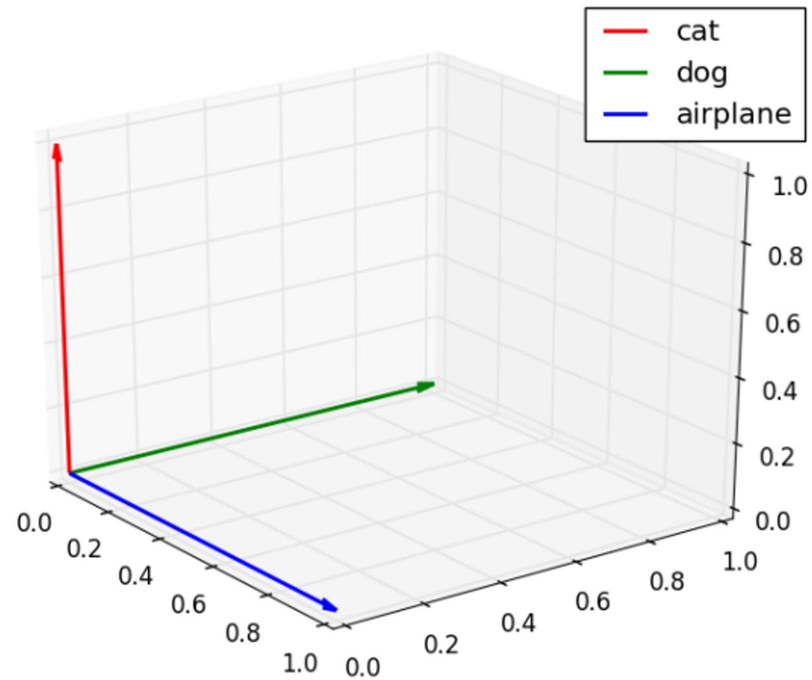
  - TF-IDF (term-frequency * inverse document frequency) is most commonly used weighting scheme

Word representations

Sparse representation

- Each word is represented by a **one-hot vector**, i.e., it is given a unique symbolic ID
- The dimension of the symbolic representation for each word is equal to the size of the vocabulary V (number of words)
- All but one dimension are equal to zero, and one is set to one

$$v_{word} = (\ldots, 0, 1, 0, \ldots)$$

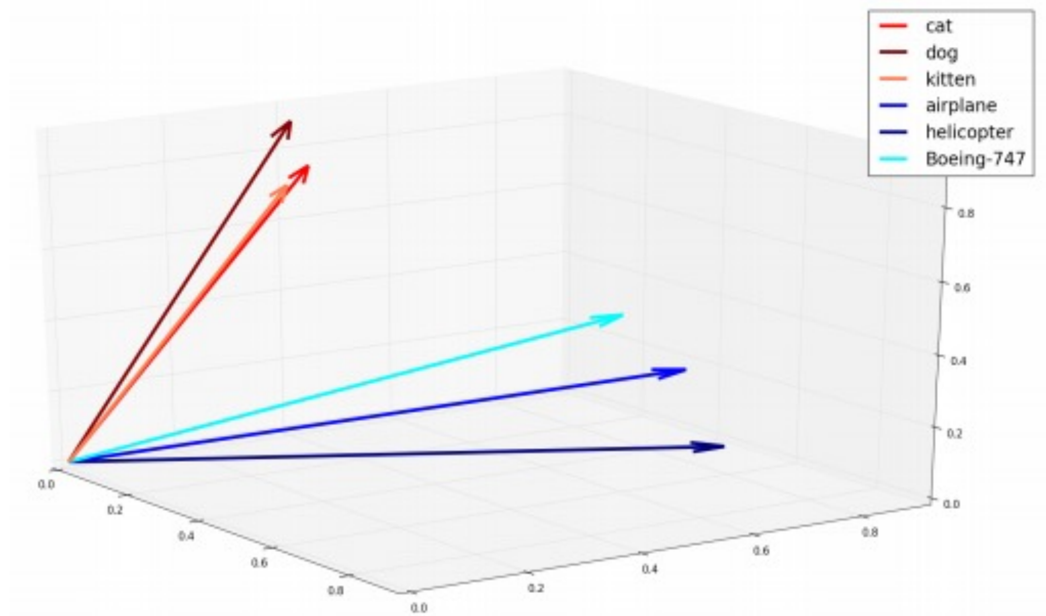


Word representations

Dense representation

- Each word is represented by a dense vector, a **point in a vector space**
- The dimension of the semantic representation d is usually much smaller than the size of the vocabulary (**d << V**)
- All dimensions contain real-valued numbers (possibly normalized between −1 and 1)

v_{word} = (..., 0.3, −0.5, 0.1,...)



Unsupervised Sentiment Classification

- If user ratings are not available, we need manual labelling for supervised machine learning methods
  - Tedious, expensive, time-consuming
- A typical unsupervised approach to sentiment classification:
  1. Extract candidate phrases (e.g., matching predefined POS patterns)
  2. For reach word/phrase, compute some association score (e.g., pointwise mutual information) with sentiment lexicon entries, on a large corpus
     - Association scores (e.g., PMI) with positive seed words
     - Association scores (e.g., PMI) with negative seed words
  3. The sentiment orientation of each phrase is computed as:

$$SO(phr) = \frac{1}{|pos|} \cdot \sum_{p \in pos} PMI(phr, p) - \frac{1}{|neg|} \cdot \sum_{n \in neg} PMI(phr, n)$$

  4. The sentiment of the document is determined by summing or averaging the sentiment orientations of phrases it contains

Aspect-Oriented Sentiment Analysis

- Sentiment classification at a document or sentence level is useful but it doesn't tell the whole story
  - Does not account for specific aspects of entities
  - Detailed aspect-based sentiment specification is needed for many applications

The service provided by the staff was great. They took us from the bus station and up to the apartment. Since the price was so low, we didn't have any high expectations, but it was nice. Found the beds uncomfortable, though. It's rather difficult to find the office on arrival.

- Unlike sentiment classification, aspect based sentiment analysis requires NLP techniques for more fine-grained analysis of text

Outline

1. Introduction to Sentiment Analysis / Opinion Mining
2. Constructing Sentiment Lexicons
3. Sentiment Classification
4. Sarcasm Detection
5. Hate Speech Detection

Sarcasm detection

- Non-transparent expressions of sentiment cause most errors in sentiment analysis and opinion mining
  - Irony and sarcasm being most salient
- Sarcasm is a sharp, bitter, or cutting expression or remark; a bitter gibe or taunt
- Sarcasm is notoriously difficult to detect in text, even for humans!

Sarcasm detection

- The variation by which sarcasm is expressed is basically unlimited
- Computational approaches focus merely on specific types of sarcasm
  - Sarcasm as contrast of negative situations and positive sentiment (Riloff et al., 2013)
- Sarcasm as contrast – examples
  - Oh how I love being ignored.
  - Thoroughly enjoyed shoveling the driveway today!
  - Absolutely adore it when my bus is late.
  - I’m so pleased mom woke me up with vacuuming this morning.

Detecting Sarcasm as Contrast

- Detecting sarcasm in tweets as contrast between negative situation and positive sentiment
- Boostrapping rule-based algorithm that automatically learns positive sentiment phrases and negative situation phrases:
  1. Start with (1) single positive sentiment word (*love*) and (2) a set of tweets with hashtag #sarcasm or #sarcastic
  2. Negative situation candidates – n-grams (1-3) that directly follow positive sentiment phrases and fulfill pre-defined POS patterns
  3. Positive sentiment candidates – n-grams (1-3) near the negative situation phrases that satisfy POS patterns
  4. Candidates are scored based on ratio of frequencies in sarcastic (with hashtags) vs. non-sarcastic tweets

Detecting Sarcasm as Contrast

- **Some extracted positive sentiment phrases:**
  - ***missed***, ***loves***, ***enjoy***, ***can't wait***, ***excited***, ***wanted***, ***can't wait***, ***appreciate***, ***loving***, ***really like***, ***looooove***, ***just keeps***, **loveee**, **...**
- **Some extracted negative situation phrases:**
  - ***being ignored***, ***being sick***, ***waking up early***, ***cleaning***, ***crying***, ***sitting at home***, ***being told***, ***not sleeping***, ***not talking***, ***doing homework***, ***being ditched***, ***falling***, ***walking home***, ***getting yelled at***, ***taking care***, **...**

- **Detection performance: 51% F1-score**
  - **On a very constrained sarcasm detection task**
  - **Just proves the difficulty of sarcasm detection**

Outline

1. Introduction to Sentiment Analysis / Opinion Mining
2. Constructing Sentiment Lexicons
3. Sentiment Classification
4. Aspect-Oriented Sentiment Analysis
5. Sarcasm Detection
6. Hate Speech Detection

Hate Speech

- **Hate speech (HS) is commonly defined as any communication that**
  - **disparages a person or a group**
  - **on the basis of some characteristic such as race, color, ethnicity, gender, sexual orientation, nationality, religion, or other.**
- **Expressions that:**
  - (i) **incite discrimination or violence due to racial hatred, xenophobia, sexual orientation and other types of intolerance;**
  - (ii) **foster hostility through prejudice and intolerance.**

J. T. Nockleby (2000). **Hate speech**. Encyclopedia of the American Constitution (2nd ed., edited by Leonard W. Levy, Kenneth L. Karst et al., New York: Macmillan), pp. 1277–1279

Hate Speech and social media

Facebook Admits It Was Used to Incite Violence in Myanmar

Rohingya refugees after crossing the Naf River, which separates Myanmar and Bangladesh, in 2017. A report commissioned by Facebook found the company failed to keep its platform from being used to "foment division and incite offline violence" in Myanmar. Adam Dean for The New York Times

https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html

Hate Speech: definitions

Source	Hate speech is to incite violence or hate	Hate speech is to attack or diminish	Hate speech has specific targets	Humour has a specific status
EU Code of conduct	Yes	No	Yes	No
ILGA	Yes	No	Yes	No
Scientific paper	No	Yes	Yes	No
Facebook	No	Yes	Yes	Yes
YouTube	Yes	No	Yes	No
Twitter	Yes	Yes	Yes	No

P. Fortuna, S. Nunes (2018). **A survey on automatic detection of hate speech in text.** ACM Computing Surveys (CSUR) 51.4

Hate Speech: more definitions!

Table 1 Glossary of terms relevant to the present survey, with their definitions from the literature

Term and definitions	Source
Hate Speech	Warner and Hirschberg (2012)
Any communication that disparages a person or a group on the basis of some characteristic such as race, ethnicity, gender, sexual orientation, nationality, religion, or other characteristic	
Use of a sexist or racial slur, attack a minority, promotes hate speech or violent crime, blatantly misrepresents truth, shows support of problematic hashtags, defends xenophobia or sexism, or contains a screen name that is offensive	Waseem and Hovy (2016)
Act of offending, insulting or threatening a person or a group of similar people on the basis of religion, race, caste, sexual orientation, gender or belongingness to a specific stereotyped community	Schmidt and Wiegand (2017)
Language that is used to express hatred towards a targeted group or is intended to be derogatory, to humiliate, or to insult the members of the group	Davidson et al. (2017)
Any communication that disparages a target group of people based on some characteristic such as race, colour, ethnicity, gender, sexual orientation, nationality, religion, or other characteristic	Nockleby (2000)

F. Poletto, V. Basile, M. Sanguinetti, C. Bosco, V. Patti. **Resources and benchmark corpora for hate speech detection: a systematic review**. Language Resources and Evaluation, 2020

Example tweets

[Example of indirect insult.]
@USER Everyone saying ***fuck*** *Russ dont know a damn thing about him or watched the interview* 🤷‍♀️🤷‍♀️🤷‍♀️

[Ex. i): offensive tweet & abusive swearing]
@USER You are an absolute ***dick*** 😡

[Ex. ii): offensive tweet & not abusive swearing]
@USER I was definitely drunk as ***shit***

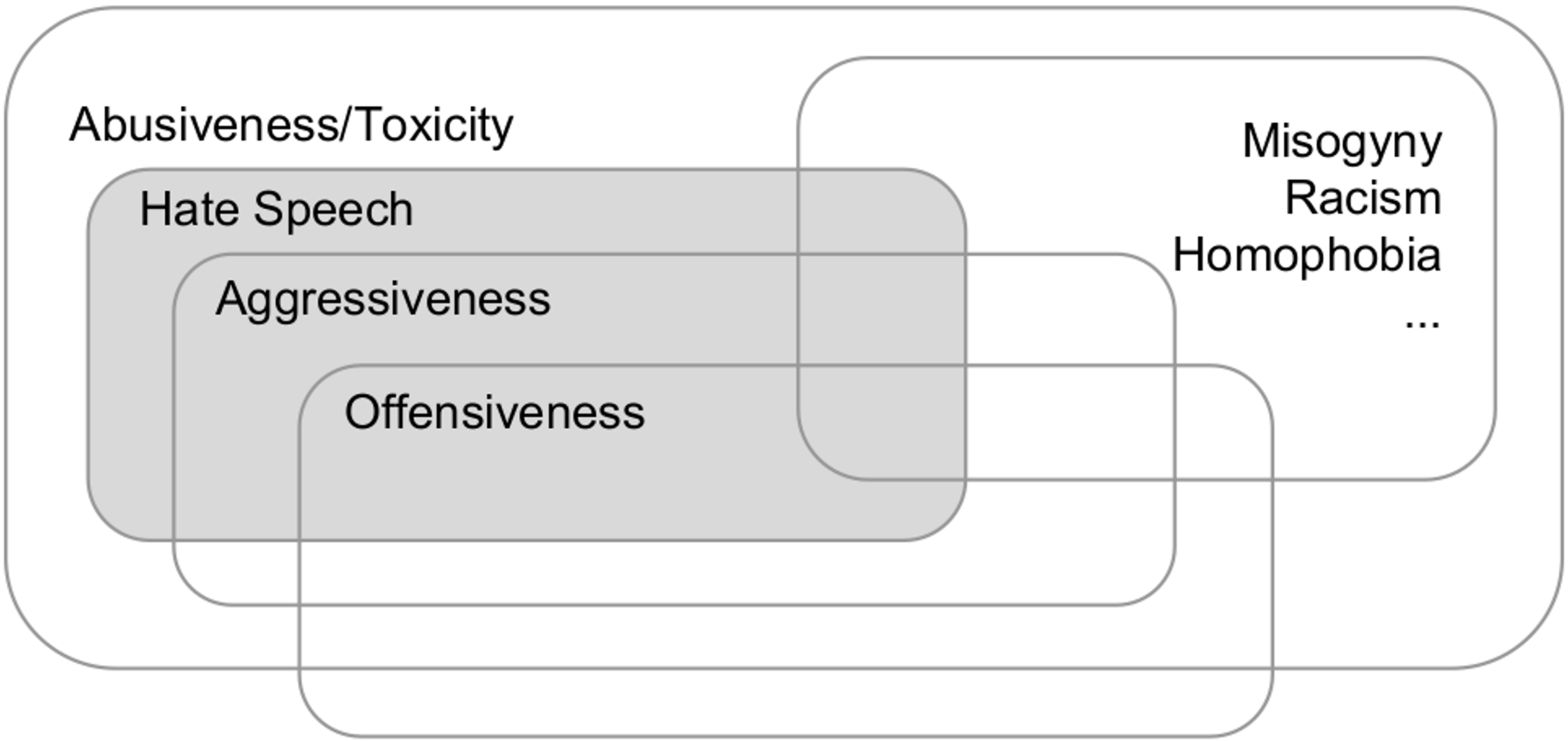
[Ex. iii): not offensive tweet & abusive swearing]
@USER ***bullshit*** *there's rich liberals too so what are you saying ???*

[Ex. iv): not offensive tweet & not abusive swearing]
@USER Haley thanx! you know how to brighten up my ***shitty*** *day* 😘

Endang Wahyu Pamungkas, Valerio Basile, and Viviana Patti. 2020. Do You Really Want to Hurt Me? Predicting Abusive Swearing in Social Media. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6237–6246, Marseille, France. European Language Resources Association.

Hate Speech, offensive language, etc.

- One of the major issues consists in the intrinsic complexity in defining HS and in a widespread vagueness in the use of related terms (such as **abusive, toxic, dangerous, offensive or aggressive language**), that often **overlap** and are prone to strongly **subjective interpretations**



Lexicons for hate speech / offensive language

- **Just like there exists sentiment lexicons we have lexicons for hurtful language**

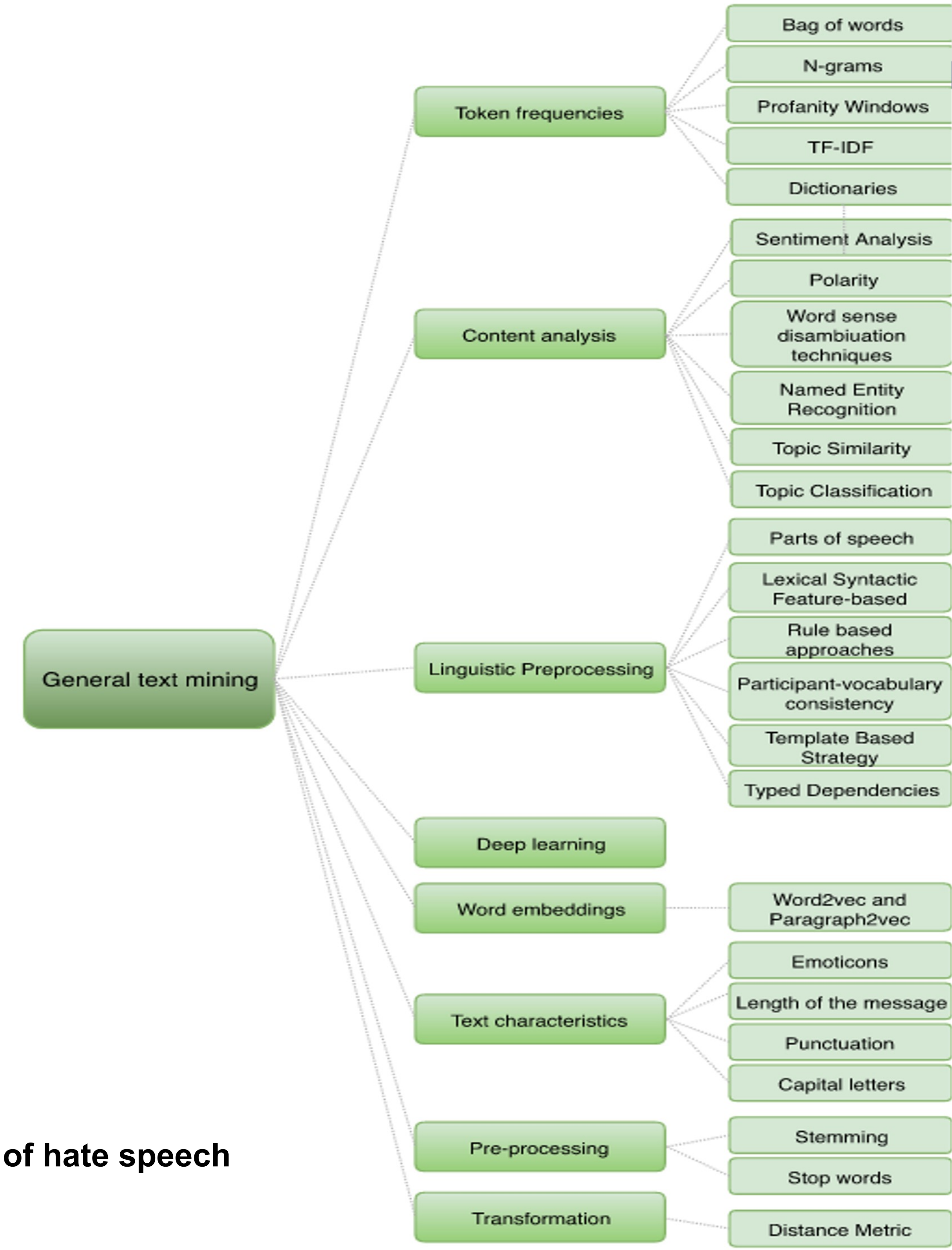
- **HurtLex (Bassignana et al., 2018)**
  - **Multilingual lexicon of “words to hurt”**
  - **53 languages**
  - **17 categories + stereotype**

HurtLex (Bassignana et al., 2018)

Category	# Terms	Examples
negative stereotypes ethnic slurs	371	barbarian, idiotic, dummy, n***oes, infertility
locations and demonyms	24	genoan, savage, barbarian, tike, boor
professions and occupations	192	wooer, politician, peasant, fishwife, academism
physical disabilities and diversity	63	handycapped, midget, worthless, invalidity, impaired
cognitive disabilities and diversity	491	artless, retarded, simple, goof, brute
moral and behavioral defects	715	close-minded, cheater, stinking, forgery, faker
words related to social and economic disadvantage	124	miscreants, miserable, wretch, pitiful, villain
plants	177	finocchio, potato, papaya whip, squash, f**ot
animals	996	b***h, t**t, goose, scoundrel, beastly
male genitalia	426	wanky, c**k, testicles, phallic, prick
female genitalia	144	babe, c**t, t**t, boob, p***y
words related to prostitution	276	s*ut, street walker, crack h*, hooker, w***e
words related to homosexuality	361	drag, crossdressing, shirtlifter, f**, qu**rio
with potential negative connotations	518	bollocks, acolyth, delirious, reject, mooch
derogatory words	2,204	scalawag, boaster, rustler, dunderheaded, pedant
felonies and words related to crime and immoral behavior	619	mafioso, roguery, robber, scalawag, rapscallion
words related to the seven deadly sins of the Christian tradition	527	concupiscience, laziness, vanity, madness, slacker

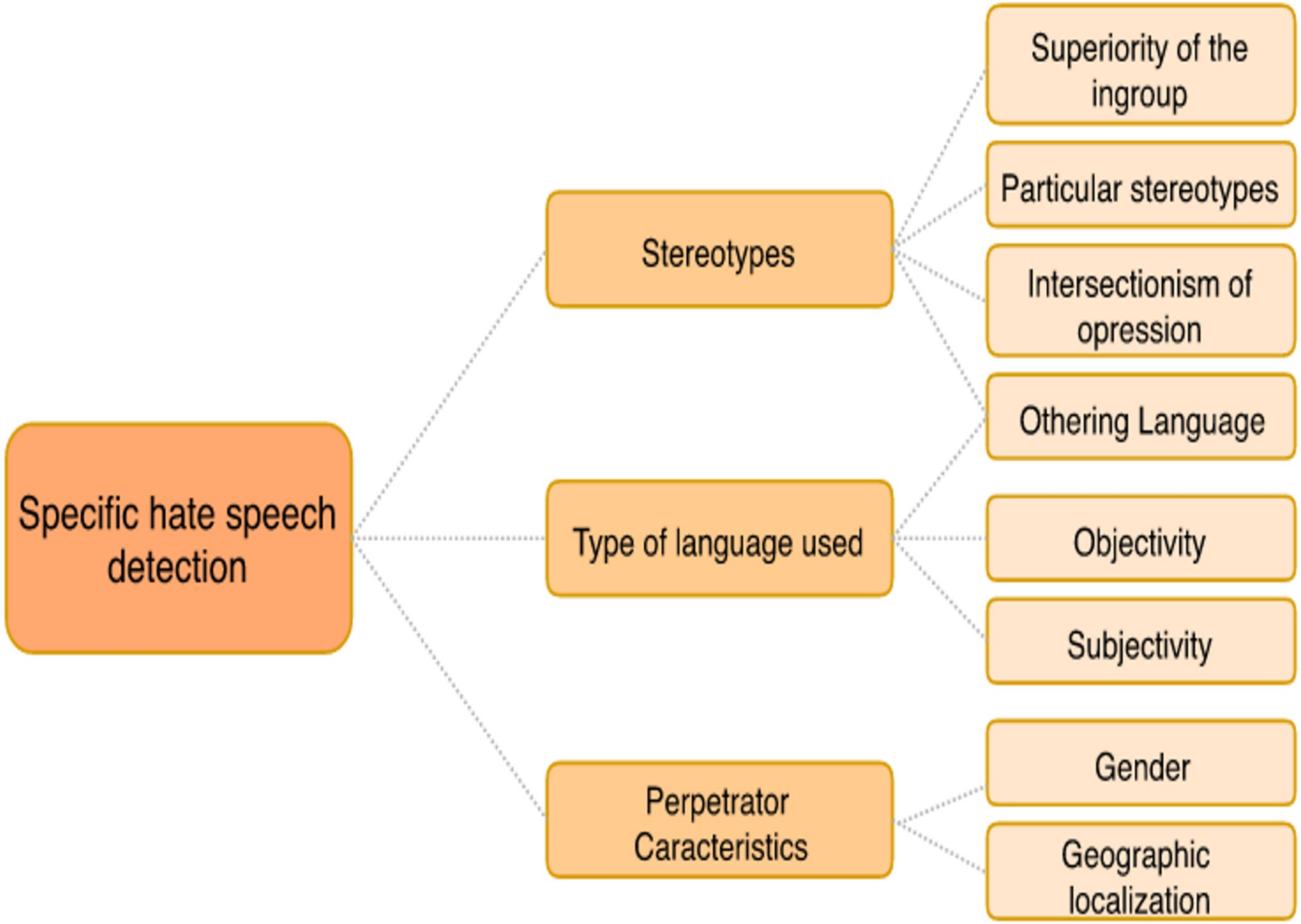
Hate Speech Detection

- **Typically addressed as a text classification task**
- **Binary or multi-label**
- **Supervised**



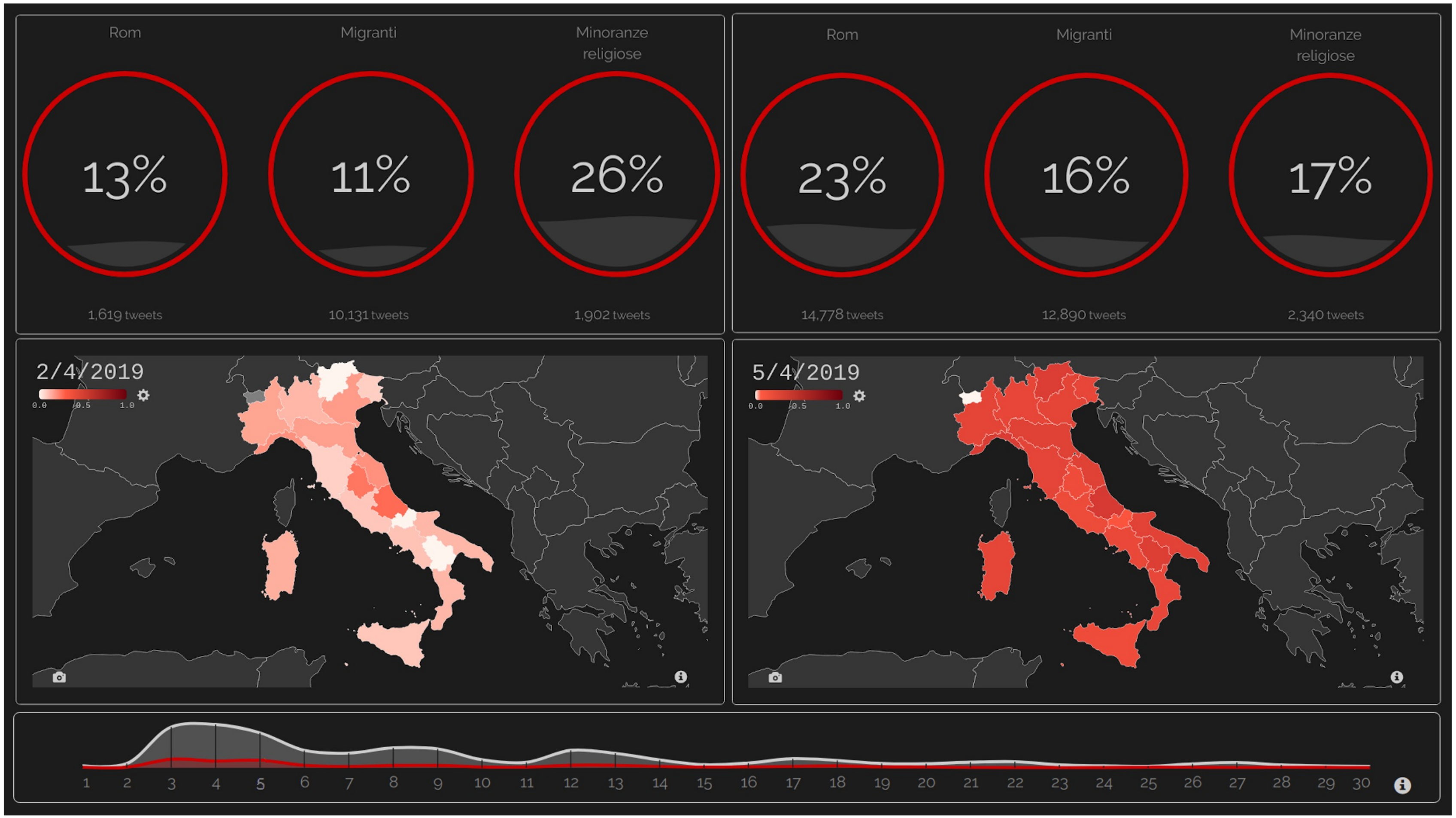
P. Fortuna, S. Nunes (2018). **A survey on automatic detection of hate speech in text.** ACM Computing Surveys (CSUR) 51.4

Specific approaches for HS detection



P. Fortuna, S. Nunes (2018). **A survey on automatic detection of hate speech in text.** ACM Computing Surveys (CSUR) 51.4

Applications: online monitoring of HS



A. T. E. Capozzi et al. (2019). **Computational linguistics against hate: hate speech detection and visualization on social media in the "Contro L'Odio" project**. In Proc. CLiC-it 2019, ceur-ws.org, vol. 2481

Summary

- Web Content Mining
  - Sentiment analysis
  - Sarcasm detection
  - Hate Speech and Offen

Next week: More Content Mining