大數據行銷研究
Big Data Marketing Research

Tamkang University
淡江大學

大數據情感分析
(Big Data Sentiment Analysis)

1051BDMR11
MIS EMBA (M2262) (8638)
Thu, 12,13,14 (19:20-22:10) (D409)

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系
http://mail. tku.edu.tw/myday/

2016-12-23

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2016/09/16	中秋節 (調整放假一天) (Mid-Autumn Festival Holiday)(Day off)
2	2016/09/23	大數據行銷研究課程介紹 (Course Orientation for Big Data Marketing Research)
3	2016/09/30	資料科學與大數據行銷 (Data Science and Big Data Marketing)
4	2016/10/07	大數據行銷分析與研究 (Big Data Marketing Analytics and Research)
5	2016/10/14	測量構念 (Measuring the Construct)
6	2016/10/21	測量與量表 (Measurement and Scaling)

課程大綱 (Syllabus)

週次 (Week) 日期 (Date) 內容 (Subject/Topics)

7 2016/10/28 大數據行銷個案分析 I
(Case Study on Big Data Marketing I)

8 2016/11/04 探索性因素分析 (Exploratory Factor Analysis)

9 2016/11/11 確認性因素分析 (Confirmatory Factor Analysis)

10 2016/11/18 期中報告 (Midterm Presentation)

11 2016/11/25 社群運算與大數據分析
(Social Computing and Big Data Analytics)

12 2016/12/02 社會網路分析 (Social Network Analysis)

課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
13	2016/12/09	大數據行銷個案分析 II (Case Study on Big Data Marketing II)
14	2016/12/16	社會網絡分析量測與實務 (Measurements and Practices of Social Network Analysis)
15	2016/12/23	大數據情感分析 (Big Data Sentiment Analysis)
16	2016/12/30	金融科技行銷研究 (FinTech Marketing Research)
17	2017/01/06	期末報告 I (Term Project Presentation I)
18	2017/01/13	期末報告 II (Term Project Presentation II)

Outline

- Sentiment Analysis
- Architectures of Sentiment Analytics
- Opinion Spam Detection
- Text Mining Techniques and Natural Language Processing

Sentiment Analysis

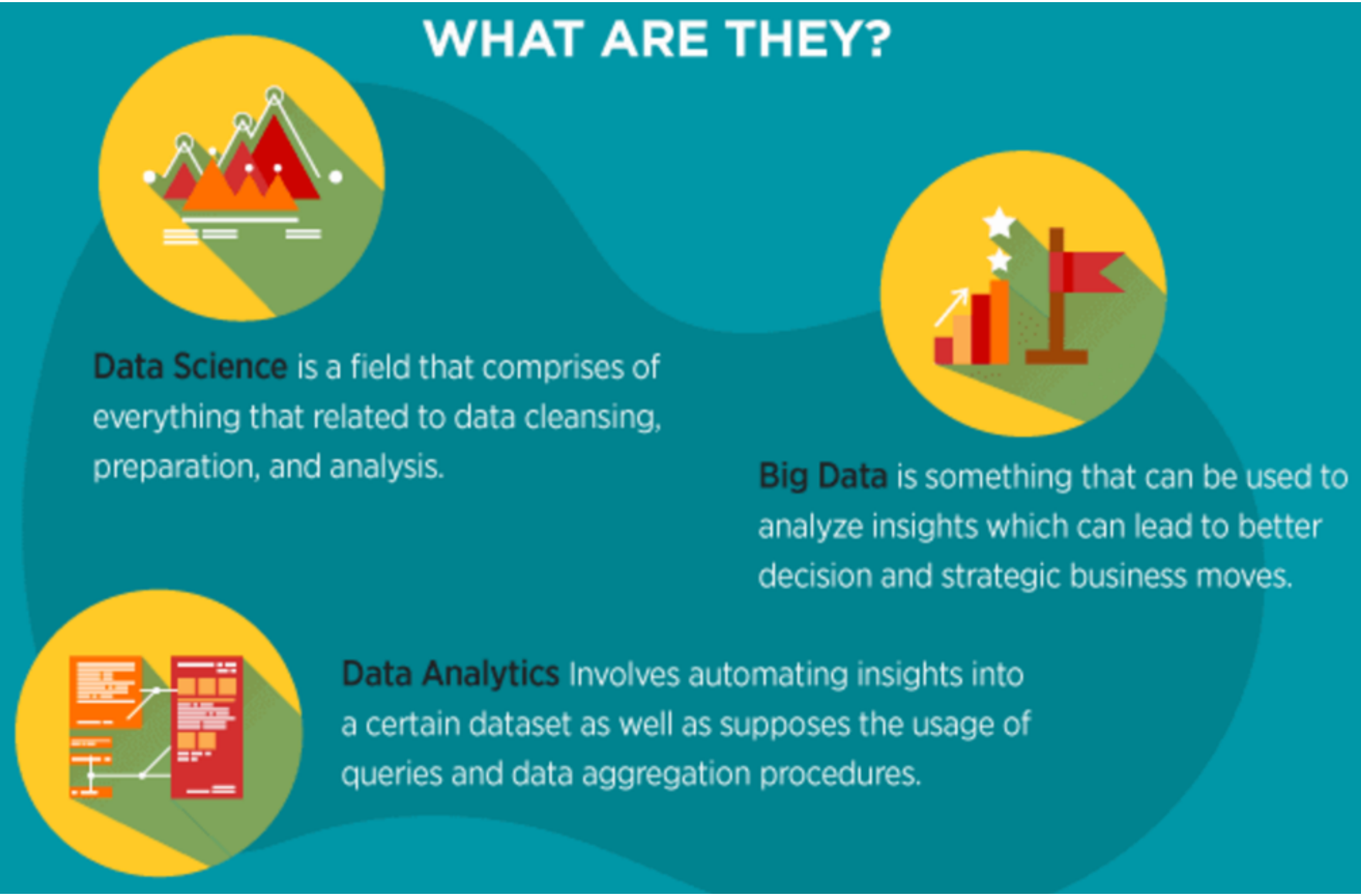
Data Scientist
資料科學家

Source: http://www.ibmbigdatahub.com/infographic/what-makes-data-scientist

Data Science vs. Big Data vs. Data Analytics

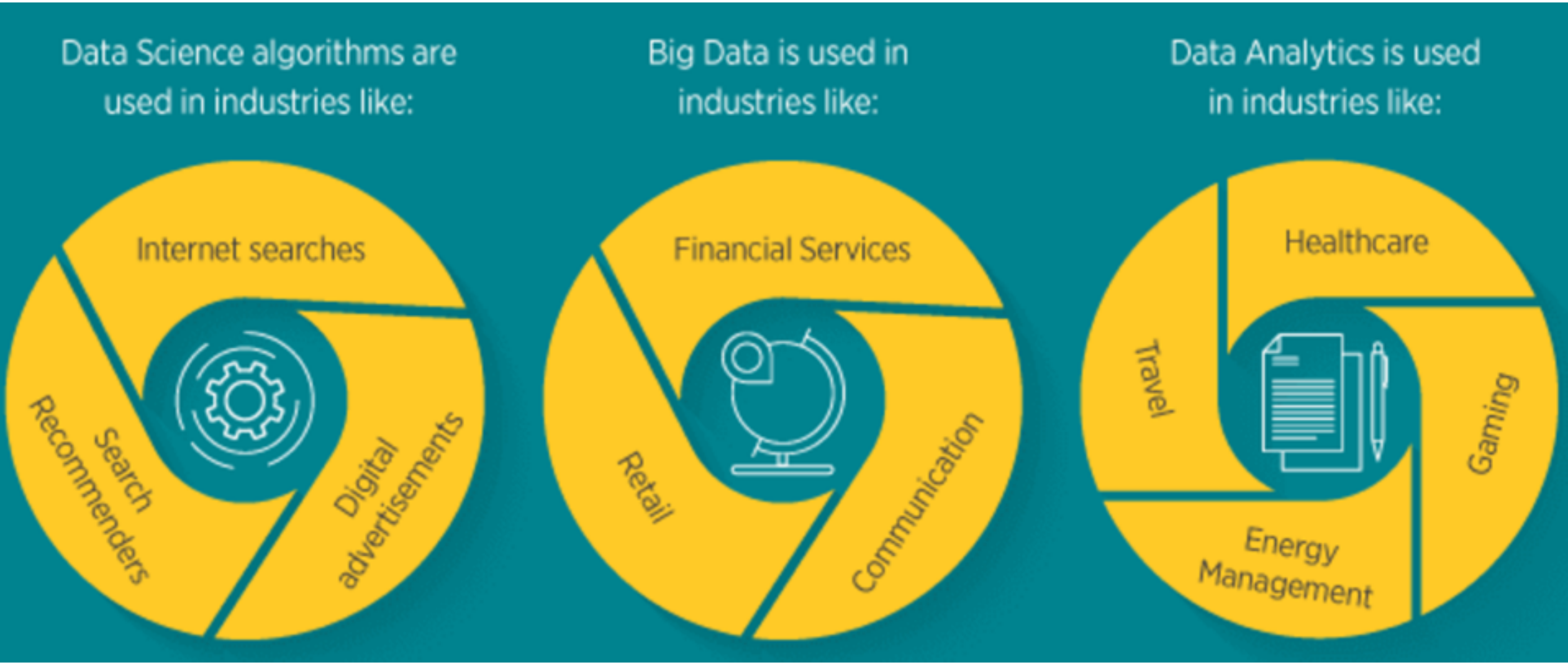
Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Data Science vs. Big Data vs. Data Analytics



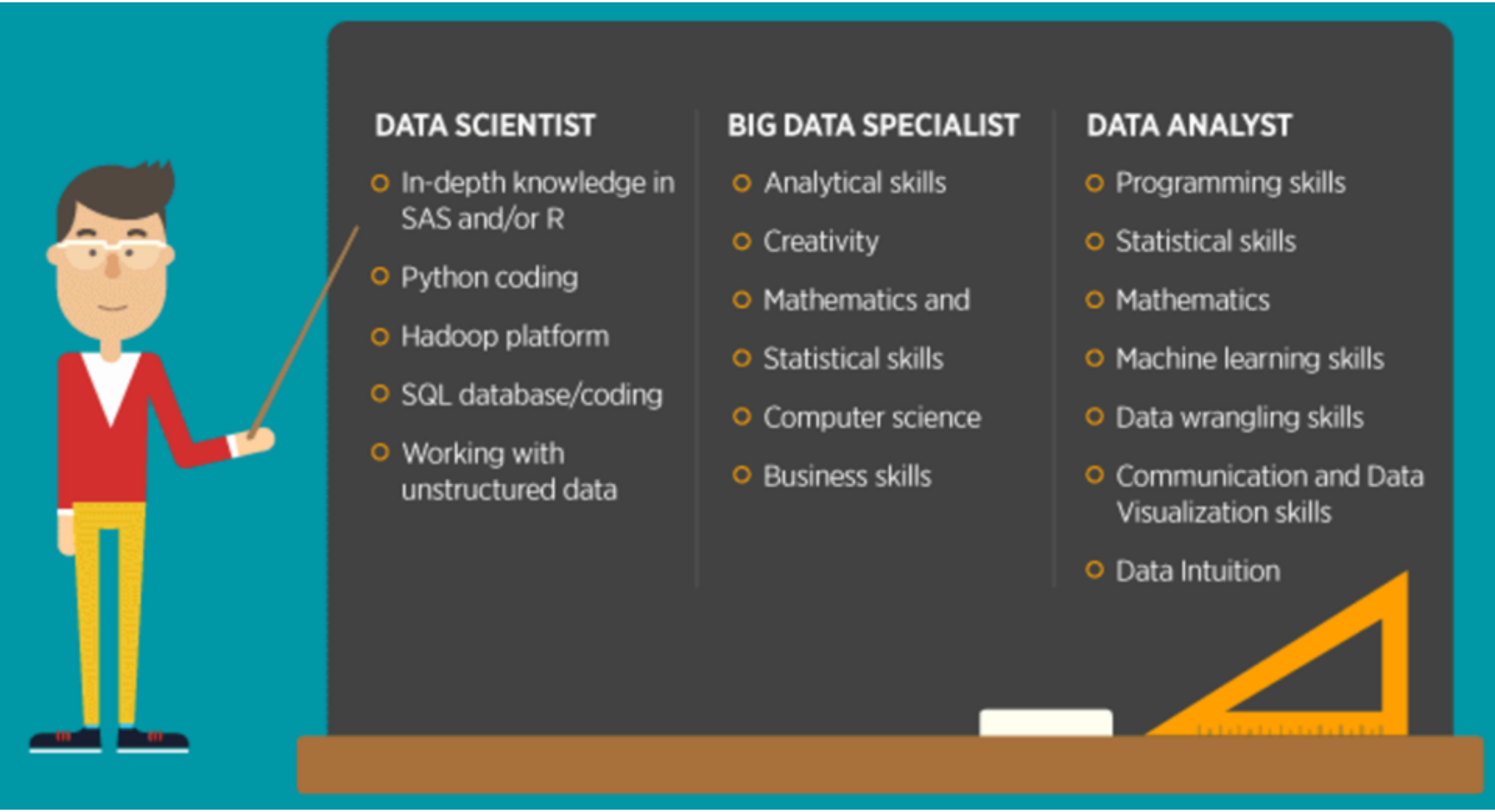
Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

What are they used?

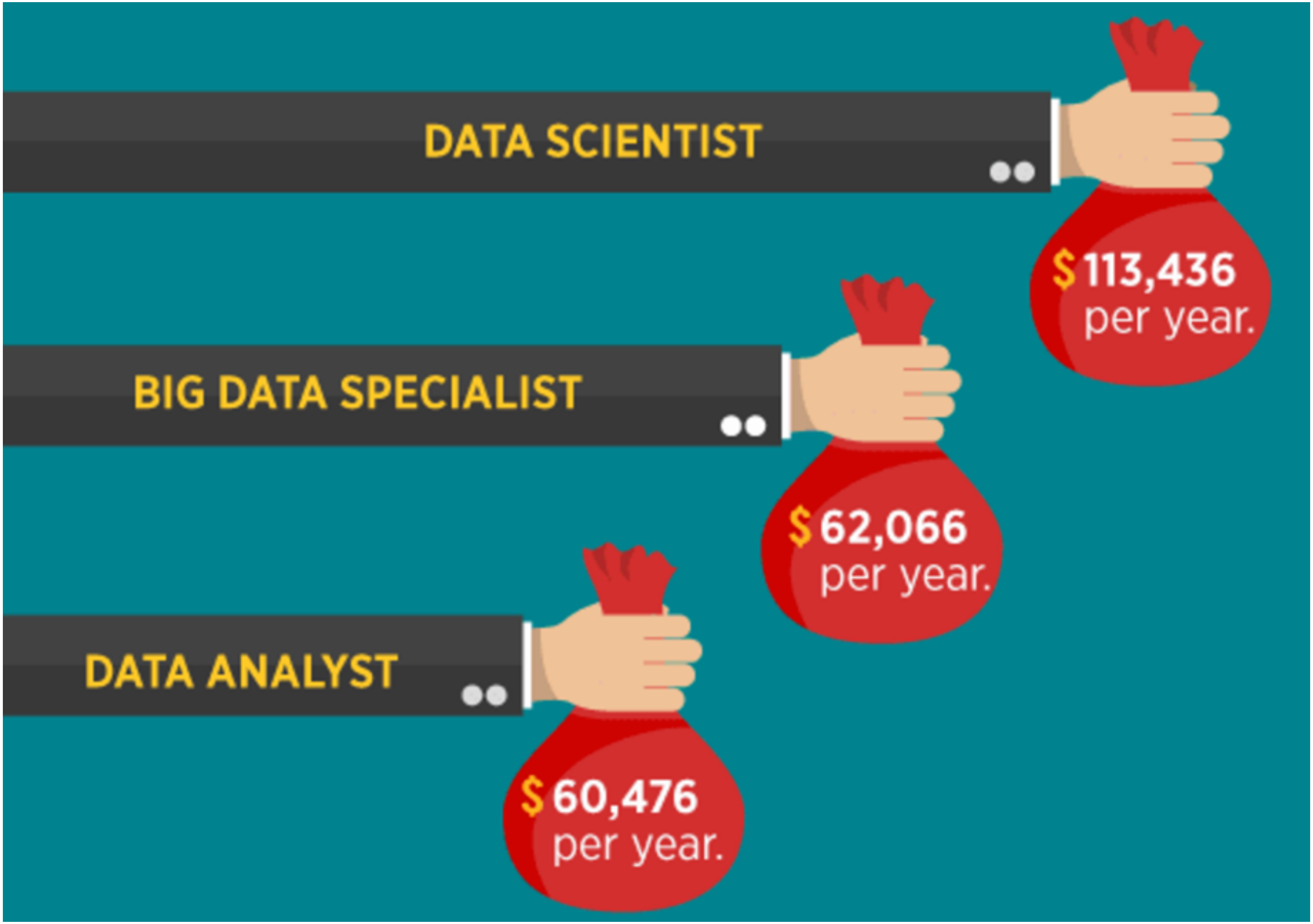


Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Data Science
What are the Skills Required?



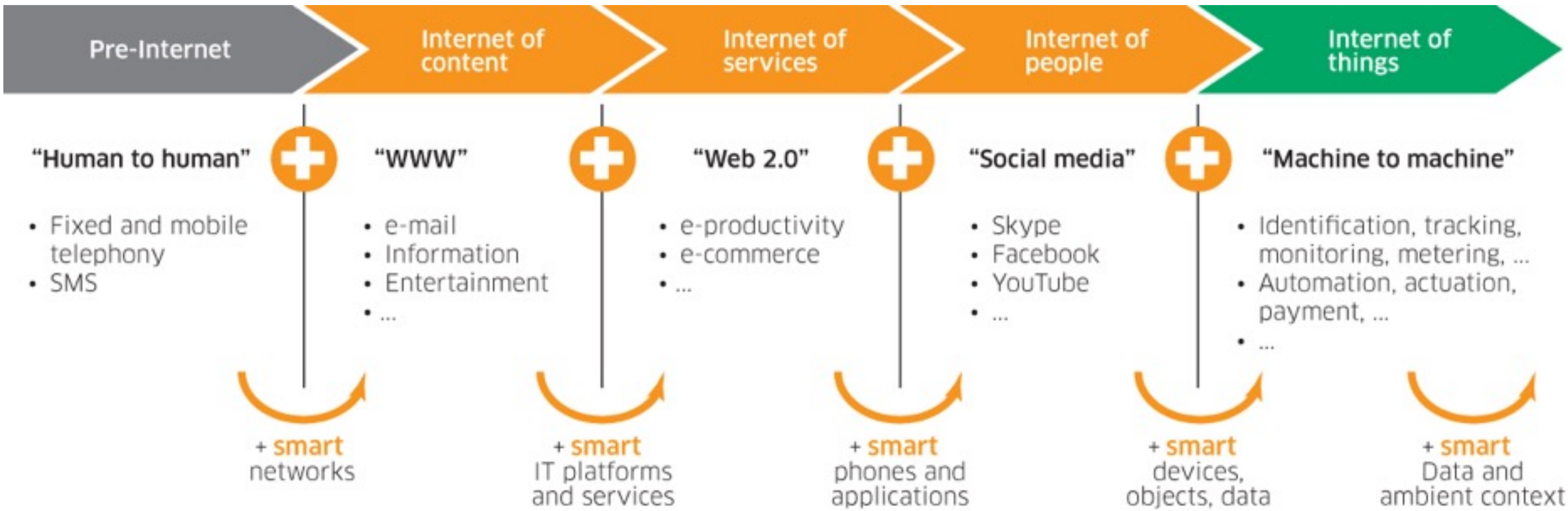
Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article



Source: https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article

Internet Evolution

Internet of People (IoP): Social Media
Internet of Things (IoT): Machine to Machine



Source: Marc Jadoul (2015), The IoT: The next step in internet evolution, March 11, 2015
http://www2.alcatel-lucent.com/techzine/iot-internet-of-things-next-step-evolution/

Social Media

Source: http://hungrywolfmarketing.com/2013/09/09/what-are-your-social-marketing-goals/

Emotions

Love

Joy

Surprise

Anger

Sadness

Fear

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Example of Opinion: review segment on iPhone

"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. … "

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Example of Opinion: review segment on iPhone

"(1) I bought an iPhone a few days ago.

(2) It was such a **nice** phone.

(3) The touch screen was really **cool**.

(4) The voice quality was **clear** too.

+Positive Opinion

(5) However, my mother was mad with me as I did not tell her before I bought it.

(6) She also thought the phone was too **expensive**, and wanted me to return it to the shop. … "

-Negative Opinion

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

How consumers
think, feel, and act

Emotions

Love

Joy

Surprise

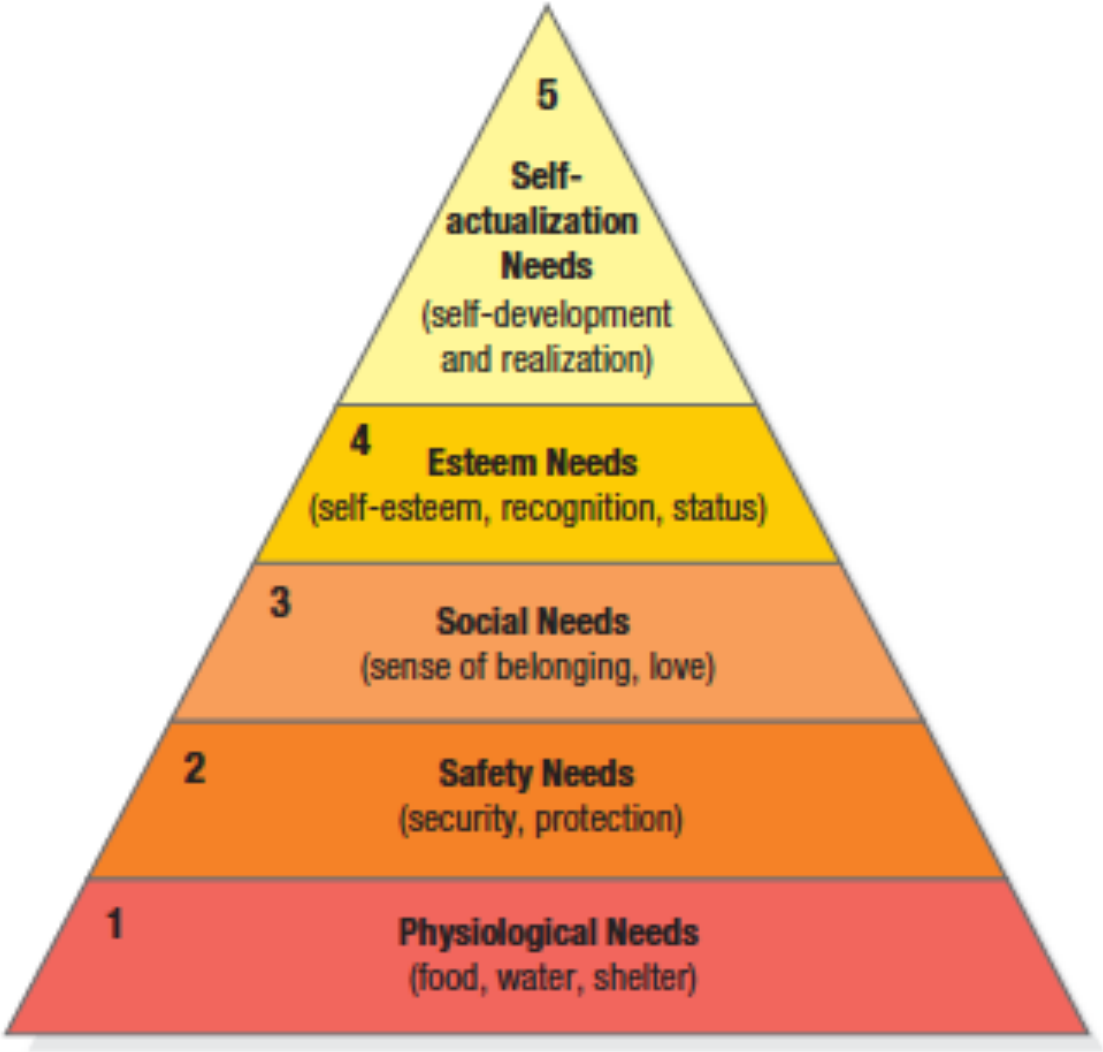
Anger

Sadness

Fear

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

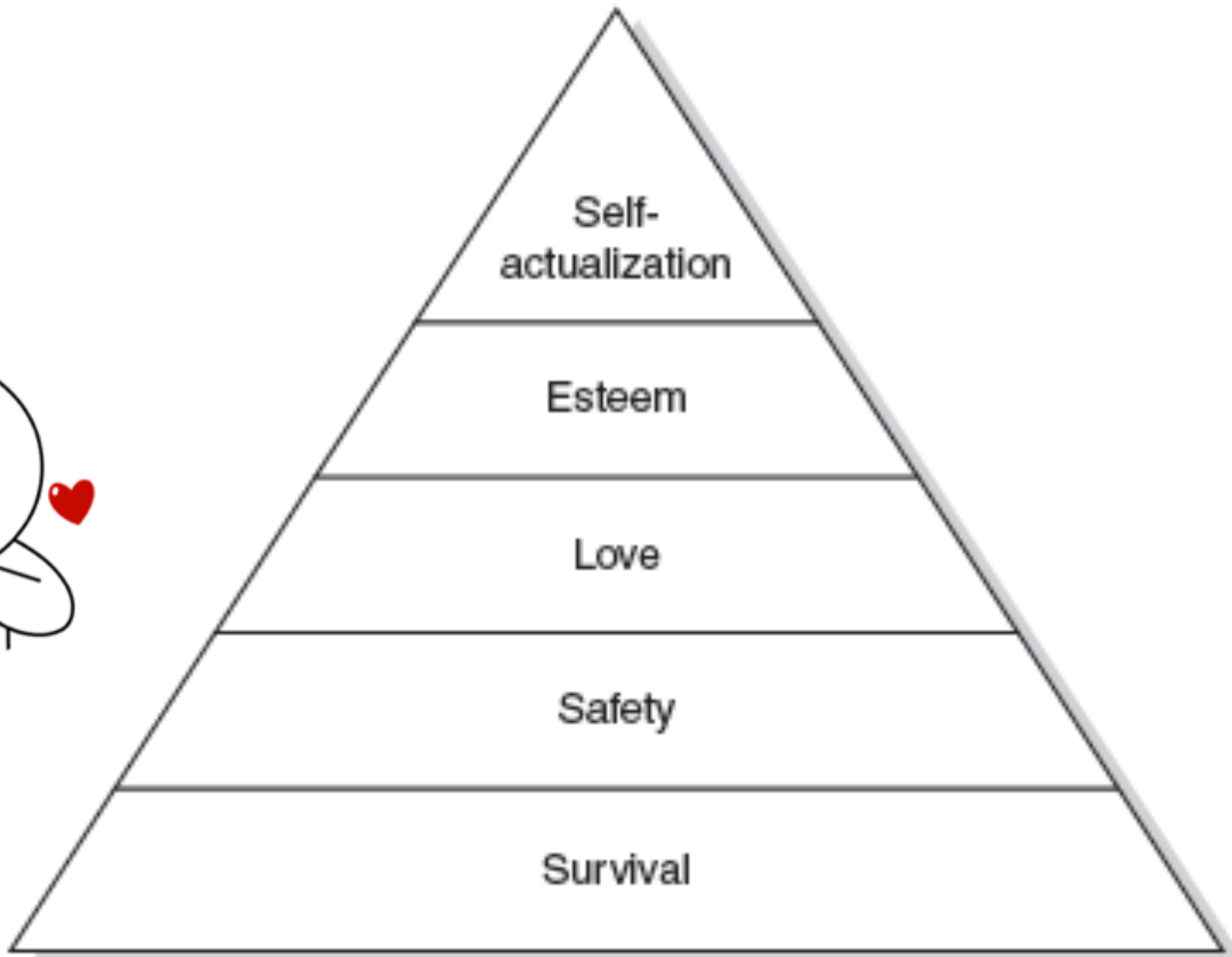
Maslow's Hierarchy of Needs



Source: Philip Kotler & Kevin Lane Keller, Marketing Management, 14th ed., Pearson, 2012

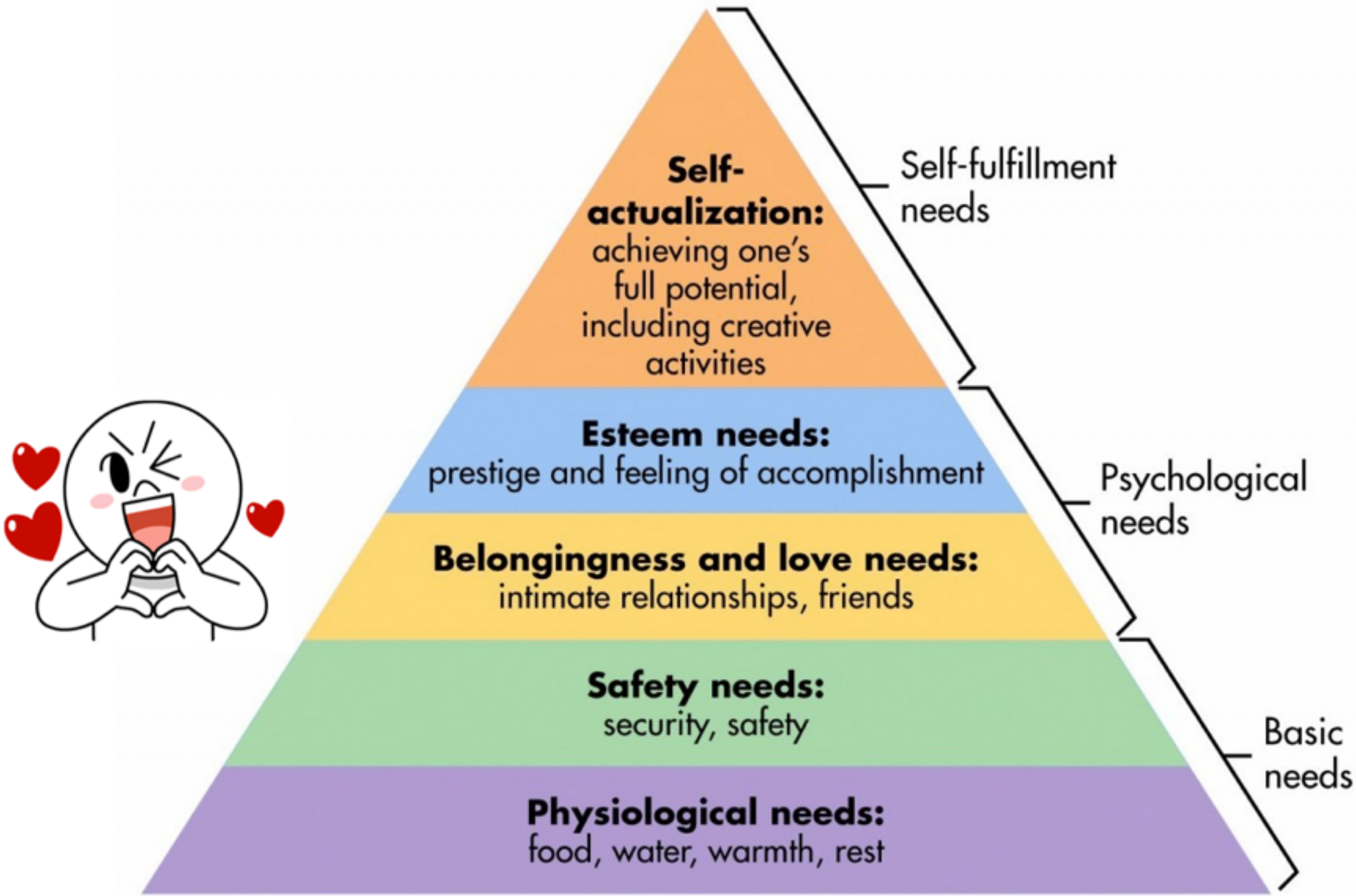
Maslow's hierarchy of human needs

(Maslow, 1943)



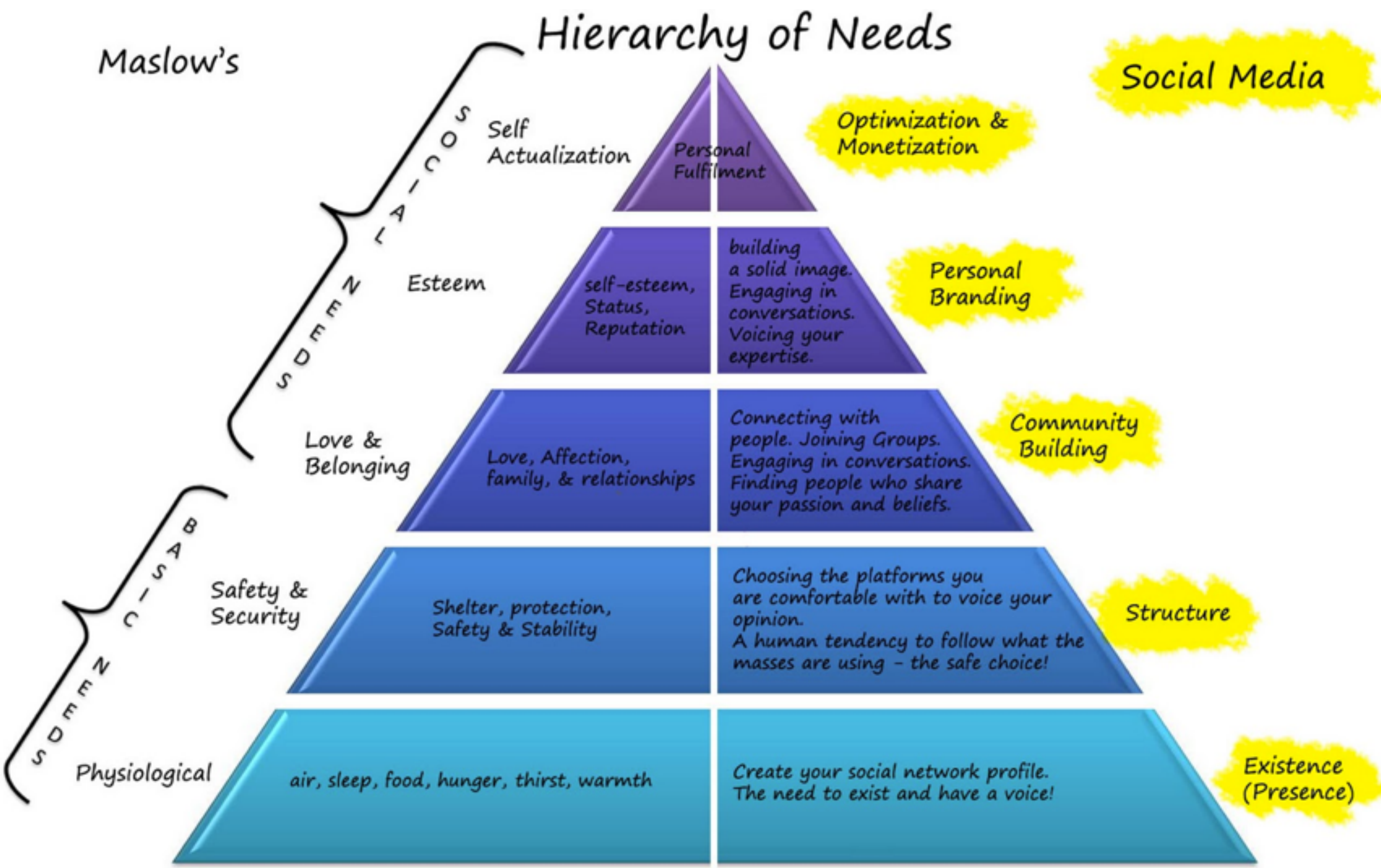
Source: Backer & Saren (2009), Marketing Theory: A Student Text, 2nd Edition, Sage

Maslow's Hierarchy of Needs



Source: http://sixstoriesup.com/social-psyche-what-makes-us-go-social/

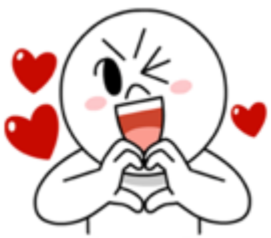
Social Media Hierarchy of Needs

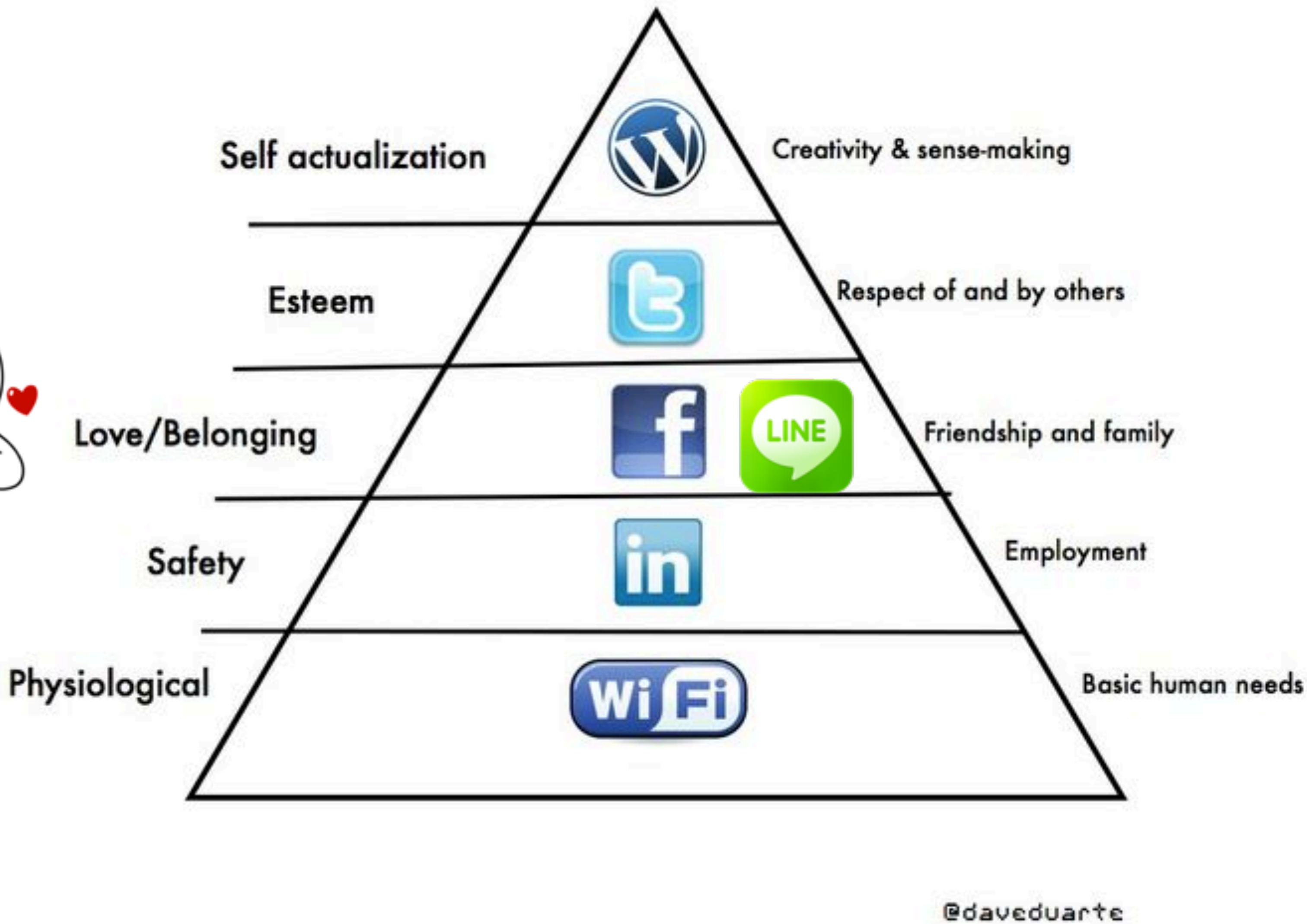


Social Media Hierarchy of Needs - by John Antonios

Source: http://2.bp.blogspot.com/_Rta1VZltiMk/TPavcanFtfI/AAAAAAAAACo/OBGnRL5arSU/s1600/social-media-heirarchy-of-needs1.jpg

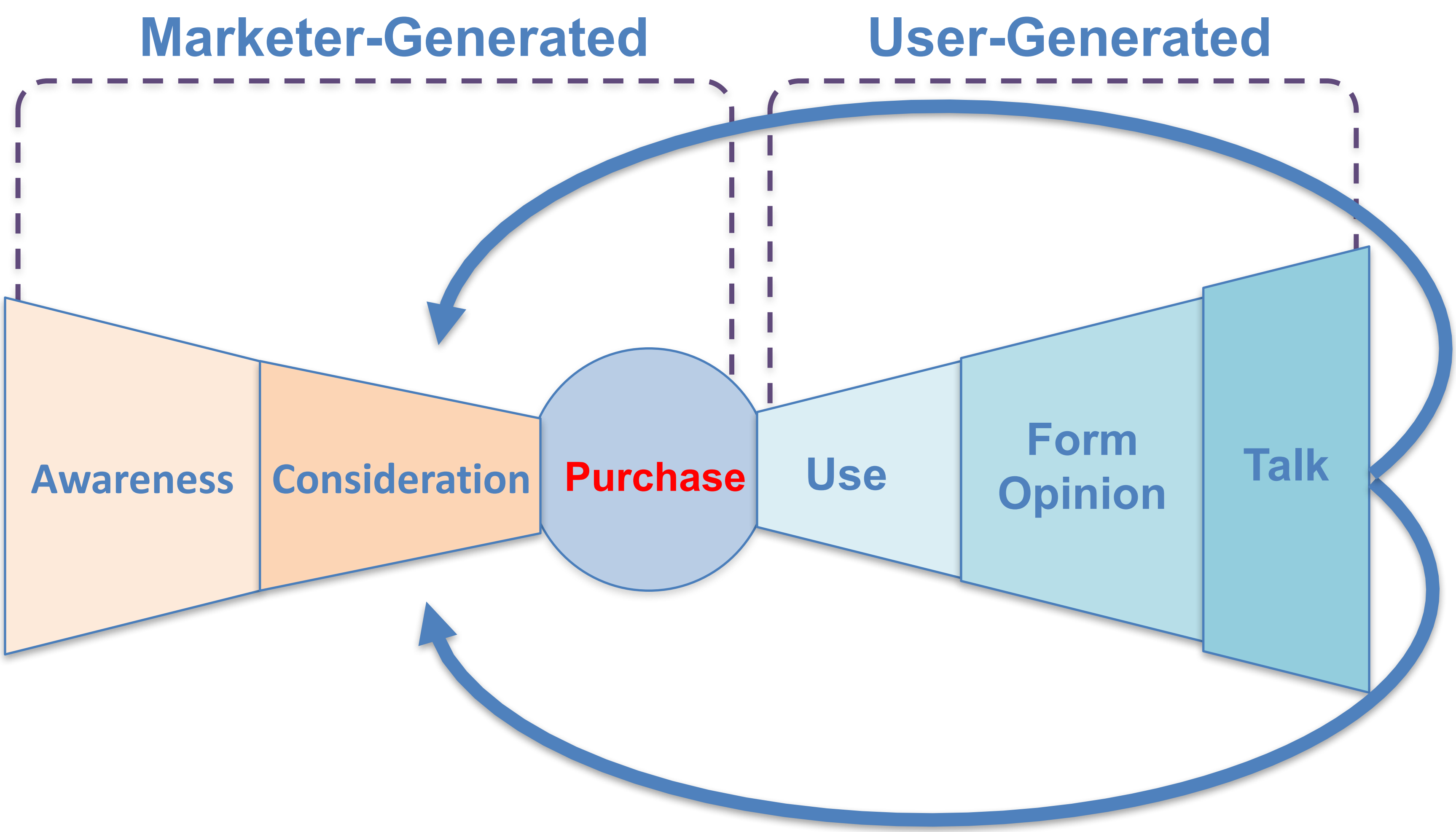
Social Media Hierarchy of Needs





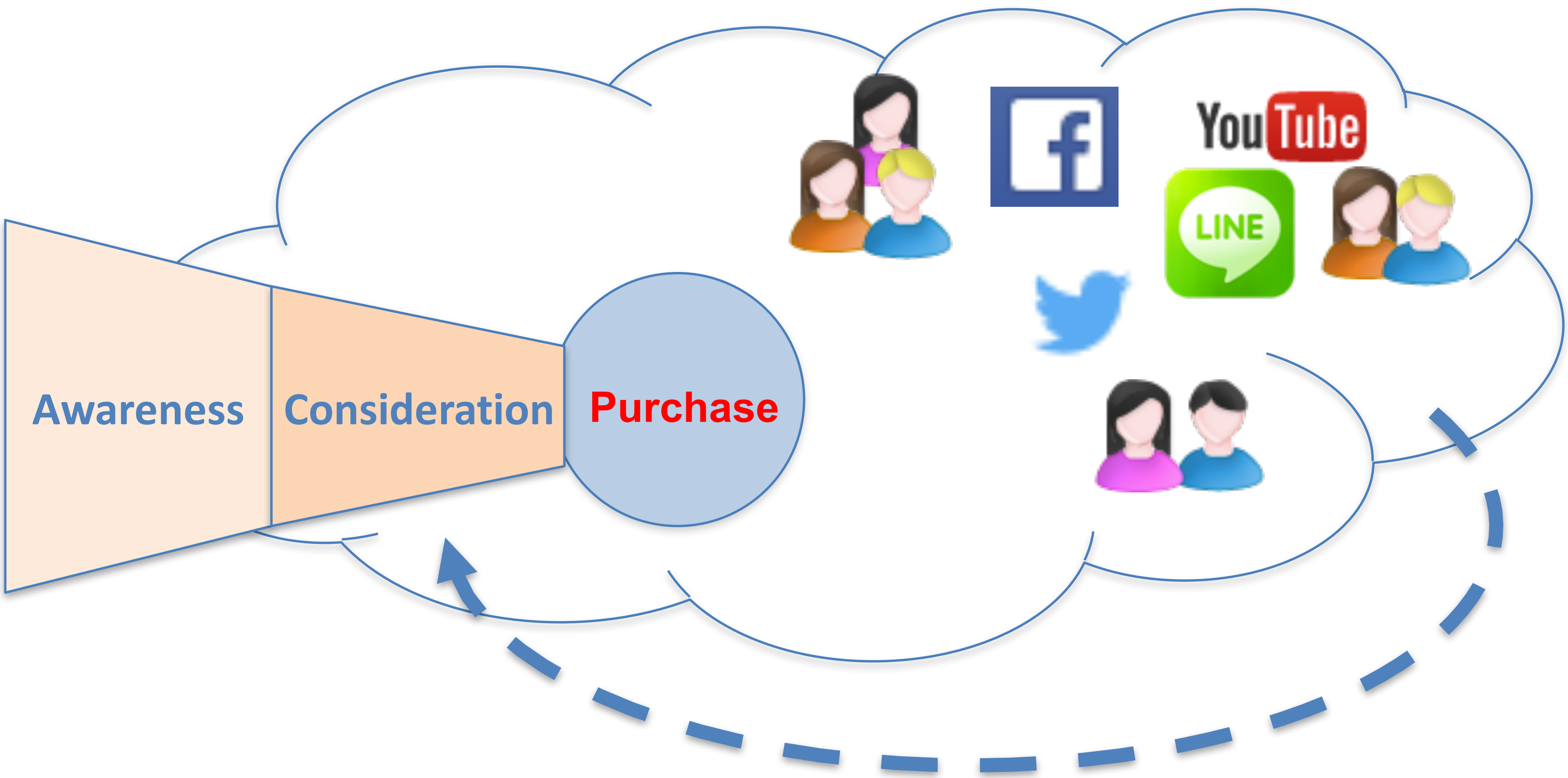
Source: http://www.pinterest.com/pin/186477859309035851/

The Social Feedback Cycle
Consumer Behavior on Social Media



Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement

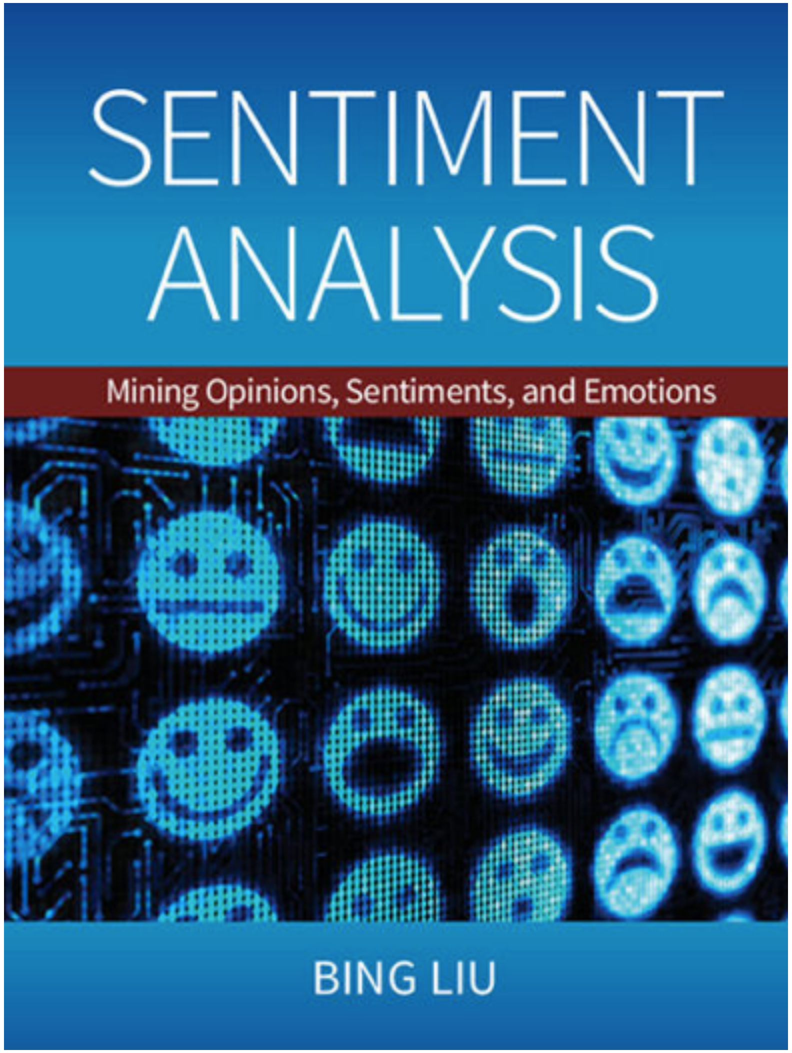
The New Customer Influence Path



Source: Evans et al. (2010), Social Media Marketing: The Next Generation of Business Engagement

Architectures of Sentiment Analytics

Bing Liu (2015),
Sentiment Analysis:
Mining Opinions, Sentiments, and Emotions,
Cambridge University Press



http://www.amazon.com/Sentiment-Analysis-Opinions-Sentiments-Emotions/dp/1107017890

Sentiment Analysis and Opinion Mining

- Computational study of
  opinions,
  sentiments,
  subjectivity,
  evaluations,
  attitudes,
  appraisal,
  affects,
  views,
  emotions,
  ets., expressed in text.
  - Reviews, blogs, discussions, news, comments, feedback, or any other documents

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Research Area of Opinion Mining

- Many names and tasks with difference objective and models
  - Sentiment analysis
  - Opinion mining
  - Sentiment mining
  - Subjectivity analysis
  - Affect analysis
  - Emotion detection
  - Opinion spam detection

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Sentiment Analysis

- Sentiment
  - A thought, view, or attitude, especially one based mainly on emotion instead of reason
- Sentiment Analysis
  - opinion mining
  - use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text

Applications of Sentiment Analysis

- Consumer information
  - Product reviews
- Marketing
  - Consumer attitudes
  - Trends
- Politics
  - Politicians want to know voters’ views
  - Voters want to know policitians’ stances and who else supports them
- Social
  - Find like-minded individuals or communities

Sentiment detection

- How to interpret features for sentiment detection?
  - Bag of words (IR)
  - Annotated lexicons (WordNet, SentiWordNet)
  - Syntactic patterns
- Which features to use?
  - Words (unigrams)
  - Phrases/n-grams
  - Sentences

Problem statement of Opinion Mining

- Two aspects of abstraction
  - Opinion definition
    - What is an opinion?
    - What is the structured definition of opinion?
  - Opinion summarization
    - Opinion are subjective
      - An opinion from a single person (unless a VIP) is often not sufficient for action
    - We need opinions from many people, and thus opinion summarization.

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

What is an opinion?

- Id: **Abc123** on **5-1-2008** *"I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …"*
- One can look at this review/blog at the
  - Document level
    - Is this review + or -?
  - Sentence level
    - Is each sentence + or -?
  - Entity and feature/aspect level

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Entity and aspect/feature level

- Id: **Abc123** on **5-1-2008** "*I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive, …*"
- What do we see?
  - Opinion targets: entities and their features/aspects
  - Sentiments: positive and negative
  - Opinion holders: persons who hold the opinions
  - Time: when opinion are expressed

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Two main types of opinions

- Regular opinions: Sentiment/Opinion expressions on some target entities
  - Direct opinions: sentiment expressions on one object:
    - "The touch screen is really cool."
    - "The picture quality of this camera is great"
  - Indirect opinions: comparisons, relations expressing similarities or differences (objective or subjective) of more than one object
    - "phone X is cheaper than phone Y." (objective)
    - "phone X is better than phone Y." (subjective)
- Comparative opinions: comparisons of more than one entity.
  - "iPhone is better than Blackberry."

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Subjective and Objective

- Objective
  - An objective sentence expresses some factual information about the world.
  - "I returned the phone yesterday."
  - Objective sentences can implicitly indicate opinions
    - "The earphone broke in two days."
- Subjective
  - A subjective sentence expresses some personal feelings or beliefs.
  - "The voice on my phone was not so clear"
  - Not every subjective sentence contains an opinion
    - "I wanted a phone with good voice quality"
- → Subjective analysis

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Sentiment Analysis vs. Subjectivity Analysis

Sentiment Analysis	Subjectivity Analysis
Positive Negative	Subjective
Neutral	Objective

A (regular) opinion

- Opinion (a restricted definition)
  – An opinion (regular opinion) is simply a positive or negative sentiment, view, attitude, emotion, or appraisal about an entity or an aspect of the entity from an opinion holder.
- Sentiment orientation of an opinion
  – Positive, negative, or neutral (no opinion)
  – Also called:
    - Opinion orientation
    - Semantic orientation
    - Sentiment polarity

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Entity and aspect

- Definition of Entity:
  - An *entity e* is a product, person, event, organization, or topic.
  - e is represented as
    - A hierarchy of components, sub-components.
    - Each node represents a components and is associated with a set of attributes of the components
- An opinion can be expressed on any node or attribute of the node
- Aspects(features)
  - represent both components and attribute

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Opinion Definition

- An opinion is a quintuple
  (e_j, a_{jk}, so_{ijkl}, h_i, t_l)
  where
  - e_j is a target entity.
  - a_{jk} is an aspect/feature of the entity e_j .
  - so_{ijkl} is the sentiment value of the opinion from the opinion holder on feature of entity at time.
    so_{ijkl} is +ve, -ve, or neu, or more granular ratings
  - h_i is an opinion holder.
  - t_l is the time when the opinion is expressed.
- (e_j, a_{jk}) is also called opinion target

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Terminologies

- Entity: object
- Aspect: feature, attribute, facet
- Opinion holder: opinion source

- Topic: entity, aspect

- Product features, political issues

Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Subjectivity and Emotion

- Sentence subjectivity
  - An objective sentence presents some factual information, while a subjective sentence expresses some personal feelings, views, emotions, or beliefs.
- Emotion
  - Emotions are people’s subjective feelings and thoughts.

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Classification Based on Supervised Learning

- Sentiment classification
  - Supervised learning Problem
  - Three classes
    - *Positive*
    - *Negative*
    - *Neutral*

Source: Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” Springer, 2nd Edition,

Opinion words in Sentiment classification

- topic-based classification
  - topic-related words are important
    - e.g., *politics, sciences, sports*
- Sentiment classification
  - topic-related words are unimportant
  - **opinion words** (also called **sentiment words)**
    - **that indicate positive or negative opinions** are important,
      e.g., *great, excellent, amazing, horrible, bad, worst*

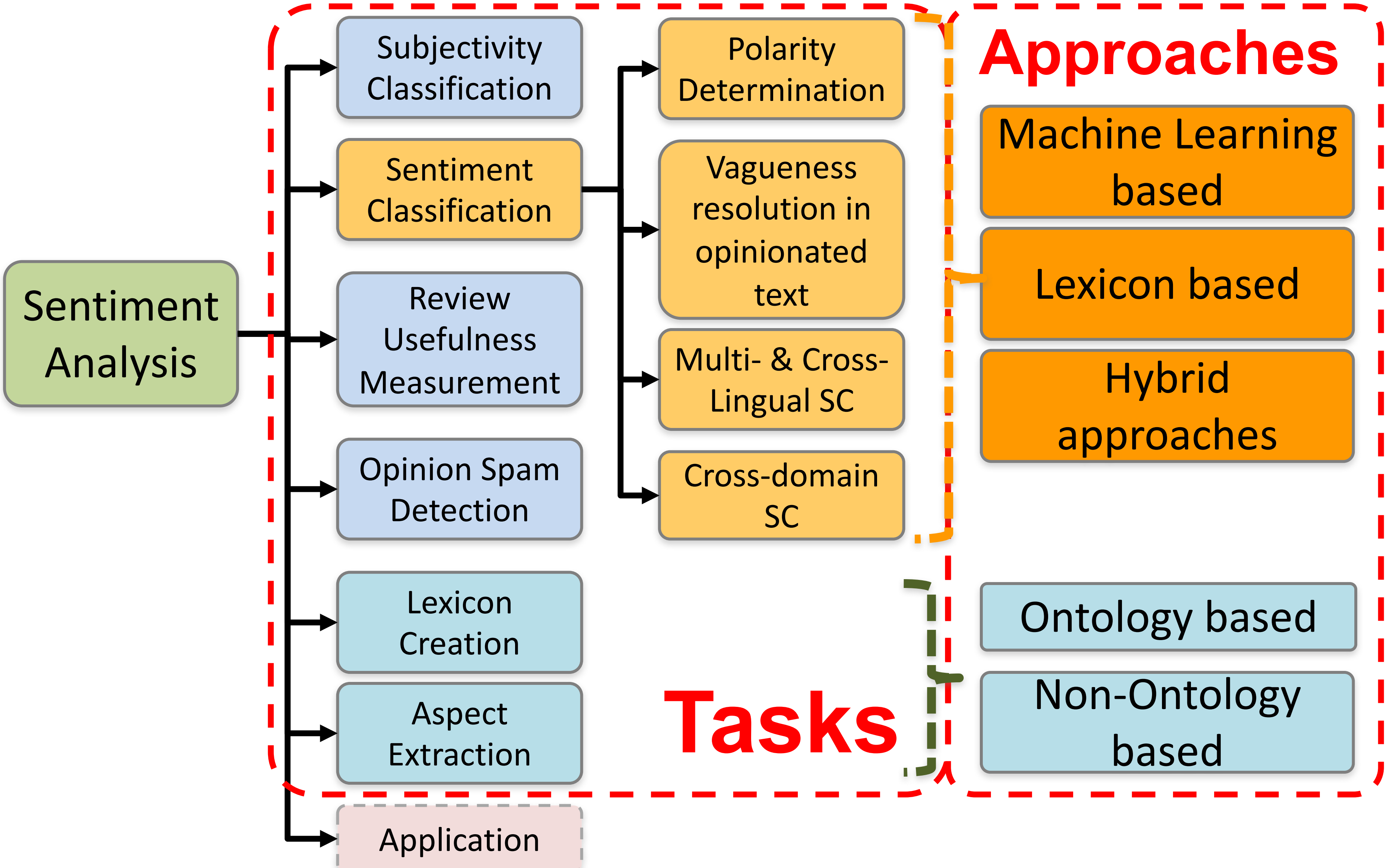
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Features in Opinion Mining

- *Terms and their frequency*
  - *TF-IDF*
- *Part of speech (POS)*
  - Adjectives
- *Opinion words and phrases*
  - *beautiful, wonderful, good, and amazing are positive opinion words*
  - *bad, poor, and terrible are negative opinion words.*
  - opinion phrases and idioms,
    e.g., *cost someone an arm and a leg*
- *Rules of opinions*
- *Negations*
- *Syntactic dependency*

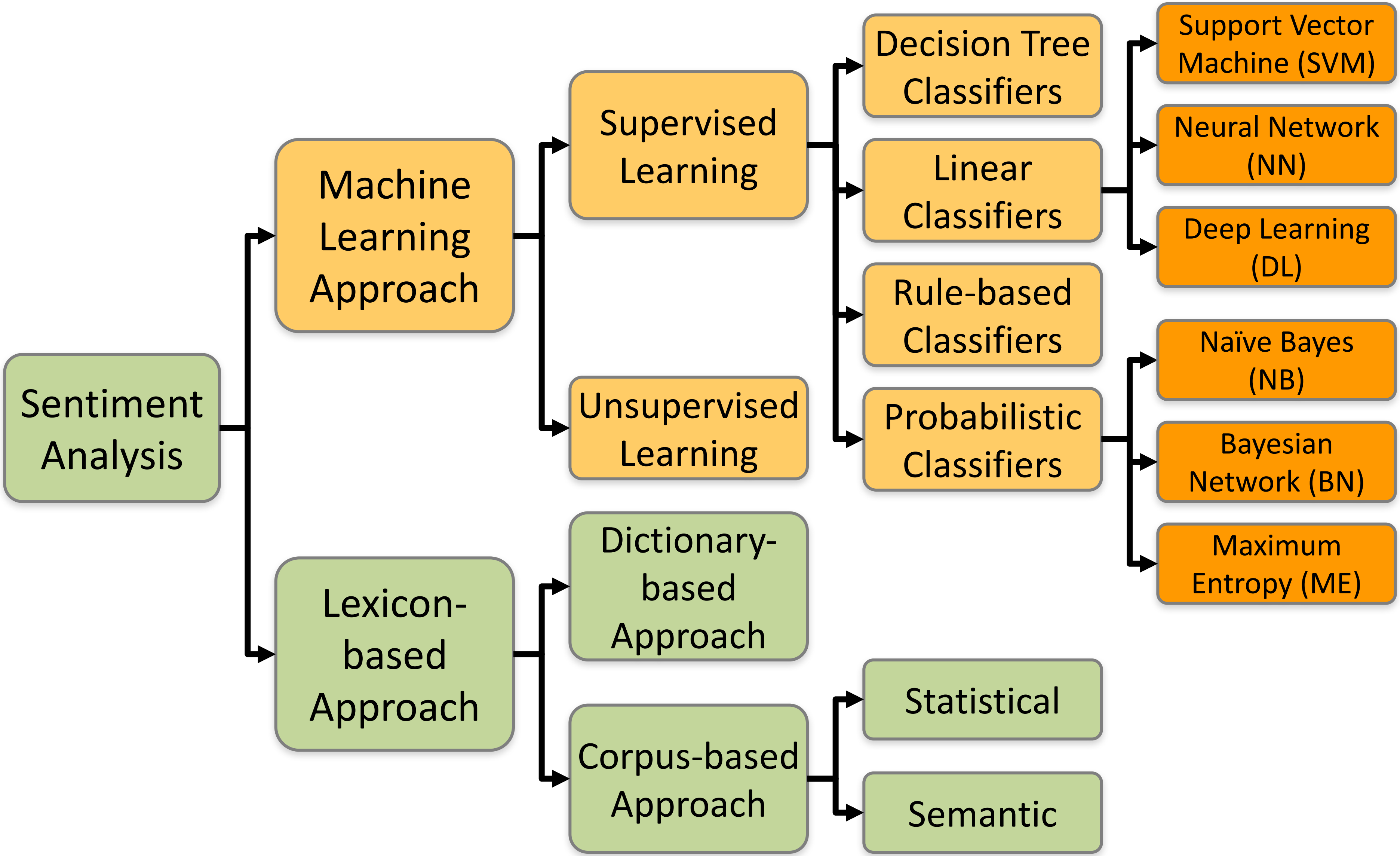
Source: Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition,

Sentiment Analysis



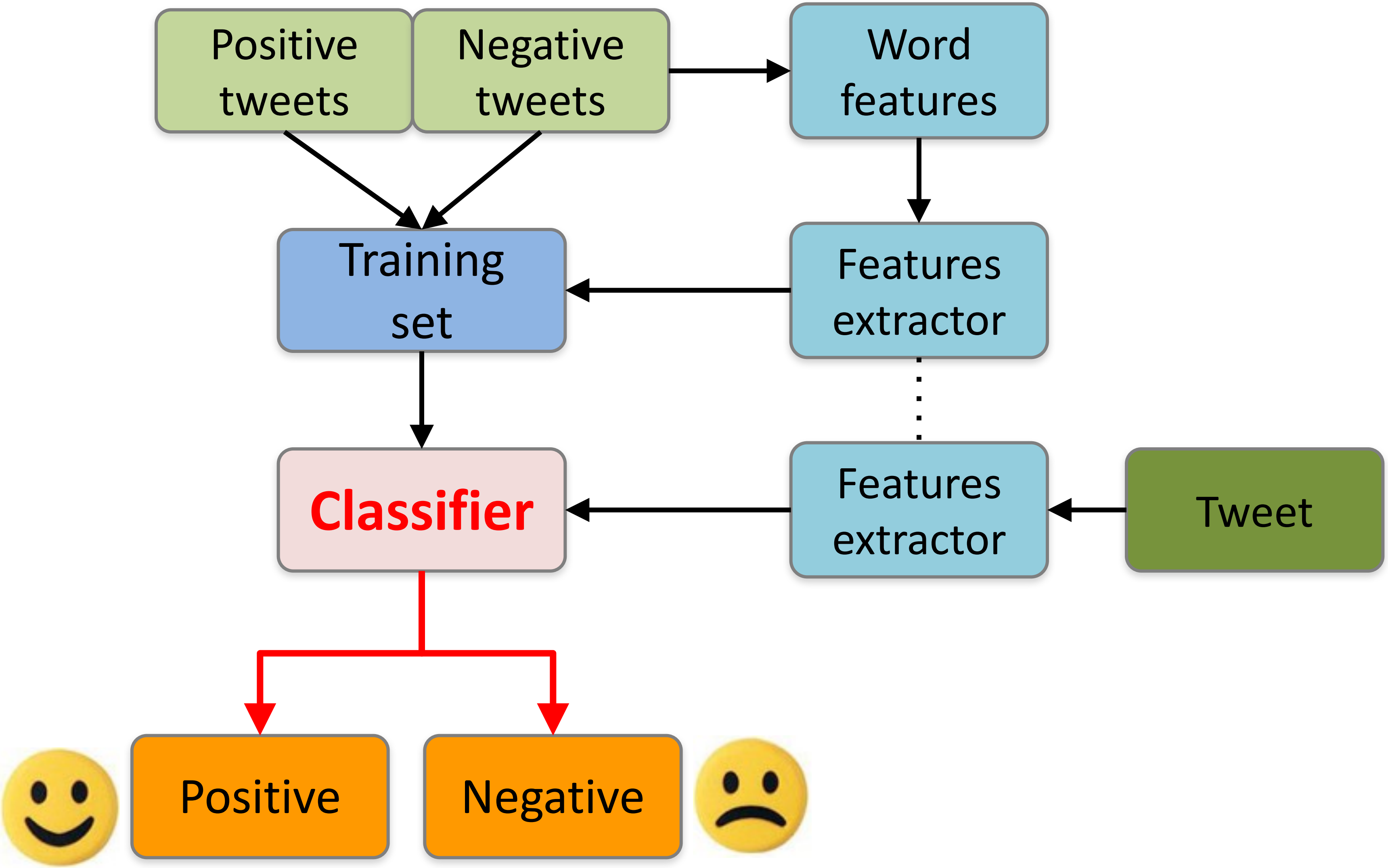
Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Classification Techniques



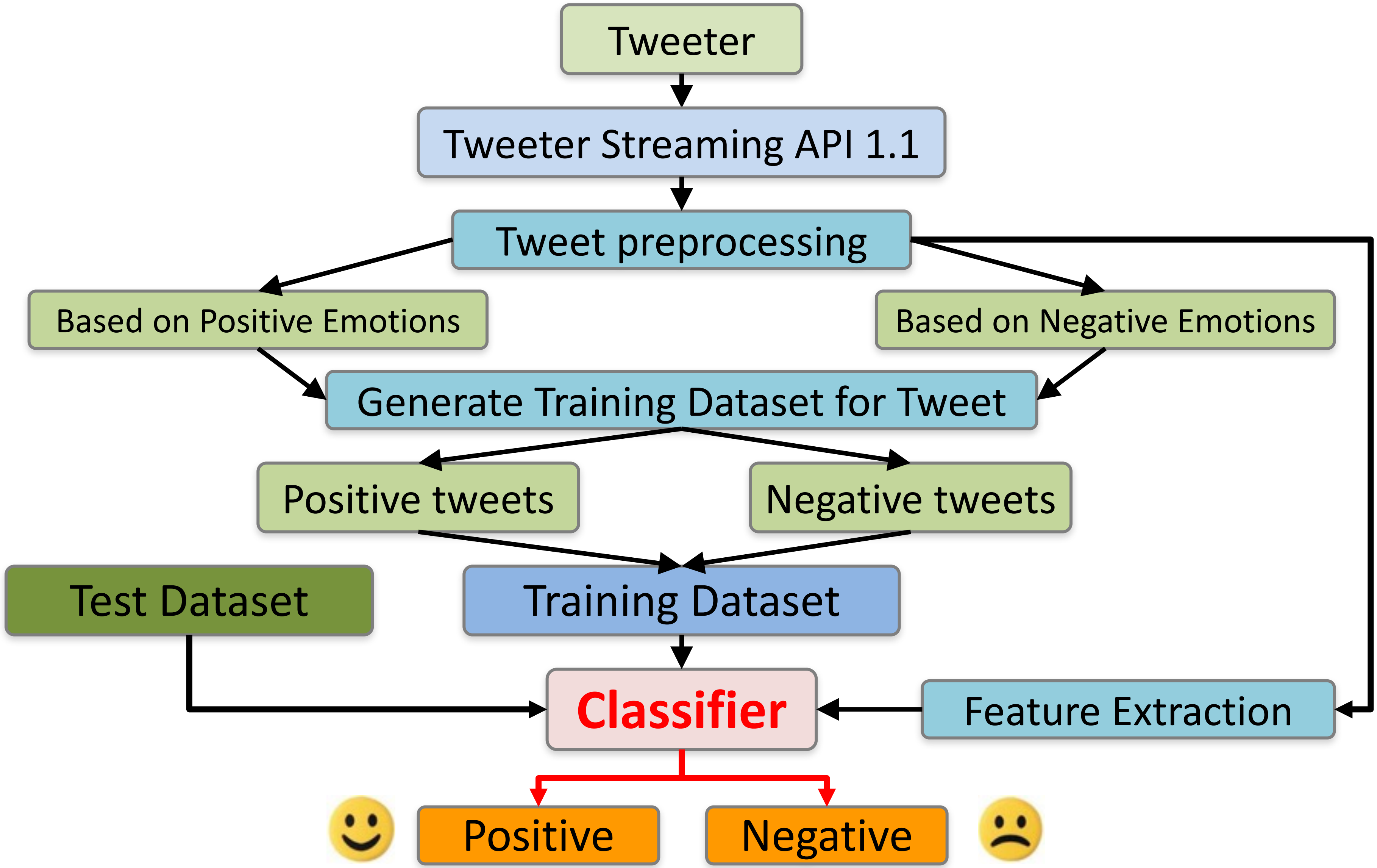
Source: Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.

Sentiment Analysis Architecture



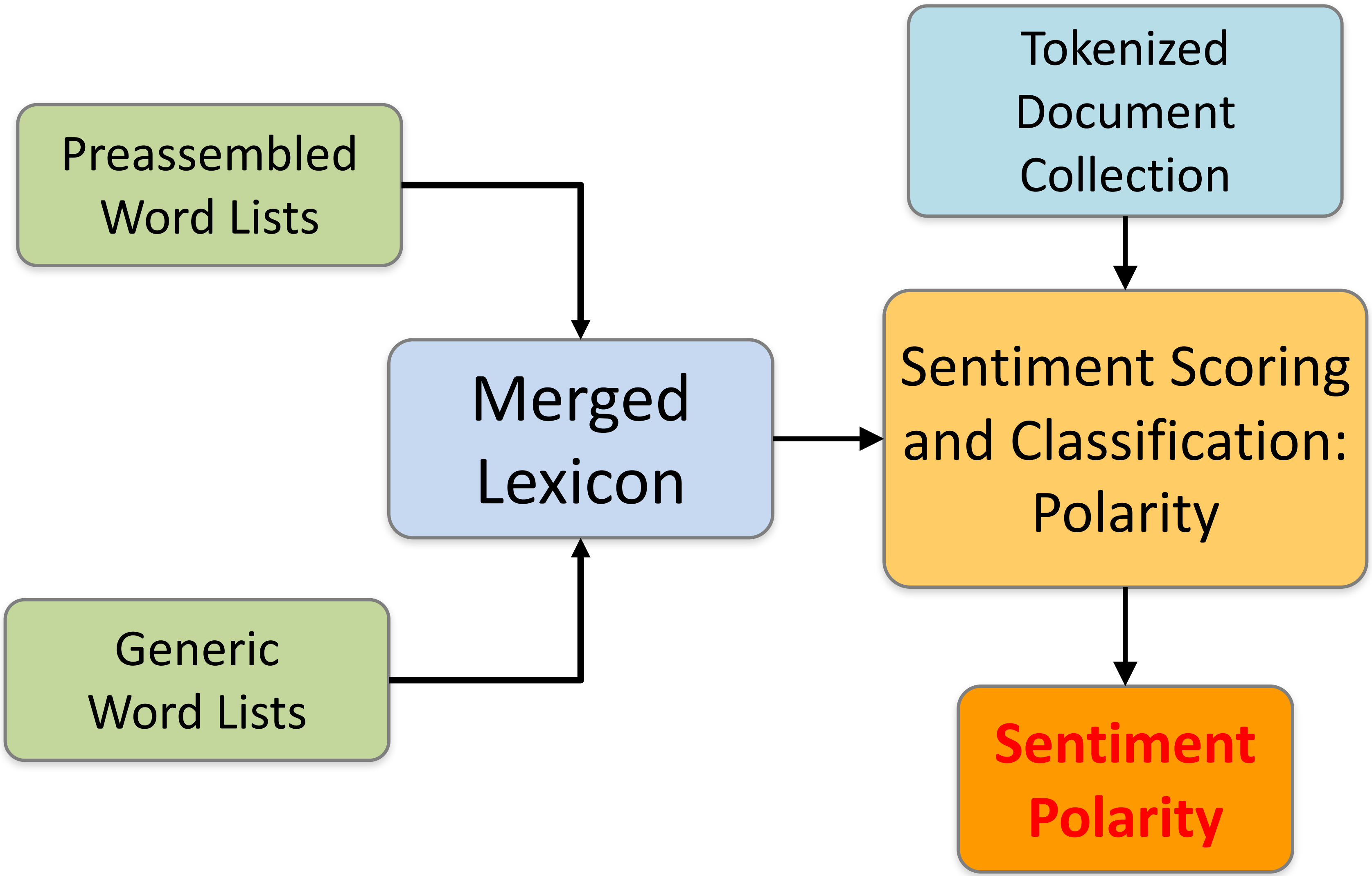
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Classification Based on Emoticons



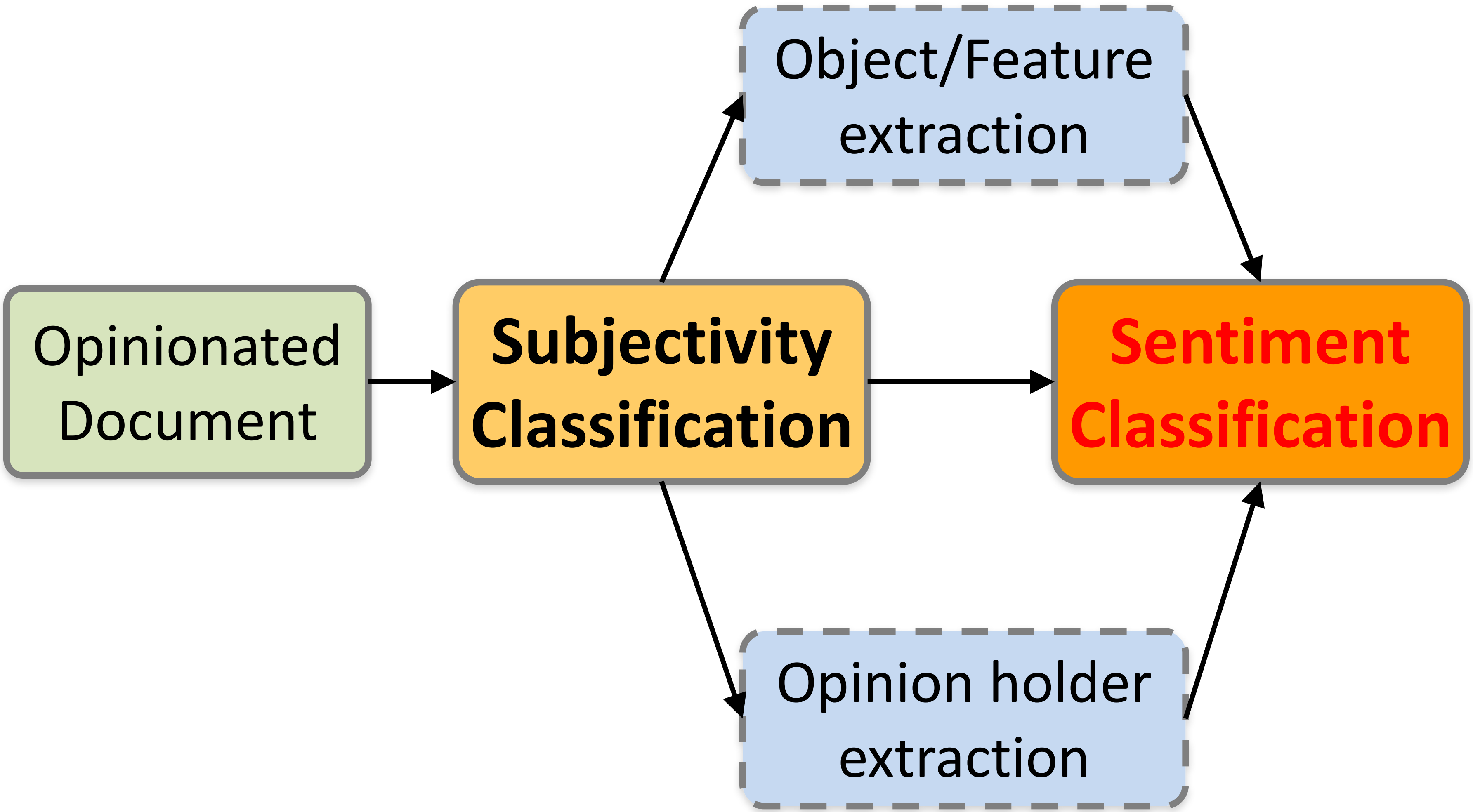
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Lexicon-Based Model



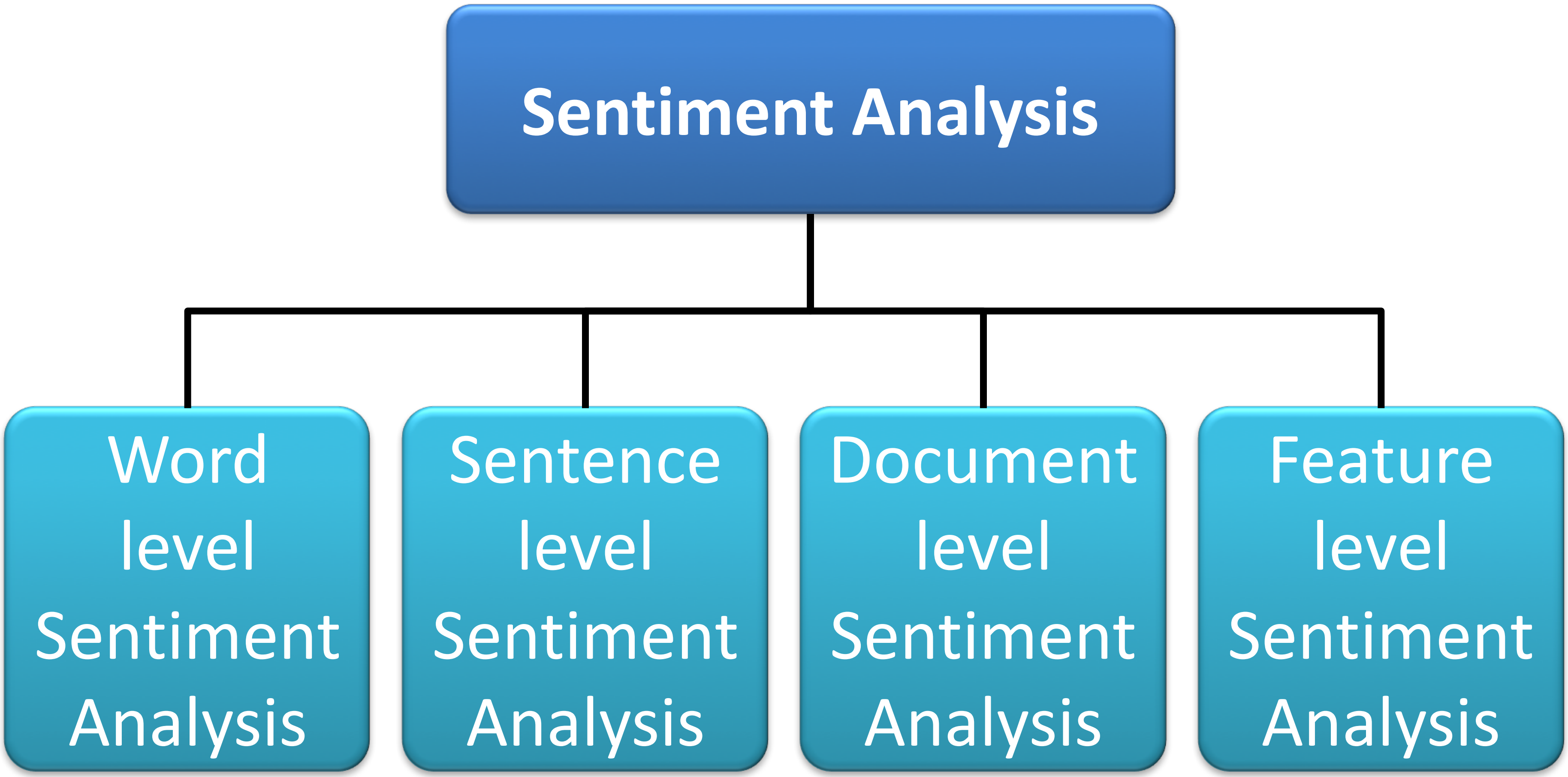
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Sentiment Analysis Tasks



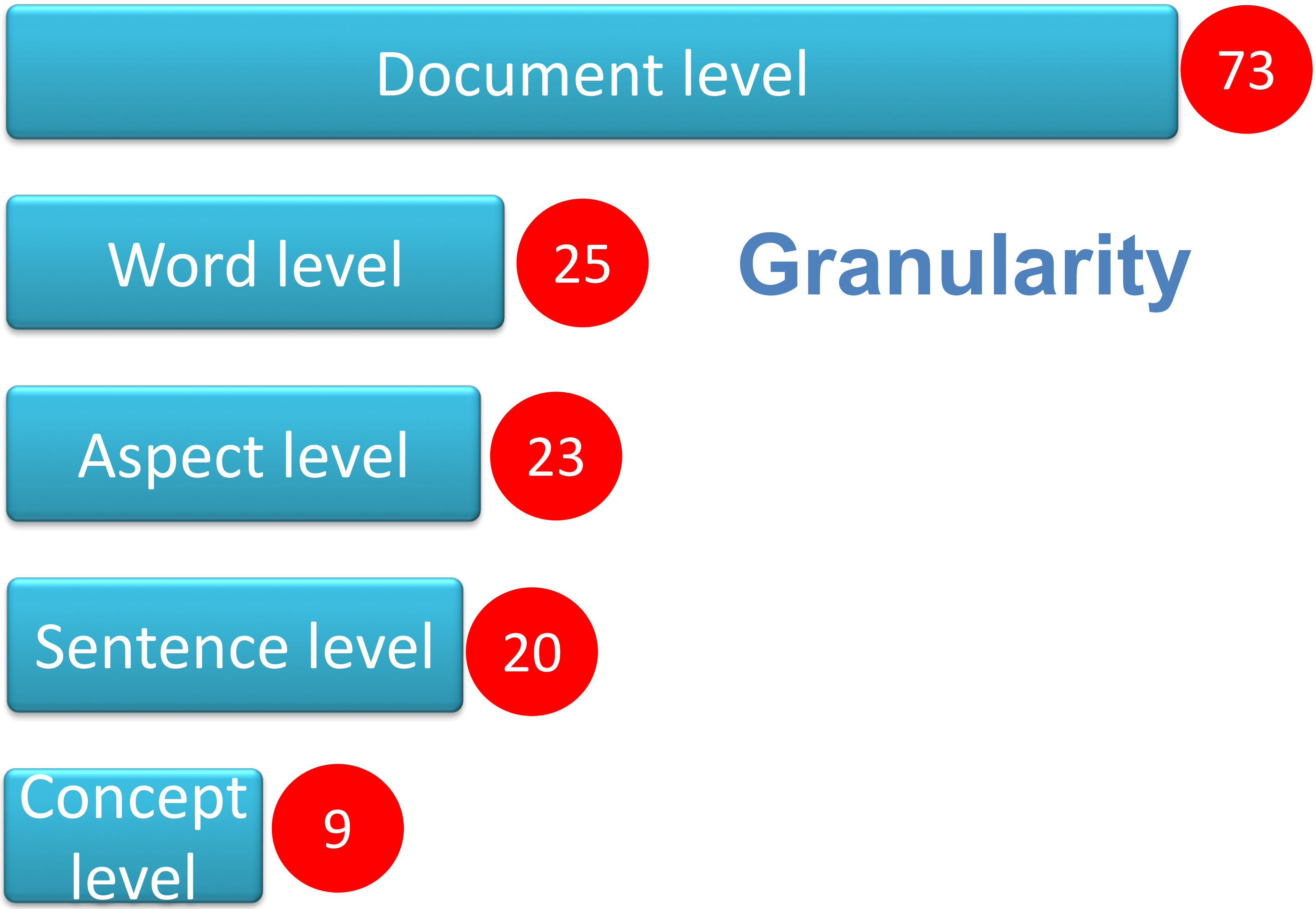
Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Levels of Sentiment Analysis



Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Levels of Sentiment Analysis



Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

A Brief Summary of Sentiment Analysis Methods

Study	Analysis Task	Sentiment Identification		Sentiment Aggregation		Nature of Measure
		Method	Level	Method	Level	
Hu and Li, 2011	Polarity	ML (Probabilistic model)	Snippet			Valence
Li and Wu, 2010	Polarity	Lexicon/Rule	Phrase	Sum	Snippet	Valence
Thelwall et al., 2010	Polarity	Lexicon/Rule	Sentence	Max & Min	Snippet	Range
Boiy and Moens, 2009	Both	ML (Cascade ensemble)	Sentence			Valence
Chung 2009	Polarity	Lexicon	Phrase	Average	Sentence	Valence
Wilson, Wiebe, and Hoffmann, 2009	Both	ML (SVM, AdaBoost, Rule, etc.)	Phrase			Valence
Zhang et al., 2009	Polarity	Lexicon/Rule	Sentence	Weighted average	Snippet	Valence
Abbasi, Chen, and Salem, 2008	Polarity	ML (GA + feature selection)	Snippet			Valence
Subrahmanian and Reforgiato, 2008	Polarity	Lexicon/Rule	Phrase	Rule	Snippet	Valence
Tan and Zhang 2008	Polarity	ML (SVM, Winnow, NB, etc.)	Snippet			Valence
Airoldi, Bai, and Padman, 2007	Polarity	ML (Markov Blanket)	Snippet			Valence
Das and Chen, 2007	Polarity	ML (Bayesian, Discriminate, etc.)	Snippet	Average	Daily	Valence
Liu et al., 2007	Polarity	ML (PLSA)	Snippet			Valence
Kennedy and Inkpen, 2006	Polarity	Lexicon/Rule, ML (SVM)	Phrase	Count	Snippet	Valence
Mishne 2006	Polarity	Lexicon	Phrase	Average	Snippet	Valence
Liu et al., 2005	Polarity	Lexicon/Rule	Phrase	Distribution	Object	Range
Mishne 2005	Polarity	ML (SVM)	Snippet			Valence
Popescu and Etzioni 2005	Polarity	Lexicon/Rule	Phrase			Valence
Efron 2004	Polarity	ML (SVN, NB)	Snippet			Valence
Wilson, Wiebe, and Hwa, 2004	Both	ML (SVM, AdaBoost, Rule, etc.)	Sentence			Valence
Nigam and Hurst 2004	Polarity	Lexicon/Rule	Chunk	Rule	Sentence	Valence
Dave, Lawrence, and Pennock, 2003	Polarity	ML (SVM, Rainbow, etc.)	Snippet			Valence
Nasukawa and Yi 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yi et al., 2003	Polarity	Lexicon/Rule	Phrase	Rule	Sentence	Valence
Yu and Hatzivassiloglou 2003	Both	ML (NB) + Lexicon/Rule	Phrase	Average	Sentence	Valence
Pang, Lee, and Vaithyanathan 2002	Polarity	ML (SVM, MaxEnt, NB)	Snippet			Valence
Subasic and Huettner 2001	Polarity	Lexicon/Fuzzy logic	Phrase	Average	Snippet	Valence
Turney 2001	Polarity	Lexicon/Rule	Phrase	Average	Snippet	Valence

(Both = Subjectivity and Polarity; ML= Machine Learning; Lexicon/Rule= Lexicon enhanced by linguistic rules)

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Word-of-Mouth (WOM)

- "This book is the best written documentary thus far, yet sadly, there is no soft cover edition."

- "This book is the best written documentary thus far, yet sadly, there is no soft cover edition."

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

This
book
is
the
best
written
documentary
thus
far
,
yet
sadly
,
there
is
no
soft
cover
edition
.

Word	POS
This	DT
book	NN
is	VBZ
the	DT
best	JJS
written	VBN
documentary	NN
thus	RB
far	RB
,	,
yet	RB
sadly	RB
,	,
there	EX
is	VBZ
no	DT
soft	JJ
cover	NN
edition	NN
.	.

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Conversion of text representation

SentiWordNet Lookup → Microstate Mapping → Probability Mapping

Word Vector (WV)	Polarity Score Vector (PSV) *pscore*	*nscore*		Microstate Sequence (MS)
This	0	0	Neutral (0)	0
book	0	0	Neutral (0)	0
is	0	0	Neutral (0)	0
the	0	0	Neutral (0)	0
best	0.75	0	Positive (0.75)	1
written	0	0	Neutral (0)	0
documentary	0	0	Neutral (0)	0
thus	0.375	0	Positive (0.375)	1
far	0.375	0	Positive (0.375)	1
,				
yet	0	0.125	Negative (0.125)	-1
sadly	0.25	0.5	Negative (0.25)	-1
,				
there	0	0	Neutral (0)	0
is	0	0	Neutral (0)	0
no	0	0.75	Negative (0.75)	-1
soft	0	0	Neutral (0)	0
cover	0	0	Neutral (0)	0
edition	0	0	Neutral (0)	0
.				

Probability Distribution (P)

P("1")=3/17

P("-1")=3/17

P("0")=11/17

Source: Zhang, Z., Li, X., and Chen, Y. (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.,

Example of SentiWordNet

POS ID PosScore NegScore SynsetTerms Gloss

a 00217728 0.75 0 beautiful#1 delighting the senses or exciting intellectual or emotional admiration; "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful theory"; "a beautiful party“

a 00227507 0.75 0 best#1 (superlative of `good') having the most positive qualities; "the best film of the year"; "the best solution"; "the best time for planting"; "wore his best suit“

r 00042614 0 0.625 unhappily#2 sadly#1 in an unfortunate way; "sadly he died before he could see his grandchild“

r 00093270 0 0.875 woefully#1 sadly#3 lamentably#1 deplorably#1 in an unfortunate or deplorable manner; "he was sadly neglected"; "it was woefully inadequate“

r 00404501 0 0.25 sadly#2 with sadness; in a sad manner; "`She died last night,' he said sadly"

SenticNet

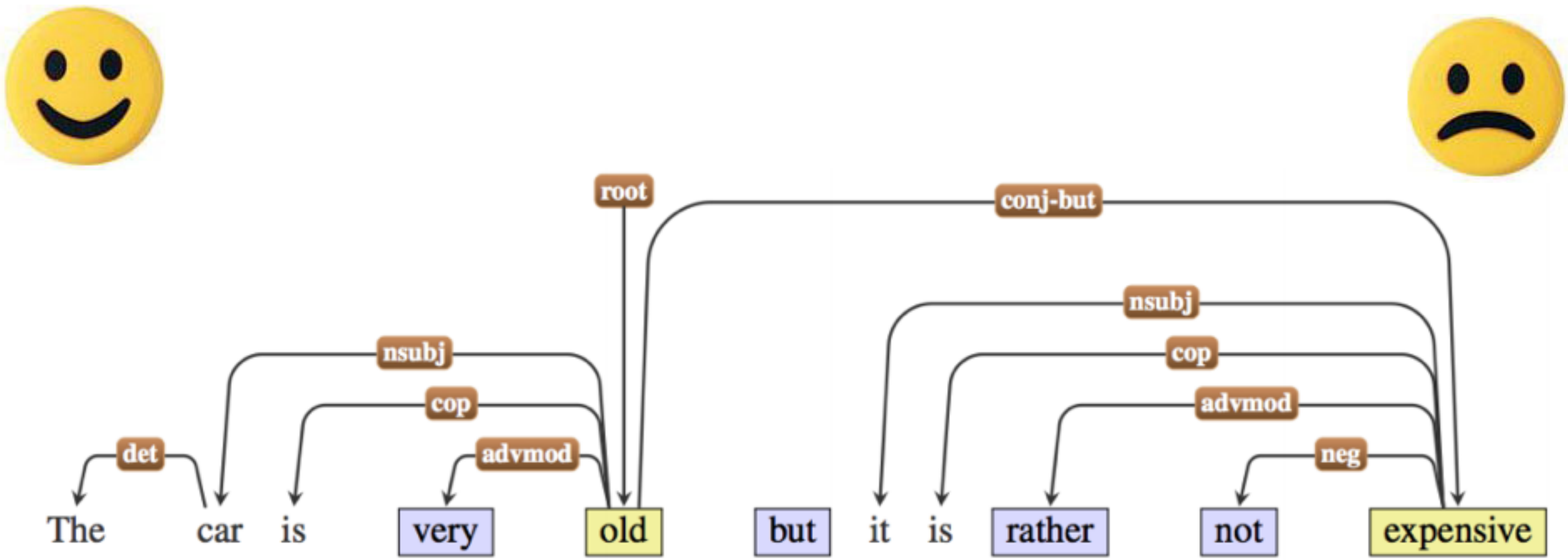
The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet

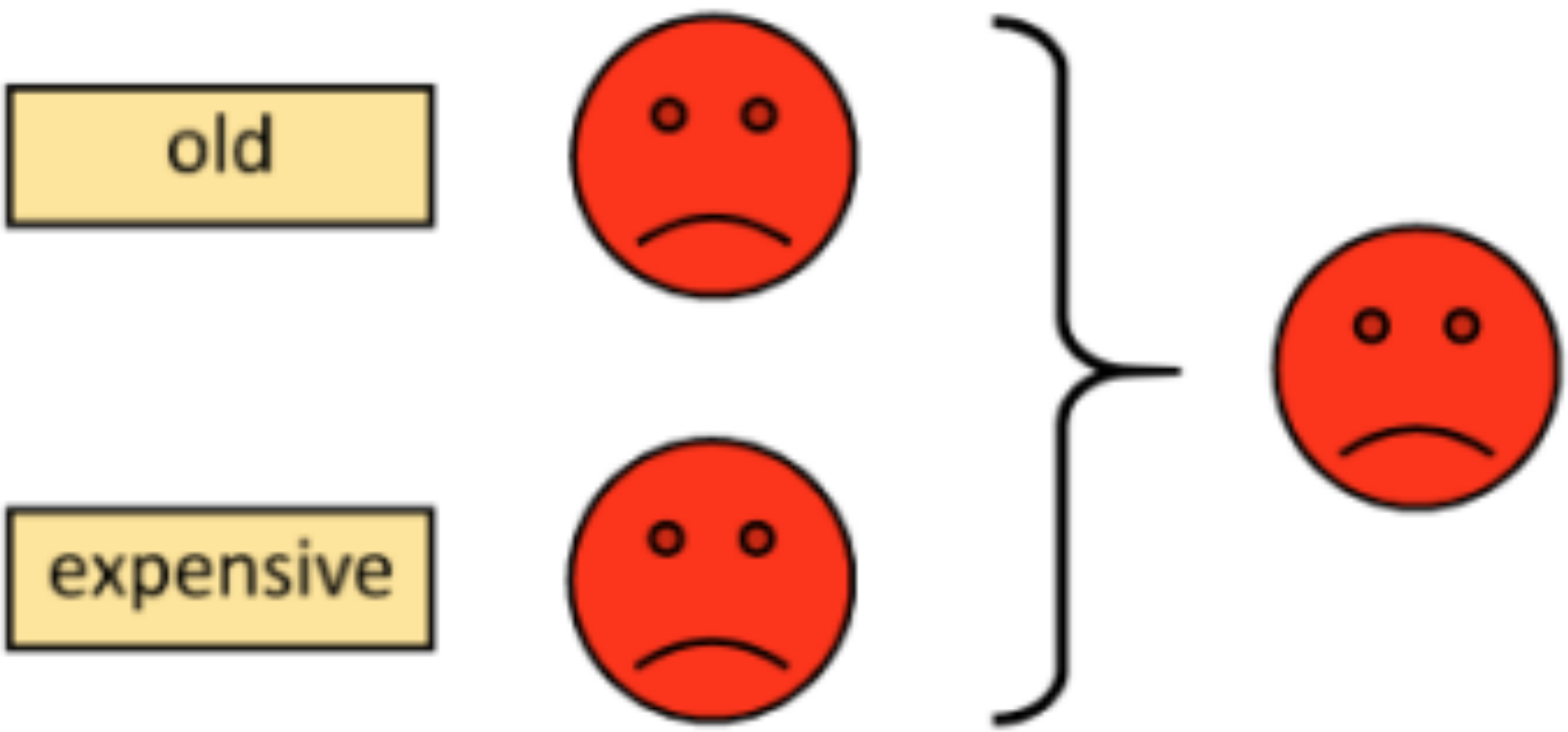


The car is very old but it is rather not expensive.

The car is very old but it is rather not expensive.

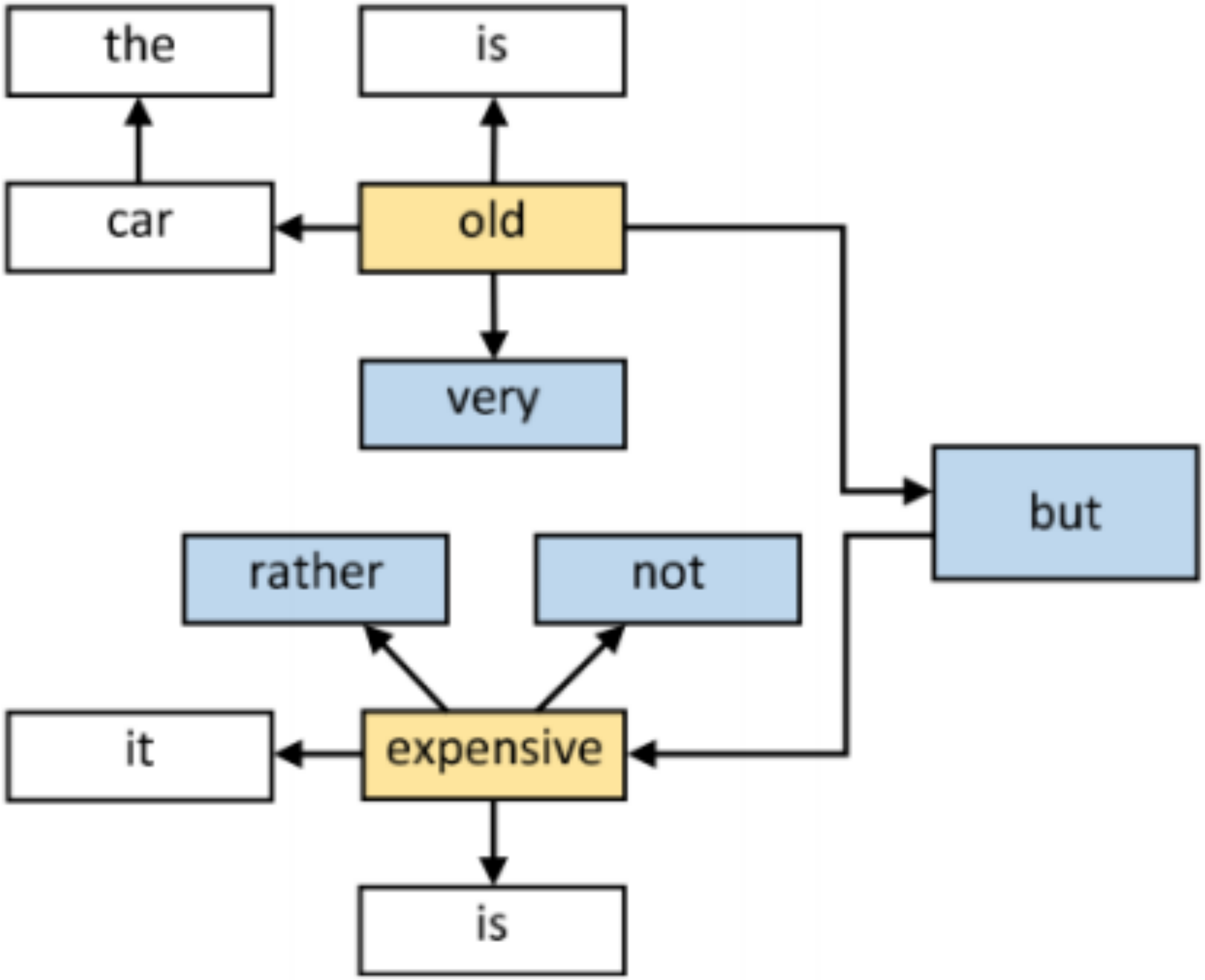
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



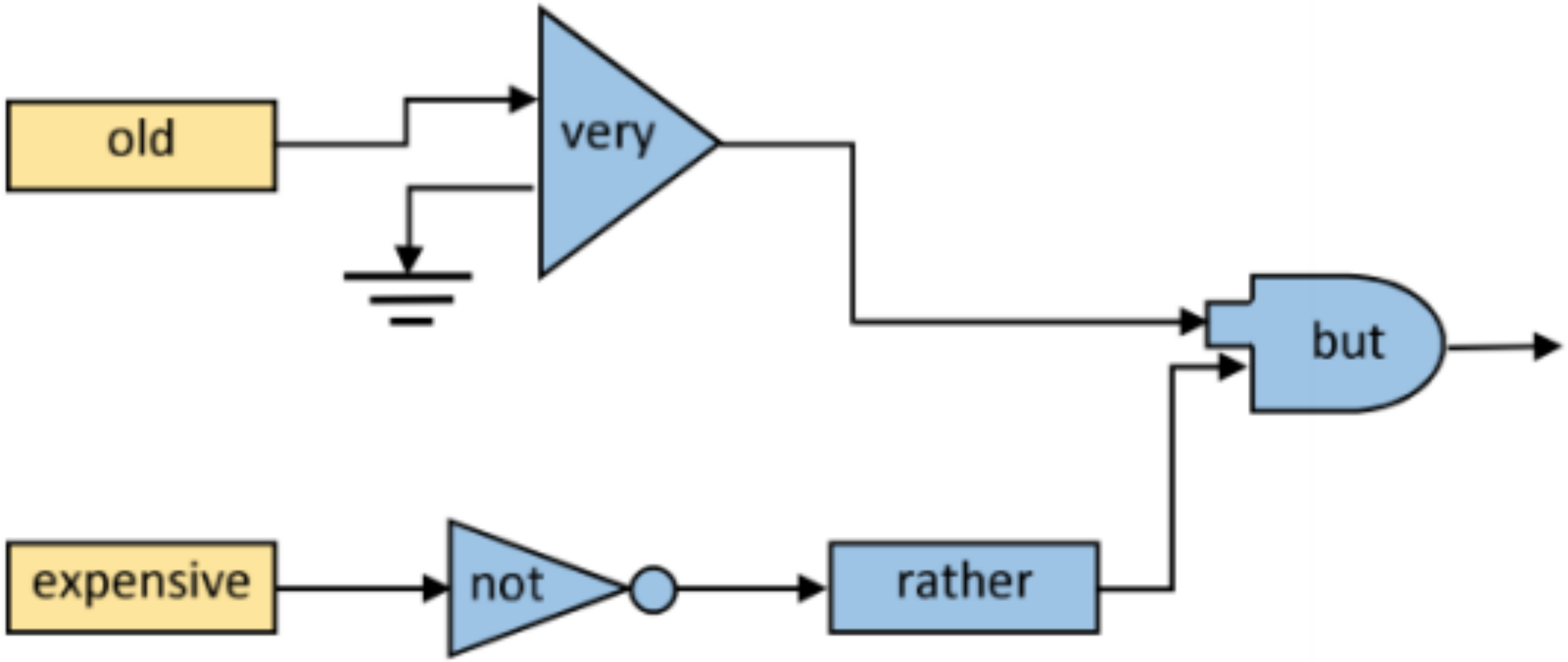
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



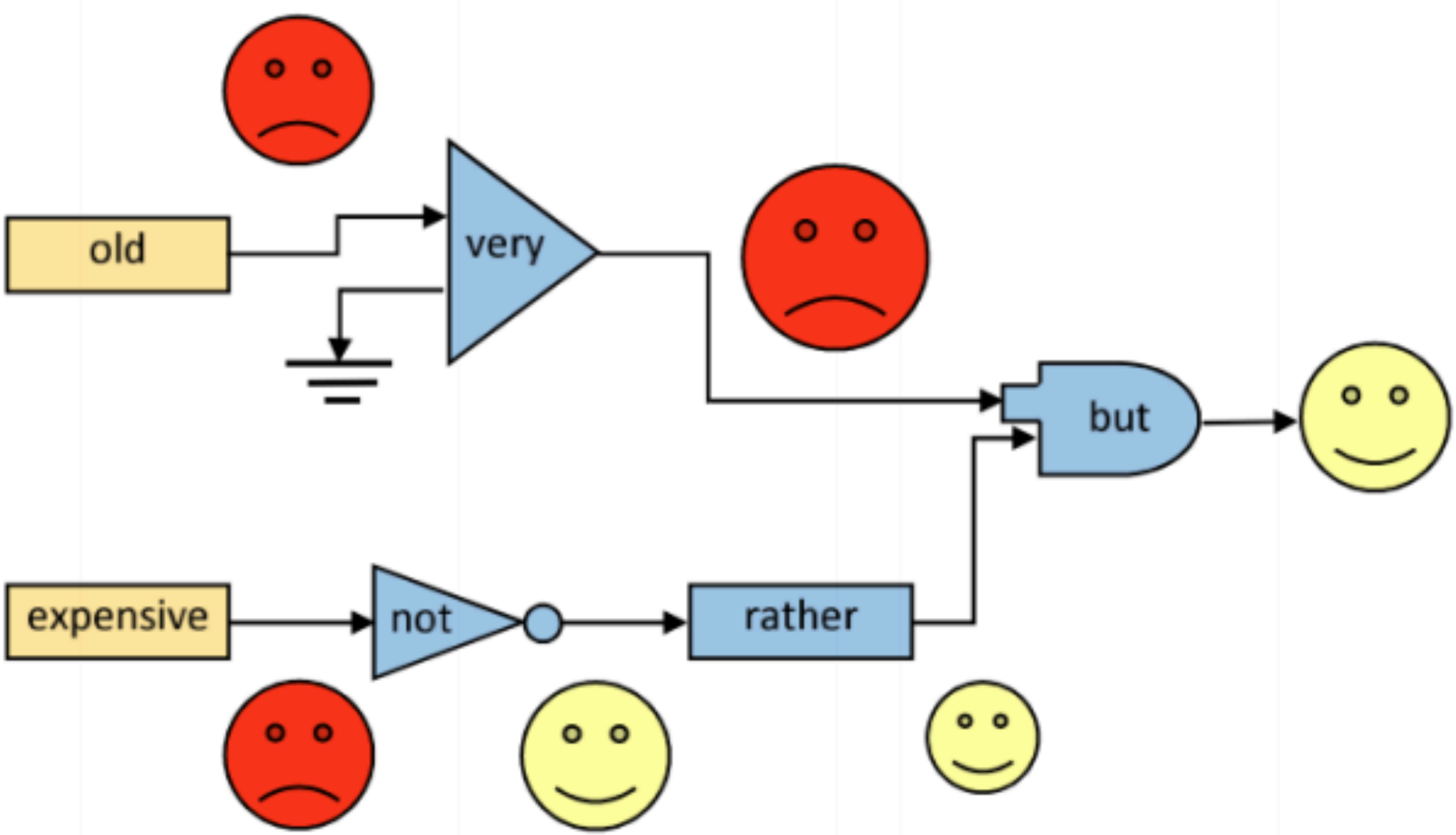
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Polarity Detection with SenticNet



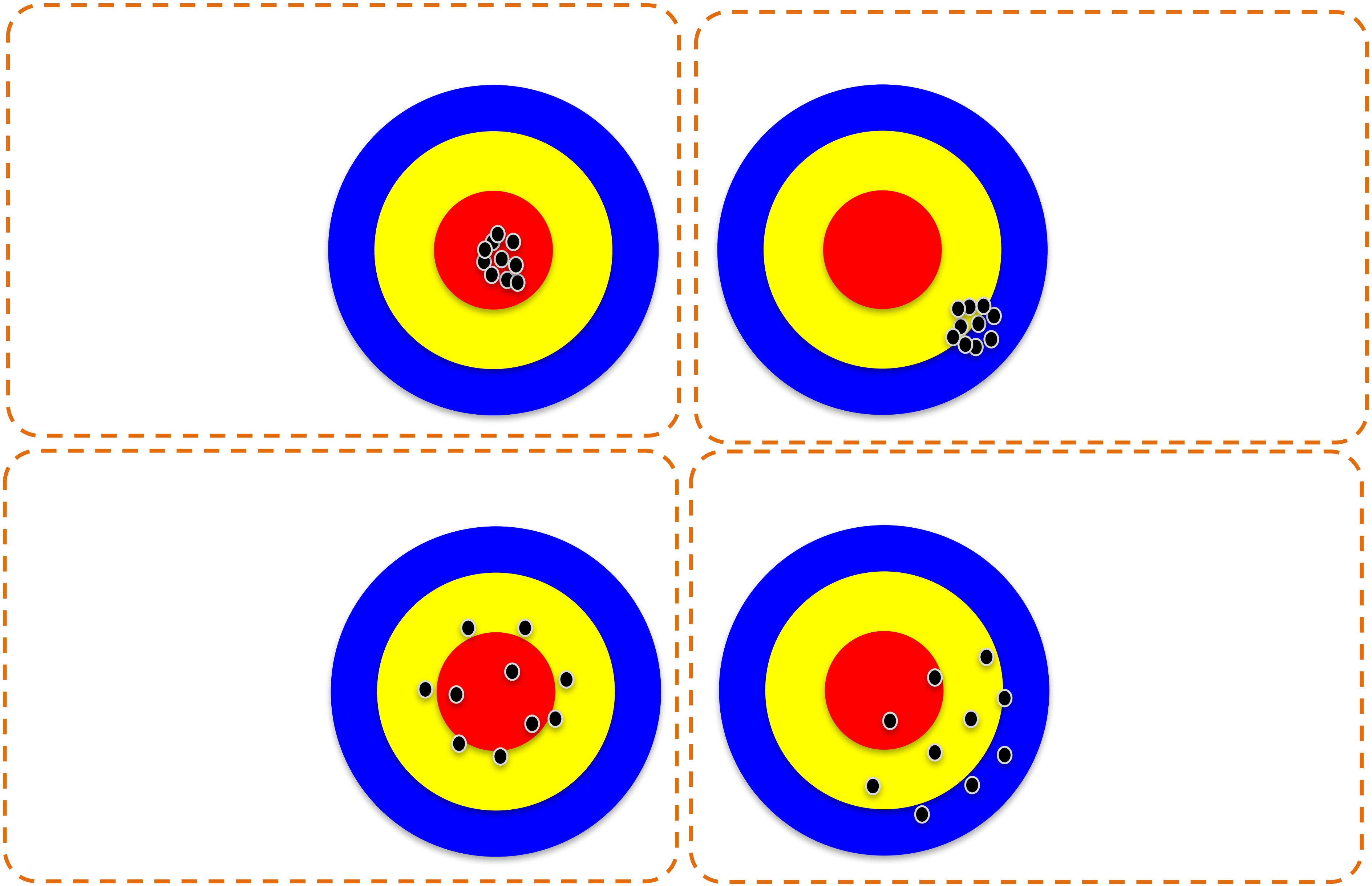
Source: Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In *the 26th International Conference on Computational Linguistics (COLING), Osaka*. 2016.

Evaluation of Text Mining and Sentiment Analysis

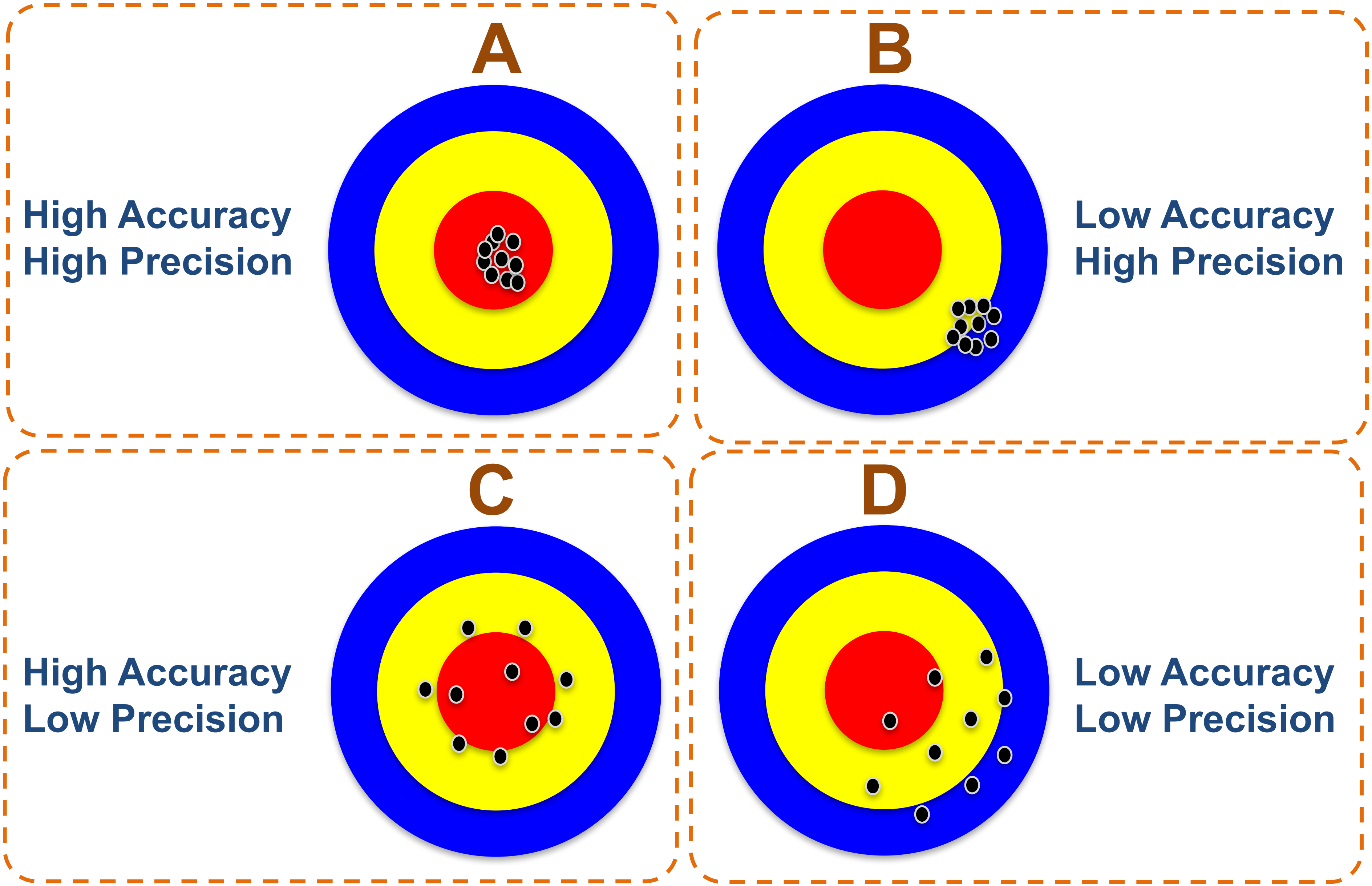
- Evaluation of Information Retrieval
- Evaluation of Classification Model (Prediction)
  - Accuracy
  - Precision
  - Recall
  - F-score

Accuracy Validity

Precision Reliability



Accuracy vs. Precision



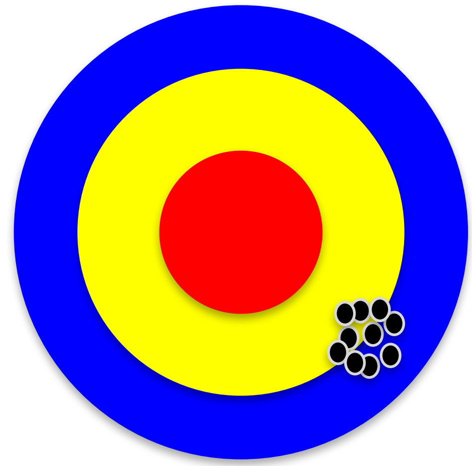
Accuracy vs. Precision

A

High Accuracy
High Precision

High Validity
High Reliability

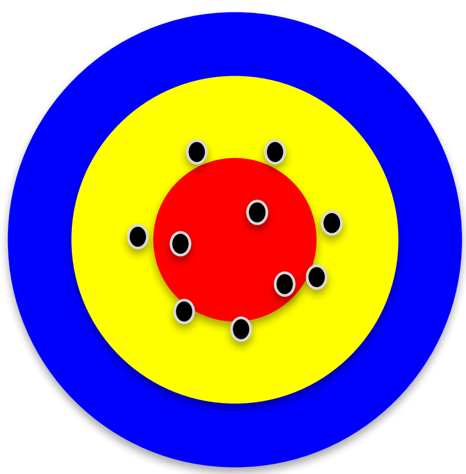
B



Low Accuracy
High Precision

Low Validity
High Reliability

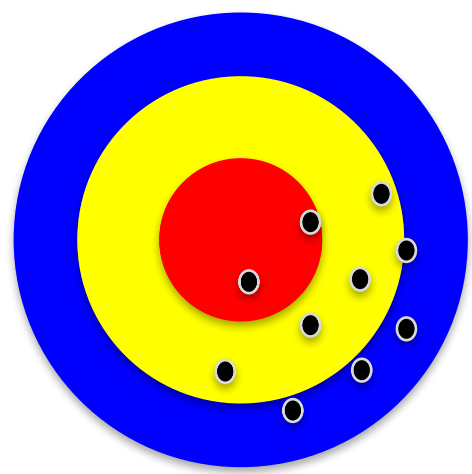
C



High Accuracy
Low Precision

High Validity
Low Reliability

D



Low Accuracy
Low Precision

Low Validity
Low Reliability

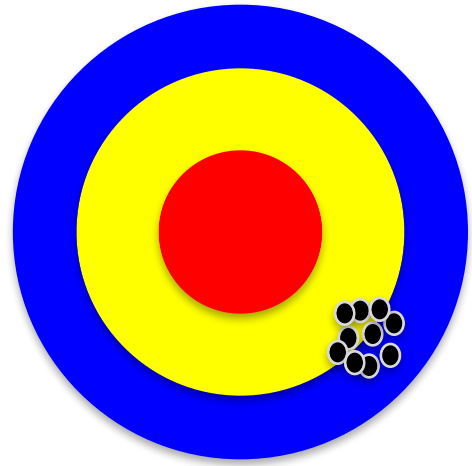
Accuracy vs. Precision

A

High Accuracy
High Precision

High Validity
High Reliability

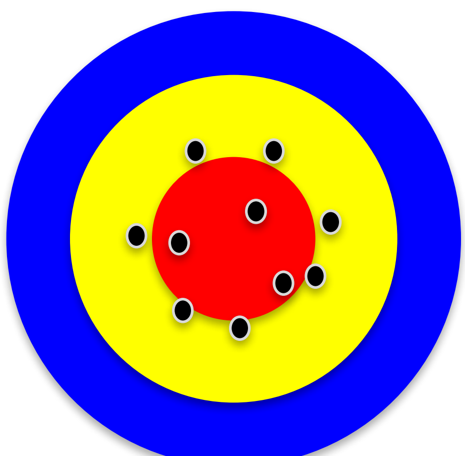
B



Low Accuracy
High Precision

Low Validity
High Reliability

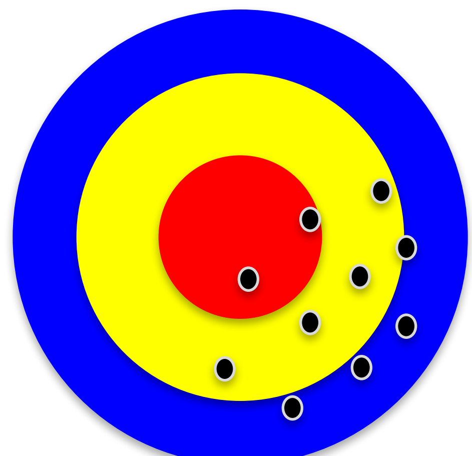
C



High Accuracy
Low Precision

High Validity
Low Reliability

D



Low Accuracy
Low Precision

Low Validity
Low Reliability

Accuracy of Classification Models

- In classification problems, the primary source for accuracy estimation is the confusion matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

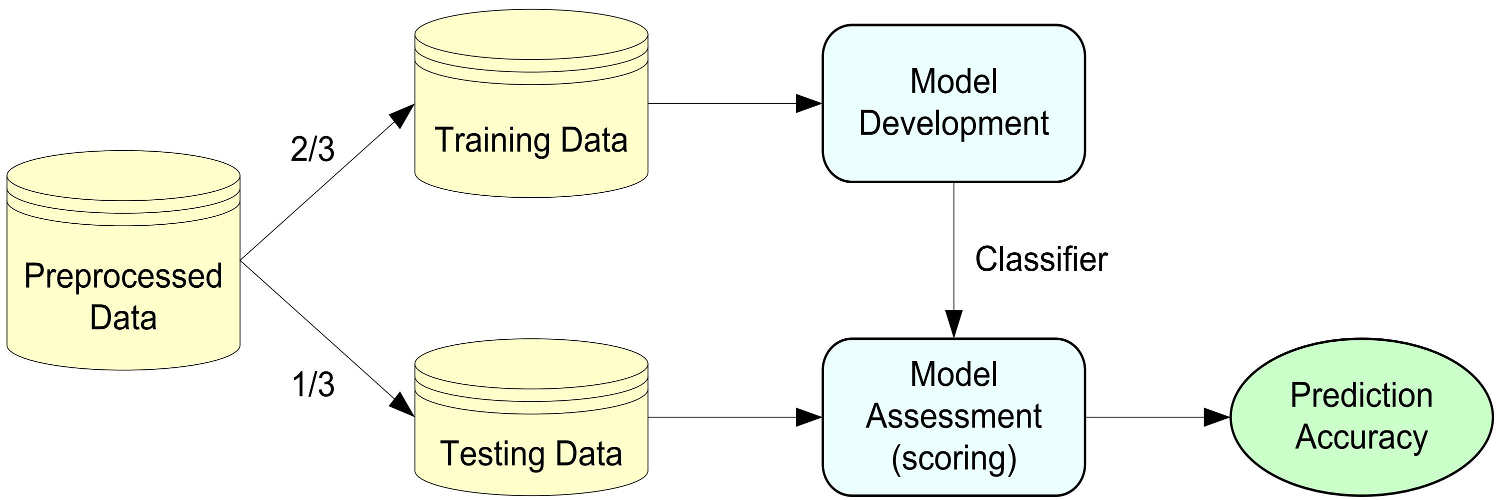
$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Estimation Methodologies for Classification

- Simple split (or holdout or test sample estimation)
  - Split the data into 2 mutually exclusive sets training (~70%) and testing (30%)



  - For ANN, the data is split into three sub-sets (training [~60%], validation [~20%], testing [~20%])

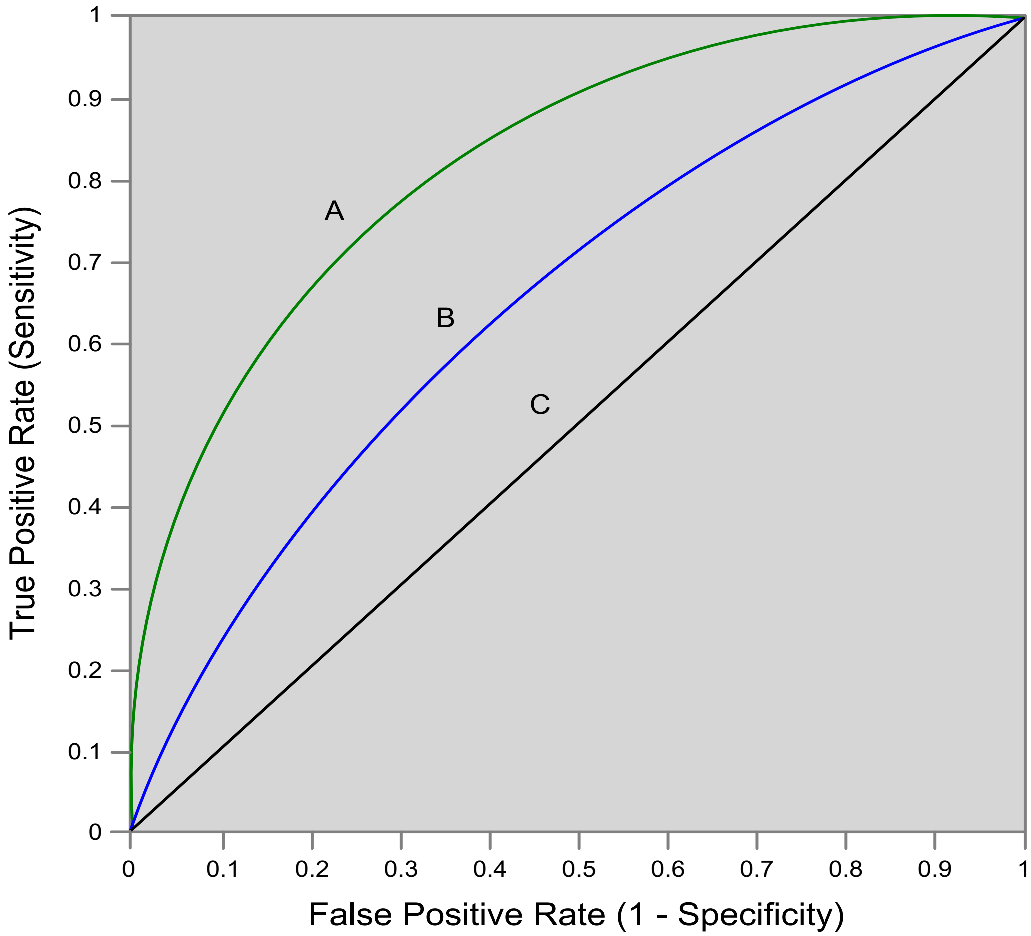
Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Estimation Methodologies for Classification

- *k*-Fold Cross Validation (rotation estimation)
  - Split the data into *k* mutually exclusive subsets
  - Use each subset as testing while using the rest of the subsets as training
  - Repeat the experimentation for *k* times
  - Aggregate the test results for true estimation of prediction accuracy training
- Other estimation methodologies
  - Leave-one-out, bootstrapping, jackknifing
  - Area under the ROC curve

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Estimation Methodologies for Classification – ROC Curve



Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Sensitivity =True Positive Rate

Specificity =True Negative Rate

		True Class (actual value)		
		Positive	Negative	total
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

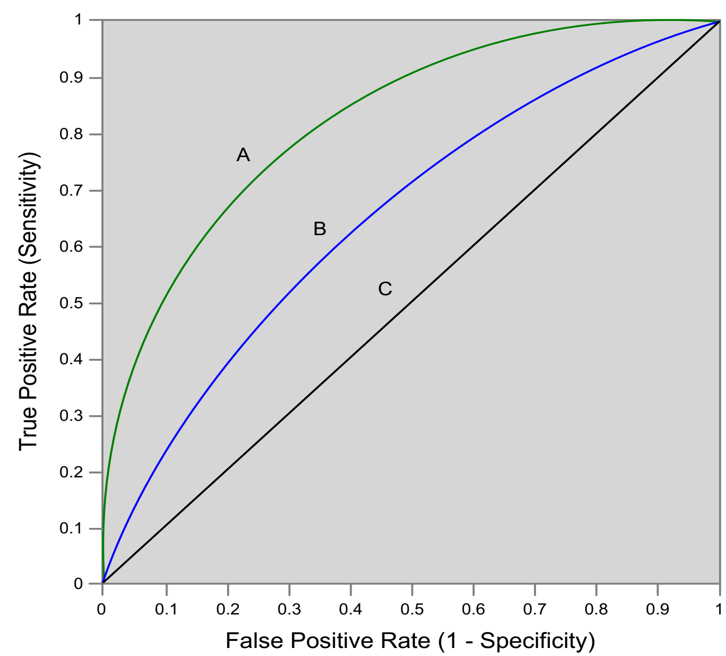
$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$True\ Positive\ Rate\ (\text{Sensitivity}) = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate\ (\text{Specificity}) = \frac{TN}{TN + FP}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN}$$

$$False\ Positive\ Rate\ (1\text{-Specificity}) = \frac{FP}{FP + TN}$$



Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

Predictive Class (prediction outcome)		True Class (actual value): Positive	True Class (actual value): Negative	total
	Positive	True Positive (TP)	False Positive (FP)	P'
	Negative	False Negative (FN)	True Negative (TN)	N'
total		P	N	

$$True\ Positive\ Rate = \frac{TP}{TP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$True\ Positive\ Rate\ (\text{Sensitivity}) = \frac{TP}{TP + FN}$$

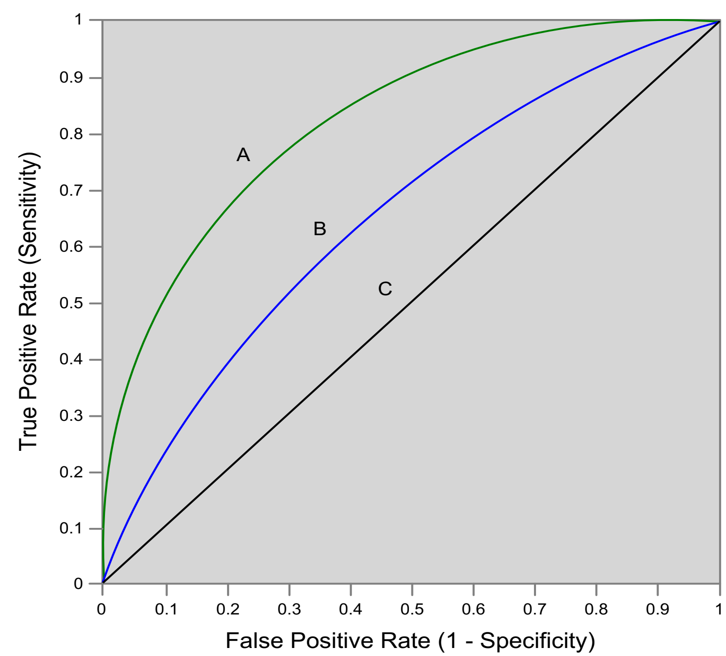
Sensitivity

= True Positive Rate

= Recall

= Hit rate

= TP / (TP + FN)



Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

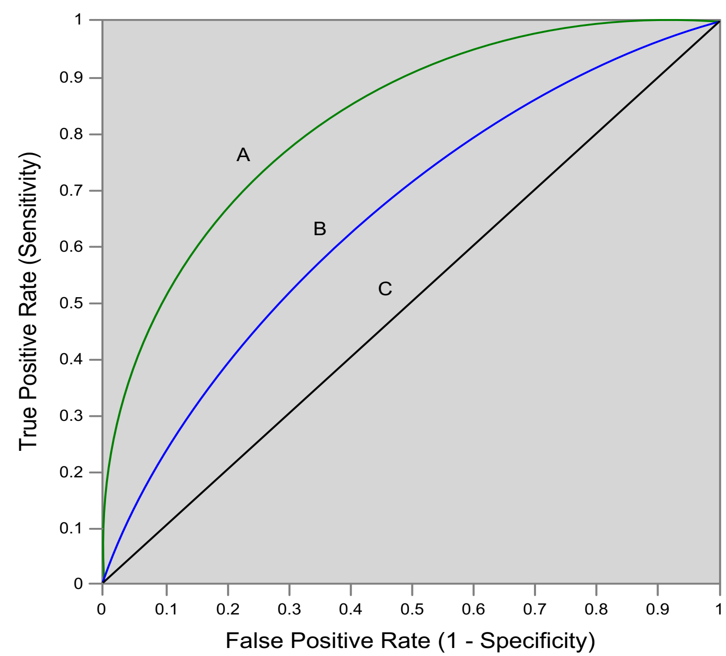
		True Class **(actual value)**		
		Positive	Negative	**total**
Predictive Class **(prediction outcome)**	Positive	True Positive (TP)	False Positive (FP)	P’
	Negative	False Negative (FN)	True Negative (TN)	N’
total		P	N	

$$True\ Negative\ Rate = \frac{TN}{TN + FP}$$

Specificity
= True Negative Rate
= TN / N
= TN / (TN+ FP)

$$True\ Negative\ Rate\ (\text{Specificity}) = \frac{TN}{TN + FP}$$

$$False\ Positive\ Rate\ (1\text{-}\text{Specificity}) = \frac{FP}{FP + TN}$$



Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

		True Class (actual value)		
		Positive	Negative	total
Predictive Class (prediction outcome)	Positive	True Positive (TP)	False Positive (FP)	P’
	Negative	False Negative (FN)	True Negative (TN)	N’
total		P	N	

Precision

= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

Recall

= True Positive Rate (TPR)

= Sensitivity

= Hit Rate

$$Recall = \frac{TP}{TP + FN}$$

F1 score (F-score)(F-measure)

is the harmonic mean of precision and recall

= 2TP / (P + P’)

= 2TP / (2TP + FP + FN)

$$F = 2 * \frac{precision * recall}{precision + recall}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

A

63 (TP)	28 (FP)	91
37 (FN)	72 (TN)	109
100	100	200

Recall
= True Positive Rate (TPR)
= Sensitivity
= Hit Rate
= TP / (TP + FN)

Specificity
= True Negative Rate
= TN / N
= TN / (TN + FP)

TPR = 0.63

$$Recall = \frac{TP}{TP + FN}$$

$$True\ Negative\ Rate\ (\text{Specificity}) = \frac{TN}{TN + FP}$$

FPR = 0.28

$$\text{False}\ Positive\ Rate\ (1\text{-Specificity}) = \frac{FP}{FP + TN}$$

PPV = 0.69
=63/(63+28)
=63/91

$$Precision = \frac{TP}{TP + FP}$$

Precision
= Positive Predictive Value (PPV)

F1 = 0.66
= 2*(0.63*0.69)/(0.63+0.69)
= (2 * 63) /(100 + 91)
= (0.63 + 0.69) / 2 =1.32 / 2 =0.66

$$F = 2 * \frac{precision * recall}{precision + recall}$$

F1 score (F-score) (F-measure)
is the harmonic mean of precision and recall
= 2TP / (P + P')
= 2TP / (2TP + FP + FN)

ACC = 0.68
= (63 + 72) / 200
= 135/200 = 67.5

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

A

63 (TP)	28 (FP)	91
37 (FN)	72 (TN)	109
100	100	200

TPR = 0.63

FPR = 0.28

PPV = 0.69
=63/(63+28)
=63/91

F1 = 0.66
= 2*(0.63*0.69)/(0.63+0.69)
= (2 * 63) /(100 + 91)
= (0.63 + 0.69) / 2 =1.32 / 2 =0.66

ACC = 0.68
= (63 + 72) / 200
= 135/200 = 67.5

B

77 (TP)	77 (FP)	154
23 (FN)	23 (TN)	46
100	100	200

TPR = 0.77

FPR = 0.77

PPV = 0.50

F1 = 0.61

ACC = 0.50

Recall
= True Positive Rate (TPR)
= Sensitivity
= Hit Rate

$$Recall = \frac{TP}{TP + FN}$$

Precision
= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

C

24 (TP)	88 (FP)	112
76 (FN)	12 (TN)	88
100	100	200

TPR = 0.24
FPR = 0.88
PPV = 0.21
F1 = 0.22
ACC = 0.18

C'

76 (TP)	12 (FP)	88
24 (FN)	88 (TN)	112
100	100	200

TPR = 0.76
FPR = 0.12
PPV = 0.86
F1 = 0.81
ACC = 0.82

Recall
= True Positive Rate (TPR)
= Sensitivity
= Hit Rate

$$Recall = \frac{TP}{TP + FN}$$

Precision
= Positive Predictive Value (PPV)

$$Precision = \frac{TP}{TP + FP}$$

Source: http://en.wikipedia.org/wiki/Receiver_operating_characteristic

LeCun, Yann,
Yoshua Bengio,
and Geoffrey Hinton.

"**Deep learning**."

Nature 521, no. 7553 (2015): 436-444.

REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

Machine-learning technology powers many aspects of modern society: from web searches to content filtering on social networks to recommendations on e-commerce websites, and it is increasingly present in consumer products such as cameras and smartphones. Machine-learning systems are used to identify objects in images, transcribe speech into text, match news items, posts or products with users' interests, and select relevant results of search. Increasingly, these applications make use of a class of techniques called deep learning.

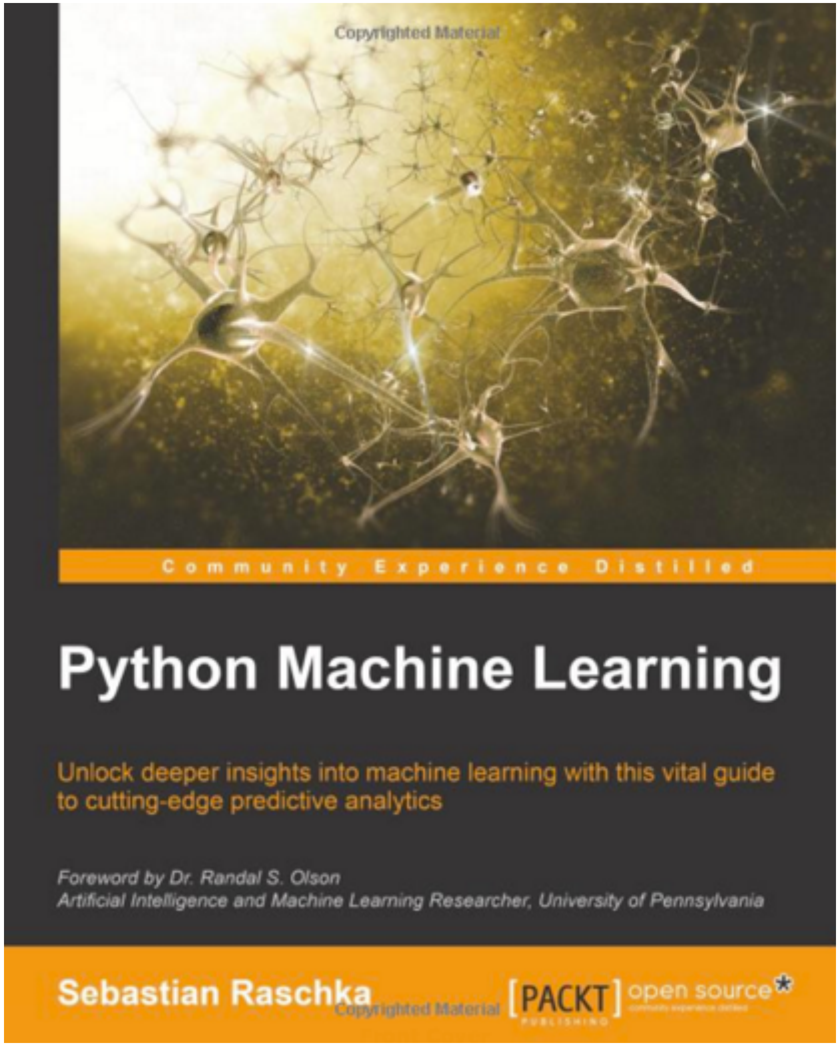
Conventional machine-learning techniques were limited in their ability to process natural data in their raw form. For decades, con- intricate structures in high-dimensional data and is therefore applicable to many domains of science, business and government. In addition to beating records in image recognition[1–4] and speech recognition[5–7], it has beaten other machine-learning techniques at predicting the activity of potential drug molecules[8], analysing particle accelerator data[9,10], reconstructing brain circuits[11], and predicting the effects of mutations in non-coding DNA on gene expression and disease[12,13]. Perhaps more surprisingly, deep learning has produced extremely promising results for various tasks in natural language understanding[14], particularly topic classification, sentiment analysis, question answering[15] and language translation[16,17].

Source: LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

Sebastian Raschka (2015),

Python Machine Learning,

Packt Publishing

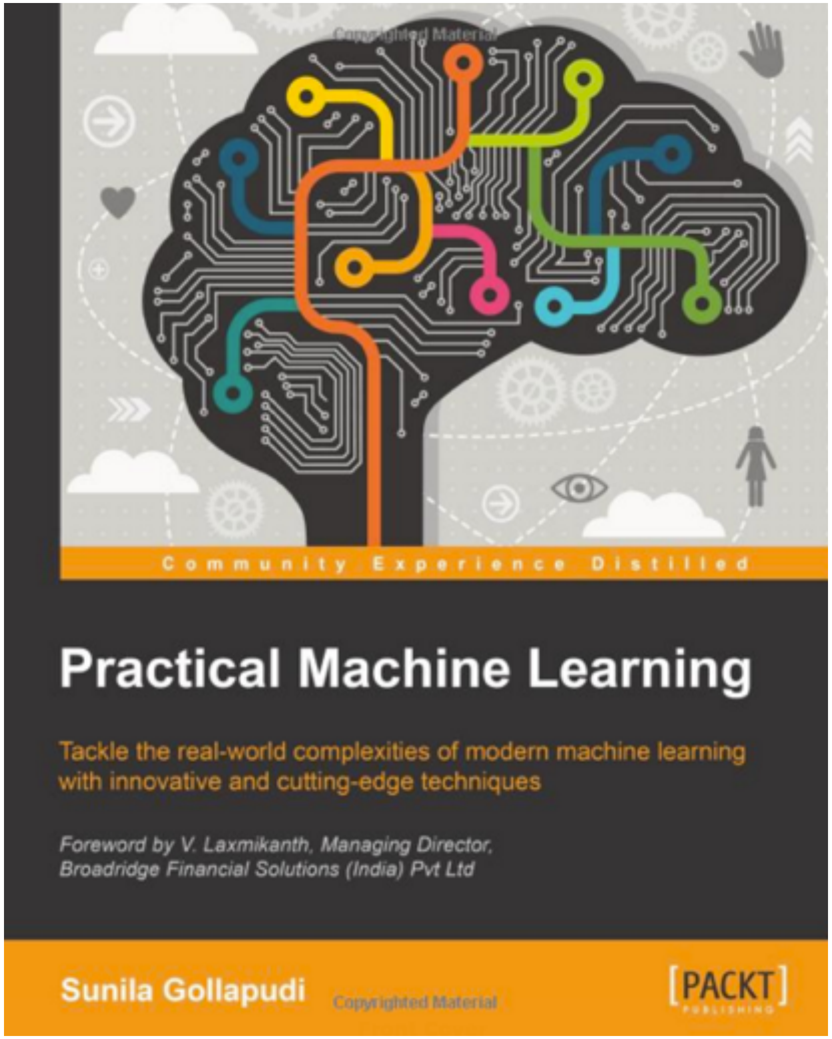


Source: http://www.amazon.com/Python-Machine-Learning-Sebastian-Raschka/dp/1783555130

Sunila Gollapudi (2016),

Practical Machine Learning,

Packt Publishing



Source: http://www.amazon.com/Practical-Machine-Learning-Sunila-Gollapudi/dp/178439968X

Machine Learning Models

- Deep Learning
- Kernel
- Association rules
- Ensemble
- Decision tree
- Dimensionality reduction
- Clustering
- Regression Analysis
- Bayesian
- Instance based

Source: Sunila Gollapudi (2016), Practical Machine Learning, Packt Publishing

Deep Learning

Intelligence from Big Data

Source: https://www.vlab.org/events/deep-learning/

Deep Learning for Sentiment Analytics

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts
Stanford University, Stanford, CA 94305, USA
richard@socher.org,{aperelyg,jcchuang,ang}@cs.stanford.edu
{jeaneis,manning,cgpotts}@stanford.edu

Abstract

Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new treebank, this model out-

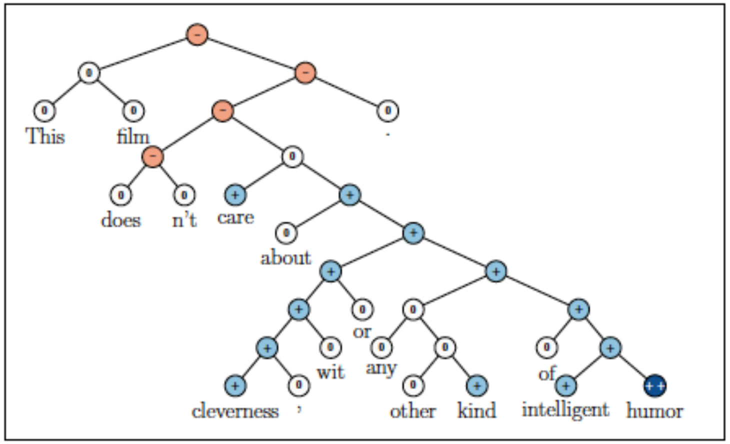
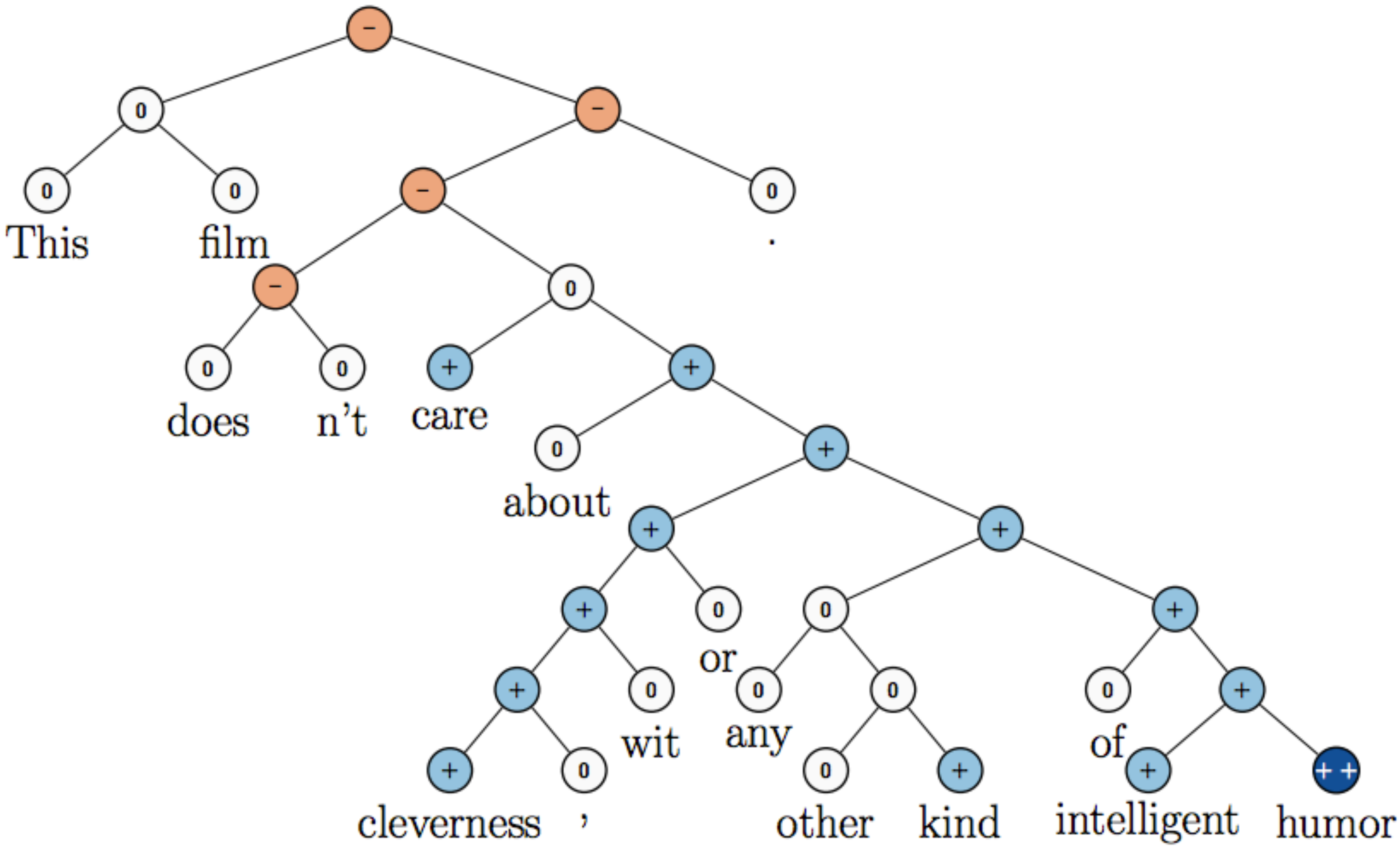


Figure 1: Example of the Recursive Neural Tensor Network accurately predicting 5 sentiment classes, very negative to very positive (– –, –, 0, +, + +), at every node of a parse tree and capturing the negation and its scope in this sentence.

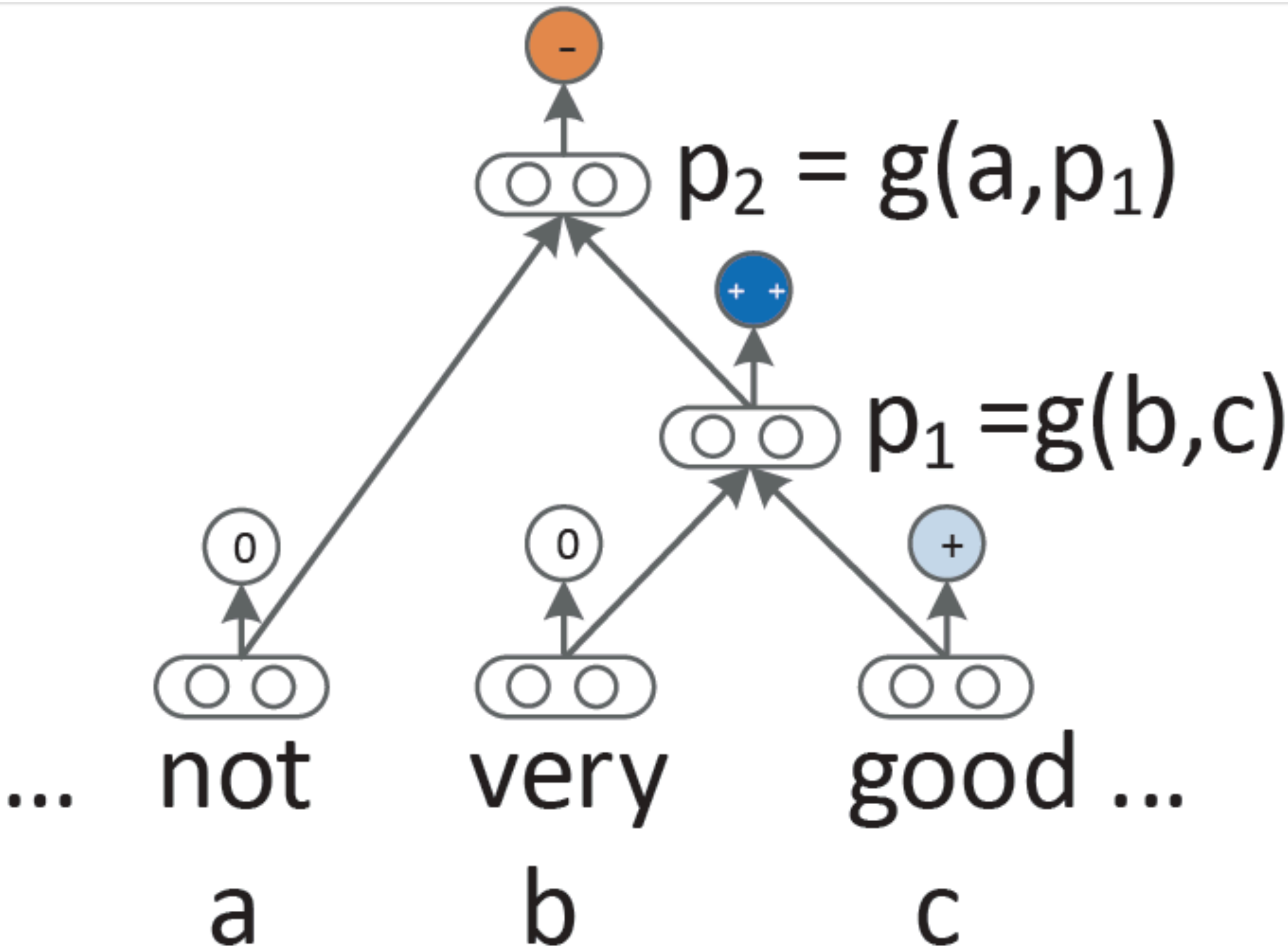
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Tensor Network (RNTN)



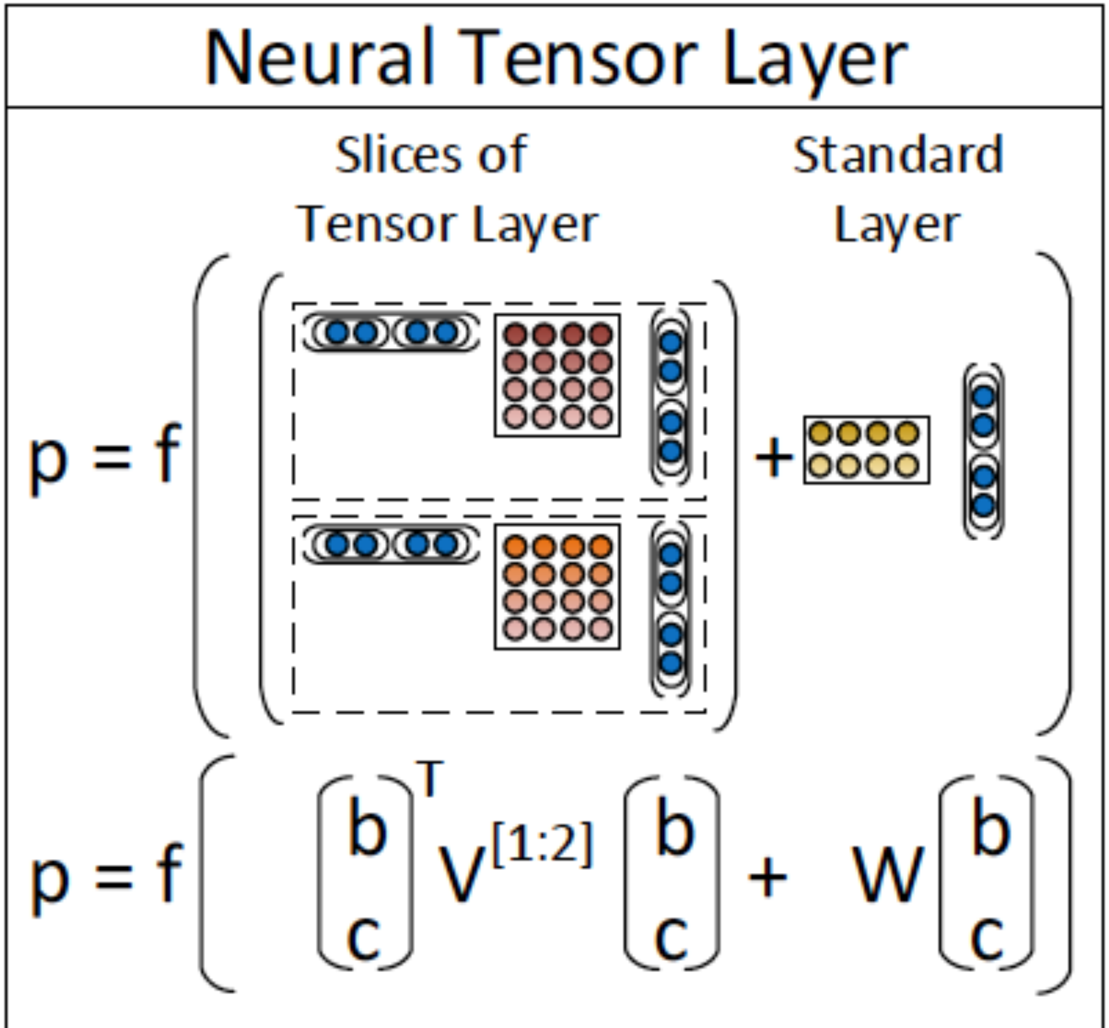
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Network (RNN) models for sentiment



Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Recursive Neural Tensor Network (RNTN)



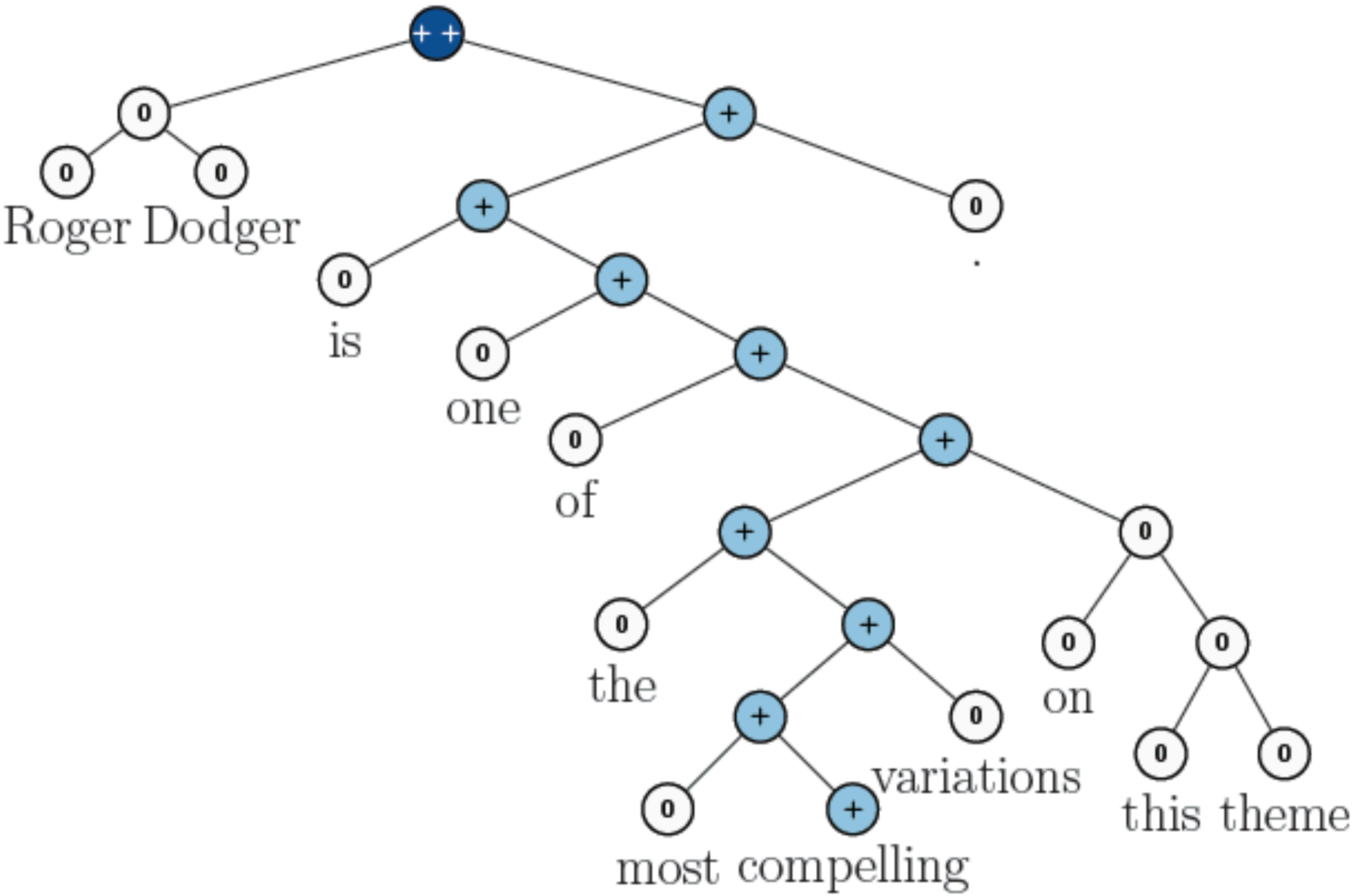
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Roger Dodger is one of the most compelling variations on this theme.

Roger Dodger is one of the least compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

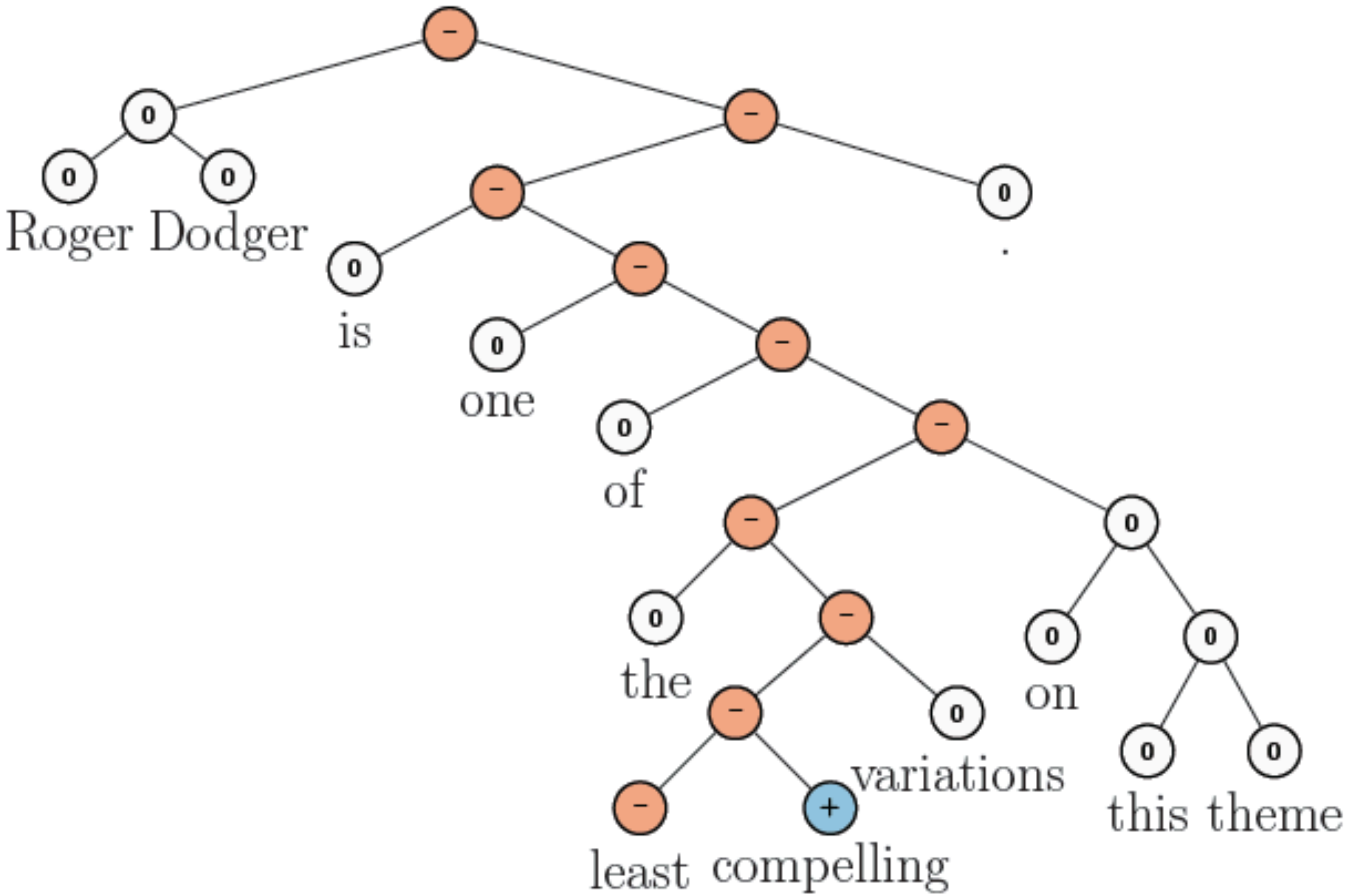
RNTN for Sentiment Analysis



Roger Dodger is one of the most compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

RNTN for Sentiment Analysis



Roger Dodger is one of the least compelling variations on this theme.

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Accuracy for fine grained (5-class) and binary predictions at the sentence level (root) and for all nodes

Model	Fine-grained		Positive/Negative	
	All	Root	All	Root
NB	67.2	41.0	82.6	81.8
SVM	64.3	40.7	84.6	79.4
BiNB	71.0	41.9	82.7	83.1
VecAvg	73.3	32.7	85.1	80.1
RNN	79.0	43.2	86.1	82.4
MV-RNN	78.7	44.4	86.8	82.9
RNTN	**80.7**	**45.7**	**87.6**	**85.4**

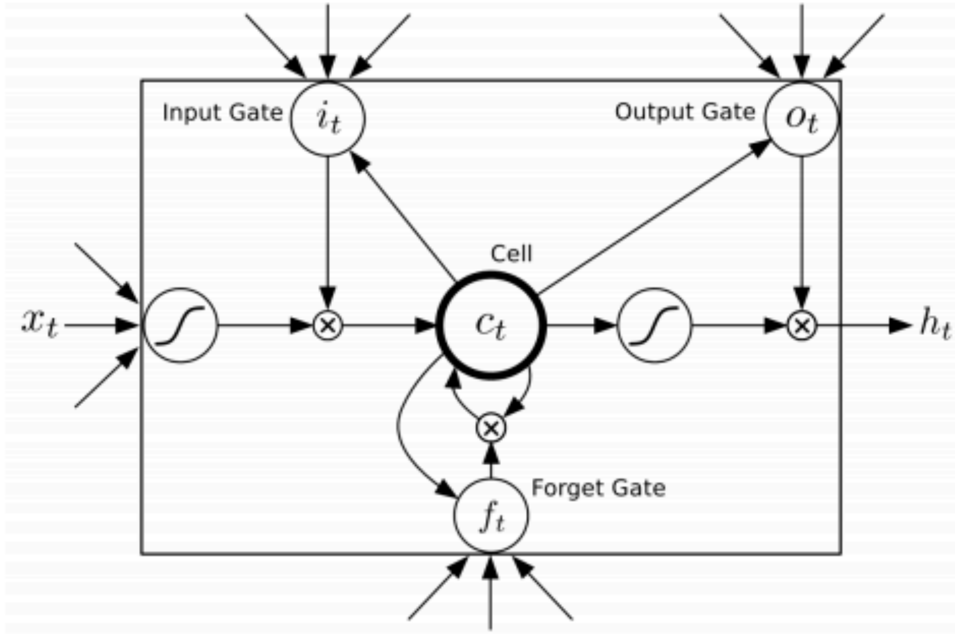
Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

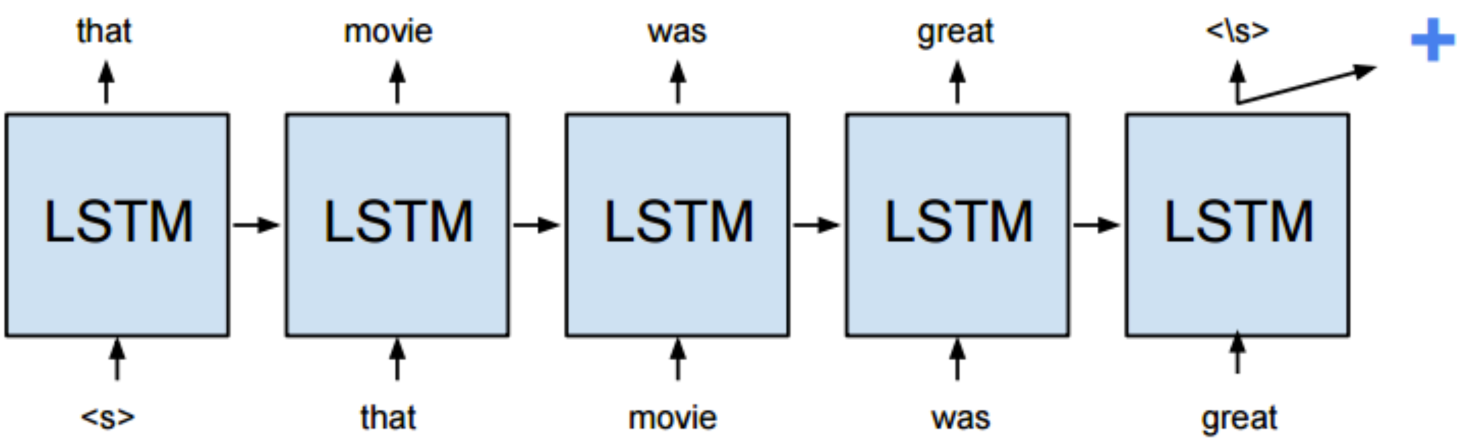
Accuracy of negation detection

Model	Accuracy	
	Negated Positive	Negated Negative
biNB	19.0	27.3
RNN	33.3	45.5
MV-RNN	52.4	54.6
RNTN	**71.4**	**81.8**

Source: Richard Socher et al. (2013) "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank", EMNLP 2013

Long Short-Term Memory (LSTM)





Source: https://cs224d.stanford.edu/reports/HongJames.pdf

Deep Learning for Sentiment Analysis CNN RNTN LSTM

Model	Fine (5-class)	Binary
DCNN (Blunsom, et al. 2014)	0.485	0.868
RNTN (Socher, et al. 2013)	0.457	0.854
CNN-non-static (Kim, 2014)	0.480	0.872
CNN-multi-channel (Kim, 2014)	0.474	0.881
DRNN w. pretrained word-embeddings (Irsoy and Cardie, 2014)	0.498	0.866
Paragraph Vector (Le and Mikolov. 2014)	0.487	0.878
Dependency Tree-LSTM (Tai, et al, 2015)	0.484	0.857
Constituency Tree-LSTM (Tai, et al, 2015)	0.439	0.820
Constituency Tree-LSTM (Glove vectors) (Tai, et al, 2015)	0.510	0.880
Paragraph Vector	0.391	0.798
LSTM	0.456	0.843
Deep Recursive-NN	0.469	0.847

Source: https://cs224d.stanford.edu/reports/HongJames.pdf

Performance Comparison of Sentiment Analysis Methods

	Method	Data Set	Acc.	Author
Machine Learning	SVM	Movie reviews	86.40%	Pang, Lee[23]
	CoTraining SVM	Twitter	82.52%	Liu[14]
	Deep learning	Stanford Sentiment Treebank	80.70%	Richard[18]
Lexical based	Corpus	Product reviews	74.00%	Turkey
	Dictionary	Amazon's Mechanical Turk	---	Taboada[20]
Cross-lingual	Ensemble	Amazon	81.00%	Wan,X[16]
	Co-Train	Amazon, ITI68	81.30%	Wan,X.[16]
	EWGA	IMDb movie review	>90%	Abbasi,A.
	CLMM	MPQA,NTCIR,ISI	83.02%	Mengi
Cross-domain	Active Learning	Book, DVD, Electronics, Kitchen	80% (avg)	Li, S
	Thesaurus			Bollegala[22]
	SFA			Pan S J[15]

Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, Vol 139, No. 11, 2016. pp.5-15

Kumar Ravi and Vadlamani Ravi (2015),

"A survey on opinion mining and sentiment analysis: tasks, approaches and applications."

Knowledge-Based Systems,

89, pp.14-46

Knowledge-Based Systems 89 (2015) 14–46

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

ELSEVIER

A survey on opinion mining and sentiment analysis: Tasks, approaches and applications

CrossMark

Kumar Ravi [a,b], Vadlamani Ravi [a,*]

[a] Center of Excellence in CRM and Analytics, Institute for Development and Research in Banking Technology, Castle Hills Road No. 1, Masab Tank, Hyderabad 500057, AP, India
[b] School of Computer & Information Sciences, University of Hyderabad, Hyderabad 500046, AP, India

Table 5
Sentiment classification accuracy reported on common datasets.

S#	Dataset	Articles	Obtained result
1	Pang and Lee [167]	[156]	92.70% accuracy
2		[112]	90.45% F_1
3		[169]	90.2% accuracy
4		[35]	89.6% accuracy
5		[54]	87.70% accuracy
6		[46]	87.4% accuracy
7		[50]	86.5% accuracy
8		[26]	85.35% accuracy
9		[162]	81% F_1
10		[124]	79% accuracy & 86% F_1
11		[61]	76.6% accuracy
12		[69]	76.37% accuracy
13		[48]	75% precision
14		[98]	79% precision
15	Pang et al. [33]	[109]	Approx. 90% accuracy
16		[165]	88.5% accuracy
17		[172]	87% accuracy
18		[33]	82.9% accuracy
19		[156]	78.08% accuracy
20		[180]	75% accuracy
21		[48]	60% precision
22		[195]	86.04%
23	Blitzer et al. [149]	[45]	84.15% accuracy
24		[99]	80.9% (Avg.) accuracy
25		[54]	85.15% (Avg.) Max. 88.65% accuracy on Kitchen reviews
28		[165]	88.7% accuracy
29		[61]	71.92% accuracy

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Classification Accuracy

S#	Dataset	Articles	Obtained result
1	Pang and Lee [167]	[156]	92.70% accuracy
2		[112]	90.45% F_1
3		[169]	90.2% accuracy
4		[35]	89.6% accuracy
5		[54]	87.70% accuracy
6		[46]	87.4% accuracy
7		[50]	86.5% accuracy
8		[26]	85.35% accuracy
9		[162]	81% F_1
10		[124]	79% accuracy & 86% F_1
11		[61]	76.6% accuracy
12		[69]	76.37% accuracy
13		[48]	75% precision
14		[98]	79% precision

B. Pang, L. Lee, A sentiment education: sentiment analysis using subjectivity summarization based on minimum cuts, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, July 2004, p. 271

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Classification Accuracy

S#	Dataset	Articles	Obtained result
15	Pang et al. [33]	[109]	Approx. 90% accuracy
16		[165]	88.5% accuracy
17		[172]	87% accuracy
18		[33]	82.9% accuracy
19		[156]	78.08% accuracy
20		[180]	75% accuracy
21		[48]	60% precision
22		[195]	86.04%

B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, Association for Computational Linguistics, 2002, pp. 79–86.

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Classification Accuracy

S#	Dataset	Articles	Obtained result
23	Blitzer et al. [149]	[45]	84.15% accuracy
24		[99]	80.9% (Avg.) accuracy
25		[54]	85.15% (Avg.) Max. 88.65% accuracy on Kitchen reviews
28		[165]	88.7% accuracy
29		[61]	71.92% accuracy

J. Blitzer, M. Dredze, F. Pereira, Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification, in: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, ACL'07, vol. 7, 2007, pp. 187–205 (13, 29).

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Techniques for Sentiment Analysis

Applied techniques	#Articles
SVM	55
Dictionary based approaches (DBA)	41
NB	28
NN	11
DT	9
Maximum entropy	8
Logistic regression	9
Linear regression	8
Ontology	8
LDA	8
Random forest	4
SVR	5
CRF and rCRP	5
Boosting	4
SVM-SMO	4
Fuzzy logic	3
Rule miner	4
EM	3
K-medoids	1
RBF NN	1

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Analysis Articles in Journals (2002-2014)

S#	Name of journals	#Articles
1	Expert Systems with Applications	33
2	Decision Support Systems	28
3	Knowledge-based Systems	17
4	IEEE Intelligent Systems	12
5	IEEE Transactions on Knowledge and Data Engineering	6
6	IEEE Transactions on Affective Computing	3
7	Information Sciences	3
8	Information Processing and Management	3
9	Computer Speech and Language	2
10	Communications of the ACM	2
11	Journal of Computer Science and Technology	2
12	Journal of Informetrics	2
13	Information Retrieval	2
14	Computer Speech and Language	2
15	Inf. Retrieval	1

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Publicly Available Datasets for Sentiment Analysis

S#	Data set	Type	Lang.	Web resource	Details
1	Stanford large movie data set	Movie Reviews	English	http://ai.stanford.edu/~amaas/data/sentiment/	Movie Reviews
2	COAE2008	Product Reviews	Chinese	http://ir-china.org.cn/coae2008.html	2739 documents for movie, education, finance, economics, house, computer, mobile phones, etc. 1525 +ve, 1214 –ve
3	Boacar	Car Reviews	Chinese	http://www.riche.com.cn/boacar/	11 type of car TradeMarks and total review 1000 words, having 578 POS, 428 –ve reviews
4	[187]	Reviews, forums	English	http://sifaka.cs.uiuc.edu/~wang296/Data/	Accessed: 27 August, 2014
5	[188]	Reviews	English	http://uilab.kaist.ac.kr/research/WSDM11	Aspect oriented dataset. Accessed: 18 December, 2014
6	Movie-v2.0	Movie Reviews	English	http://www.cs.cornell.edu/people/pabo/movie-review-data/	Data size: 2000 Positive: 1000 Negative: 1000
7	Multi-domain	Multi-domain	English	http://www.cs.jhu.edu/~mdreze/datasets/sentiment	
8	SkyDrive de Hermit Dave	Spanish Word Lists	Spanish	https://skydrive.live.com/?cid=3732e80b128d016f&id=3732E80B128D016F%213584	
9	TripAdvisor	Reviews	Spanish	http://clic.ub.edu/corpus/es/node/106	18,000 customer reviews on hotels and restaurants from Hopinion
10	[38]	Multi-Domain	English	www2.cs.uic.edu/~liub/FBS/sentiment-analysis.html	6800 opinion words on 10 different products
11	TBOD [144]	Reviews	English		Product Review on Cars, Headphones, Hotels
12	[68]	Product Reviews	English	http://www.lsi.us.es/_fermin/index.php/Datasets	Product Reviews from Epinion.com on headphones 587 reviews, hotels 988 reviews and cars 972 reviews
13	[148]	Movie Reviews	Turkish	http://www.win.tue.nl/~mpechen/projects/smm/#Datasets	5331 positive and 5331 negative reviews on movie
14	[148]	Product Reviews	Turkish	http://www.win.tue.nl/~mpechen/projects/smm/#Datasets	700 +ve &700 –ve reviews on books, DVD, electronics, kitchen appliances
15	ISEAR	English sentences	English	www.affective-sciences.org/system/files/page/2636/ISEAR.zip	The dataset contains 7666 such statements, which include 18,146 sentences, 449,060 running words.
16	[149]	Product Reviews	English	http://www.cs.jhu.edu/~mdredze/datasets/sentiment/	Amazon reviews on 4 domain (books, DVDs, electronics, kitchen appliances)
17	DUC data, NIST	Texts	English	http://www-nlpir.nist.gov/projects/duc/data.html, http://www.nist.gov/tac/data/index.html	Text summarization data
18	[70]	Restaurant and Hotel Reviews	English	http://uilab.kaist.ac.kr/research/WSDM11	Restaurant and Hotel Reviews from Amazon and Yelp
19	[114]	Restaurant Reviews	Cantonese	http://www.openrice.com	Reviews on restaurant
20	[125]	Biographical Articles	Dutch	http://www.iisg.nl/bwsa	574 Biographical articles
21	Spinn3r dataset	Multi-Domain	English	http://www.icwsm.org/2011/data.php	
22	[86]	Ironic Dataset	English	http://users.dsic.upv.es/grupos/nle/	3163 ironic reviews on five products
23	HASH [179]	Tweets	English	http://demeter.inf.ed.ac.uk	31,861 Pos tweets, 64,850 Neg tweets, 125,859 Neu tweets
24	EMOT [179]	Tweets and Emoticons	English	http://twittersentiment.appspot.com	230,811 Pos & 150,570 Neg tweets
25	ISIEVE [179]	Tweets	English	www.i-sieve.com	1520 Pos tweets, 200 Neg tweets, 2295 Neu tweets
26	[177]	Tweets	English	e-mail: apoorv@cs.columbia.edu	11,875 tweets
27	[52]	Opinions	English	http://patientopinion.org.uk	2000 patient opinions
28	[96]	Tweets	English	http://goo.gl/UQvdx	667 tweets
29	[39]	Movie Reviews	English	http://ai.stanford.edu/~amaas/data/sentiment/	50,000 movie reviews
30	[164]	Tweets	English	http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip	
31	[210]	Spam Reviews	English	http://myleott.com/op_spam	400 deceptive and 400 truthful reviews in positive and negative category. Last Accessed by: 12 April, 2015
32	[230]	Sarcasm and nasty reviews	English	https://nlds.soe.ucsc.edu/iac	1000 discussions, ~390,000 posts, and some ~73,000,000 words

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Analysis Datasets

- Product Reviews (PR)
- Movie Reviews (MR)
- Restaurant Reviews (RR)
- Micro-blog (MB)
- Global domain (G)

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Sentiment Analysis Dictionary

- SenticNet (SN)
- WordNet (WN)
- ConceptNet (CN)
- WordNet-Affect (WNA)
- Bing Liu Opinion Lexicon (OL)

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Summary of reviewed articles

Ref.	Concepts and techniques utilized	P	L	Type of data	Dictionary
[8]	Page rank, Gradient descent, Linear regression	2	E	PR	
[11]	Link mining, Collective classification	NA	E	MB	
[12]	AdaBoost.HM	2	E	G	GI
[13]	DBA	5	E	News Comments	New Lexicon
[18]	DBA, SOFNN, Linear regression	2, 7	E	MB, DJIA data	OF, GPOMS
[21]	Regression, Random walk, SVM	4, 2	E		ANEW, CN
[22]	Cohen's K coefficient	6, 2	I	MB	SN
[23]	Fuzzy clustering, PMI, DBA	6, 2	E	G	WNA, SN, WN.
[24]	DBA	NA	D	G	Dutch WN
[25]	Association Miner CBA, DBA	2	E	PR	WN
[26]	SVM	2	E	PR	
[27]	Markov-Chain Monte Carlo (MCMC)	NA	E	Online discussion	
[29]	SVM with Gaussian Kernel	3, 2			MPQA
[30]	Ontology, K-means	2	E		ReiAction [122],[a] Family Relation[b]
[32]	PMI-IR	2	E	Multi-domain	
[33]	NB, SVM, ME	2	E	MR	
[35]	Ontology, DBA	2	E	MR	SWN
[36]	New Algorithm, DBA	2	E	MR, Book, Mobile	11 dictionaries
[37]	CRF	NA		PR	
[40]	Multinomial inverse regression	3	E	MB	
[41]	FFCA, Lattice	2	E	PR	
[43]	Analytic hierarchy process	NA	C	MB	
[44]	Fisher's discriminant ratio, SVM	2	C	PR	
[45]	Semantic orientation, SVM	3, 2	E	PR	SWN
[46]	MNB, ME, SVM	3, 2	E, D, F	Forum, Blog, PR	
[47]	DBA	2	D, E	News	
[48]	Semantic orientation and BackProp	2	E	Blogs, PR	
[49]	Probabilistic Matrix Factorization	NA	C	MB	
[50]	NB, SVM, NN	2	E	PR	
[51]	SVM, NN	NA	C	MB	
[52]	DNN, CNN, K-medoids, KNN	NA	E	G	CN, WNA, AffectiveSpace
[53]	SVM, NN, MLP, DT, GA, Stepwise LR, RBC	2	E	News	
[54]	NB, ME, SVM	2	E	PR	
[55]	DBA	5, 2	E	MB	
[56]	NB, EM	NA	E	PR	WN
[57]	SVM, NN	5, 2	E	MB	

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Summary of reviewed articles

Ref.	Concepts and techniques utilized	P	L	Type of data	Dictionary
[58]	SVM	NA	E	Suicide Notes	WN, SWN.
[59]	EM	NA	E, S	PR	fullStrengthLexicon[c]
[60]	ME	NA	E	MB	
[61]	Bayesian Model, LDA	2	E	PRMPQA, Appraisal Lexicons[d]	
[62]	Fuzzy Set, Ontology	2	C	PR	
[63]	ME, Bootstrapping, IG	3, 2	C	PR	Hownet, NEUCSP[e]
[64]	DBA	NA	E	e-mail, book	Roget Thesaurus[f]
[66]	NB, ME, DT, KNN, SVM	NA	C, E	PR, Forums	
[67]	SVM, DBA	2	E	PR	GI
[68]	DBA, Random walk algorithm	2	E	PR	
[69]	DBA	2	E	PR	
[70]	Linear Regression	NA	C	PR, social network	
[73]	BayesNet, J48, Jrip, SVM, NB, ZeroR, Random	5, 2	E	News, Magazine	
[74]	Semantic relationships	2	E		SWN, GI
[75]	Multilingual bootstrapping and cross-lingual bootstrapping, linear regression, IG	NA	E, R		WN
[76]	Bootstrapping, DT, MLP, PCA, SLR, SMO-SVM	2	E	Phone Reviews	WN
[77]	LR, SVM, RF	2	B	e-mails	
[78]	Discretionary accrual model	NA	E	Book Reviews	
[80]	Bayes-Nash equilibria	NA	E	MB	
[81]	RF	NA	E	PR	
[85]	DBA	3, 2	E	MB	SWN
[86]	Semantic, NB, SVM, DT	NA		PR	WN, MSOL, WNA
[88]	SVM, LR, CRF	NA	E	PR	
[90]	SVM, NB	NA	E	MB	
[91]	K-means, SVM	NA	C	Forums	
[92]	HMM-LDA	NA	E	PR	
[93]	Two level CRF	NA	E	PR	
[94]	Corpus based approach, SVM, NB, C4.5, BBR	5, 2	E, S	PR	SWN, Tree Tagger
[95]	SVM	NA	E		WNA, LIWC, VerbOcean, CN
[96]	DBA, Ontology	2	E	MB	
[97]	SMO-SVM, DBA	2	E	MR	SWN, WN
[98]	NB and Ontology	2	E	PR, MR	WN
[99]	Cosine similarity, L1 regularized logistic regression	2	E	PR	WN and SWN
[100]	Association miner CBA	NA	C	PR	
[101]	NN, C4.5, CART, SVM, NB	2	E	MB	

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Summary of reviewed articles

Ref.	Concepts and techniques utilized	P	L	Type of data	Dictionary
[102]	SVM	2	C	HR, PR	TU lexicon[g]
[107]	LDA, DBA	2	E	RR, HR	MPQA, SWN
[108]	SVM	2	A	Dialects, MB, Wiki Talks, Forums	
[109]	Rule-based multivariate features, SVM	2	E	MR, PR, Automobile	
[110]	DBA	2	S	MR	BLEL, WN
[111]	NB, SVM	2	E	RR	SWN
[112]	DBA, RBC, SVM	2	E	MR, Product, MySpace texts	WN, GI
[114]	IG, DBA	2	CT	RR	
[115]	SVM, Statistical approach	2	E, C	HR, Mobile	
[116]	DBA, SVM, NB, LR, J48, Jrip, AdaBoost, Decision Table, MLP, NB.	2	E	MySpace	SentiStrength
[117]	DBA	2	E	MB	SWN
[118]	SMO-SVM, LR, AdaBoost, SVR, DT, NB, J48, Jrip	2	E	Social Media	SentiStrength
[121]	Adaptive-NB	NA	C	PR	
[123]	SVR	6, 2	C	Sina-Wiebo	
[124]	NB	2	E	Social & Mass media	
[125]	Lexical features, NB, Linear SVM, Jrip, KNN	2	D	Biographies	Brouwers thesaurus
[126]	DBA	2	E	MB	OL
[127]	DBA	5, 2	E	G	SentiStrength
[130]	SVR, RBF	NA			
[131]	SVM, NB	3	E	MB, PR	
[132]	New Algorithm	NA		PR	
[148]	SVM, NB, ME	2	E, T		
[154]	New algorithm, Lexical features	3	E	PR	
[155]	SP-LSA, AR, EM, ε-SVR	2	E	MR	2030 appraisal words
[156]	Tabu search, MB, NB, SVM, ME	2	E	MR and News	
[157]	PSO and SVM	2	E	MB	
[158]	DBA	3, 2	E	Mobile Reviews	Moreo et al. [13]
[160]	EWGA, SVM, Bootstrapping	2	E, A	Forums	
[162]	Class sequential rules	3	E	MR	SWN
[163]	DBA, SVM, NB, Logistic, NN	2	E	MB	10 dictionaries
[165]	Semantic, GI, Chi-square, SVM	2	E	MR and PR	
[166]	Semantic	2	C	HR	
[167]	NB, SVM, Min.-cut in the graph	2	E	MR	
[168]	Linear classifiers, Clique, MIRA classifier	2	E	PR	
[169]	DBA, SVM, and SMO-SVM	2	E	MR	WN
[170]	DBA	3	J	MR and PR	Yi et al. [7] lexicon

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Summary of reviewed articles

Ref.	Concepts and techniques utilized	P	L	Type of data	Dictionary
[171]	DBA	2	E	Web pages, News	
[172]	SVM, Osgoodian values, PMI	2	E	MR	WN
[173]	Transfer-based machine translation	2	J	Camera Review	
[174]	ME	2	E	MR	
[175]	DBA, Sigmoid scoring	2	C	Blogs	Hownet
[176]	SVM, PMI	2	E	MB	GI
[177]	Convolution kernels [152], SVM, DBA	2, 3	E	MB	WN, DAL [151]
[178]	Statistical method of OASYS [8]	C	E	News articles	OASYS
[179]	Boosting, SVM	3	E	MB	MPQA, NetLingo
[180]	Bipartite graph, Regularization operator	2	E	Blogs	
[182]	LDA, Ontology, MCMC	2	E	Multi-domain	OF
[183]	SVM, TF-IDF	2	E	News headlines, Forex Rate	SWN
[184]	Vector space model	3	E	News articles	Harvard IV
[185]	Modified LDA	5	E	PR	
[186]	Recursive Chinese Restaurant Process	2	E	PR	
[189]	LDA incorporated with domain knowledge	NA	E	Camera and HR	
[190]	CRF, syntactic and semantic features	2	E	PR, Facebook text	
[191]	LDA, Appraisal expression pattern	NA	E	HR, RR, PR	
[192]	PMI, TF-IDF	2	E	PR	GI
[193]	TF-IDF, Domain relevance	2	C	HR, Cellphone	
[194]	Ontology	2	E	Automobile, PR, SW	SWN, GI, OL
[195]	Ontology	2	E	MR	WN
[196]	Ontology, Maximum-Likelihood	2	E	MR	GI
[197]	PCA, SVM, LR, Bayesian Boosting, Bagged SVM	2	E	PR	
[200]	SVM	2	E	PR	
[202]	DBA, Graphical Techniques	2	E	G	CN, DBPedia, WN
[203]	DBA	2	E	MB	CN, WN, JMDict, Verbosity
[205]	Graphical techniques	2	GE	MB	SWN, SN 3
[206]	DBA	8	E	Google n-grams	SN 3, WNANRC, SAT
[207]	Ontology, DBA	4	E	PR, MR	CN
[209]	SVM, NB, J48	3	S	Facebook text	Spanish LIWC
[210]	SVM, RF	3	S	Apontador	
[211]	DBA	2	S	MB	SN 3, WeFeelFine
[212]	NB, SVM, DBA	2	E	PR	LIWC
[213]	Ontology, DBA, ELM	2	E	G	AffectiveSpace
[214]	Ontology, DBA, SVM, FCM	2	E	G	SN 3, WNA, AffectiveSpace
[216]	DBA, Ontology	2	E	PR, MR	WN, CN
[217]	Rule base classifier, NB	2	E	Dialogue	SN 3

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Summary of reviewed articles

Ref.	Concepts and techniques utilized	P	L	Type of data	Dictionary
[218]	Bootstrapping, PMI, DBA	NA	E	PR	
[220]	DBA, Binomial LR	NA	E	PR	LIWC
[221]	Product, Review & Reviewer Information	NA	E	PR	
[222]	Linear Regression	2	E	PR	
[223]	Linear Regression	NA	E	PR	
[224]	Linear Regression	NA	E	PR	
[225]	SVM	NA	E	PR	
[226]	MLP	NA	E	PR	
[227]	RFM, SVR	NA	E	PR	
[228]	RF, NB, SVM	NA	E	PR	
[229]	DBA	2	E	PR	
[231]	Linear Regression	NA	E	PR	
[232]	PU-learning	NA	E	PR	
[240]	LDA, SVM, PMI	NA	C	PR	
[241]	PageRank algorithm, DBA	NA	C	PR	
[243]	PMI-IR, RCut, Apriori Algo.	NA	C	PR	

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

Social Media Monitoring/Analysis

Existing Tools
(“Social Media Monitoring/Analysis")

- Radian 6
- Social Mention
- Overtone OpenMic
- Microsoft Dynamics Social Networking Accelerator
- SAS Social Media Analytics
- Lithium Social Media Monitoring
- RightNow Cloud Monitor

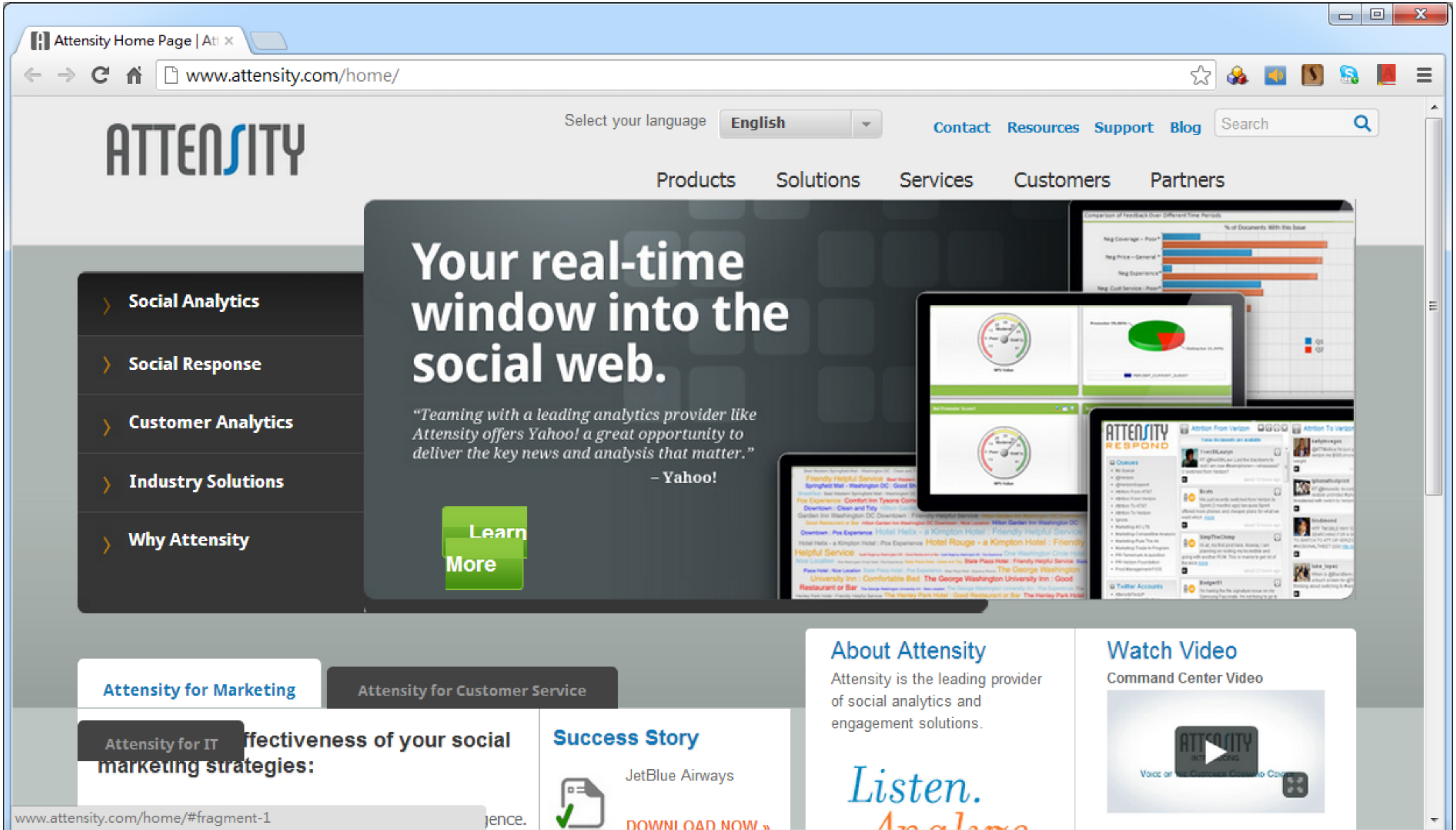
Source: Wiltrud Kessler (2012), Introduction to Sentiment Analysis

Word-of-mouth
Voice of the Customer

- 1. Attensity
  - Track social sentiment across brands and competitors
  - http://www.attensity.com/home/
- 2. Clarabridge
  - Sentiment and Text Analytics Software
  - http://www.clarabridge.com/

Attensity: Track social sentiment across brands and competitors

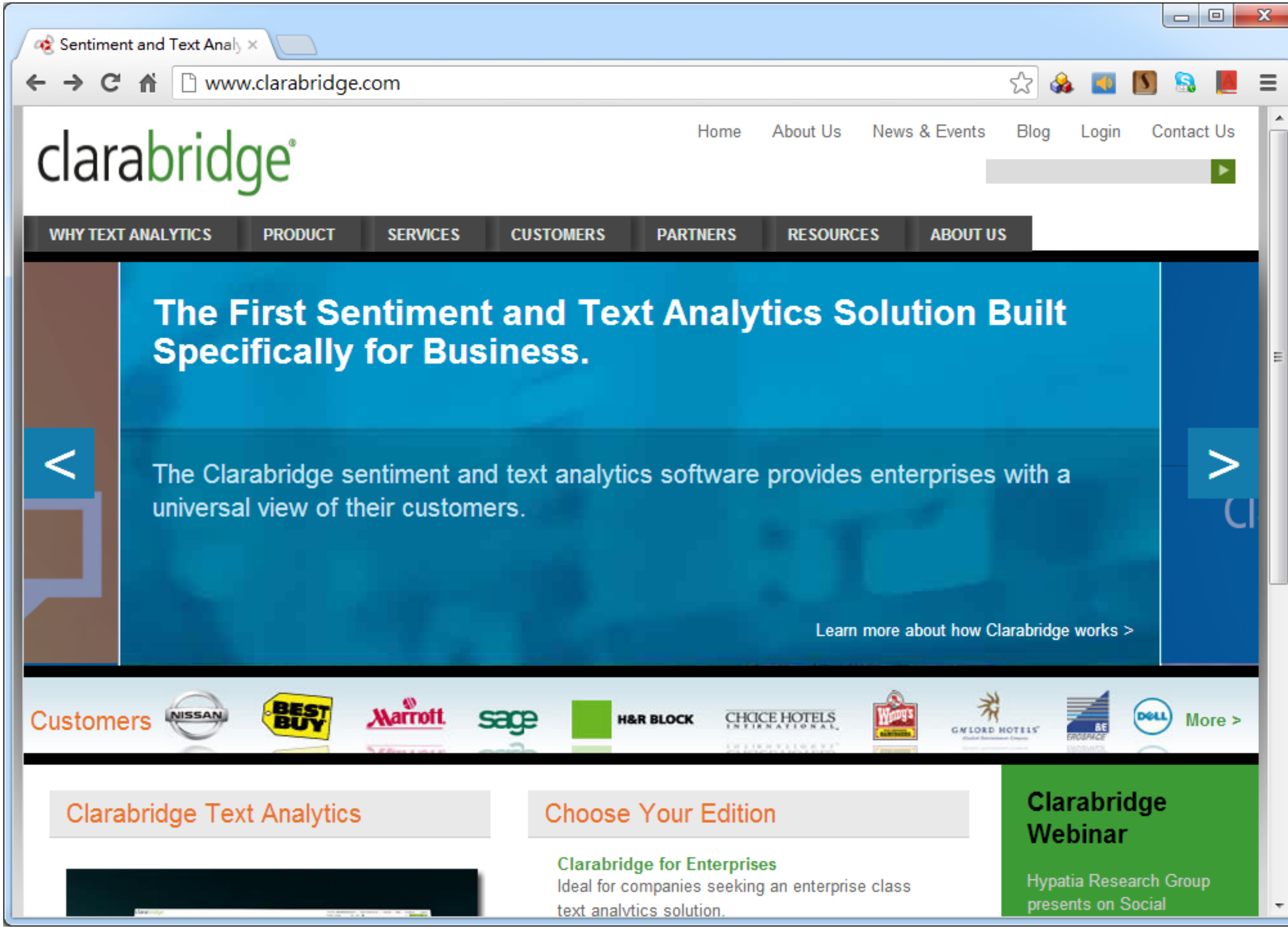
http://www.attensity.com/



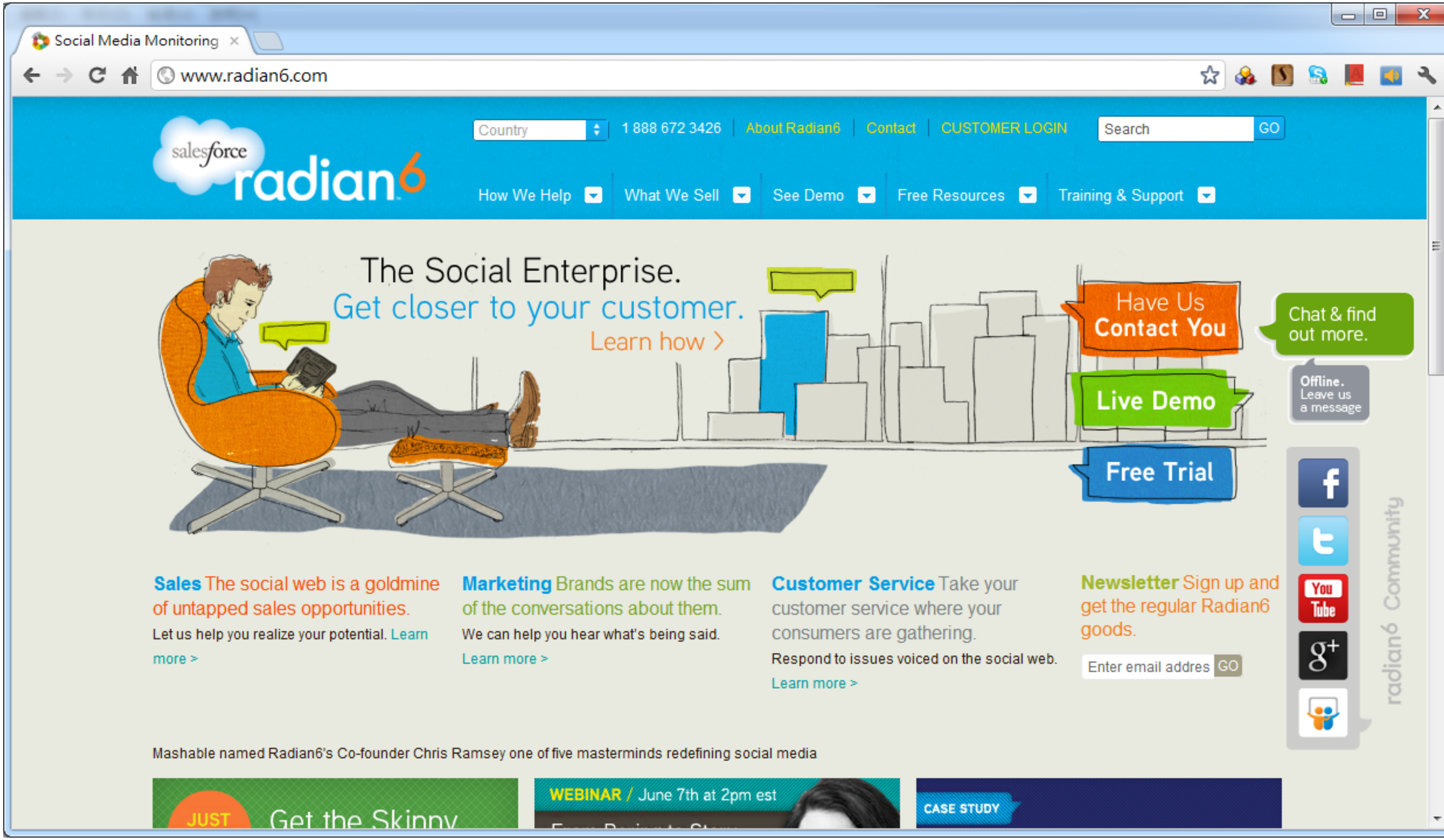
http://www.youtube.com/watch?v=4goxmBEg2Iw#!

Clarabridge: Sentiment and Text Analytics Software

http://www.clarabridge.com/

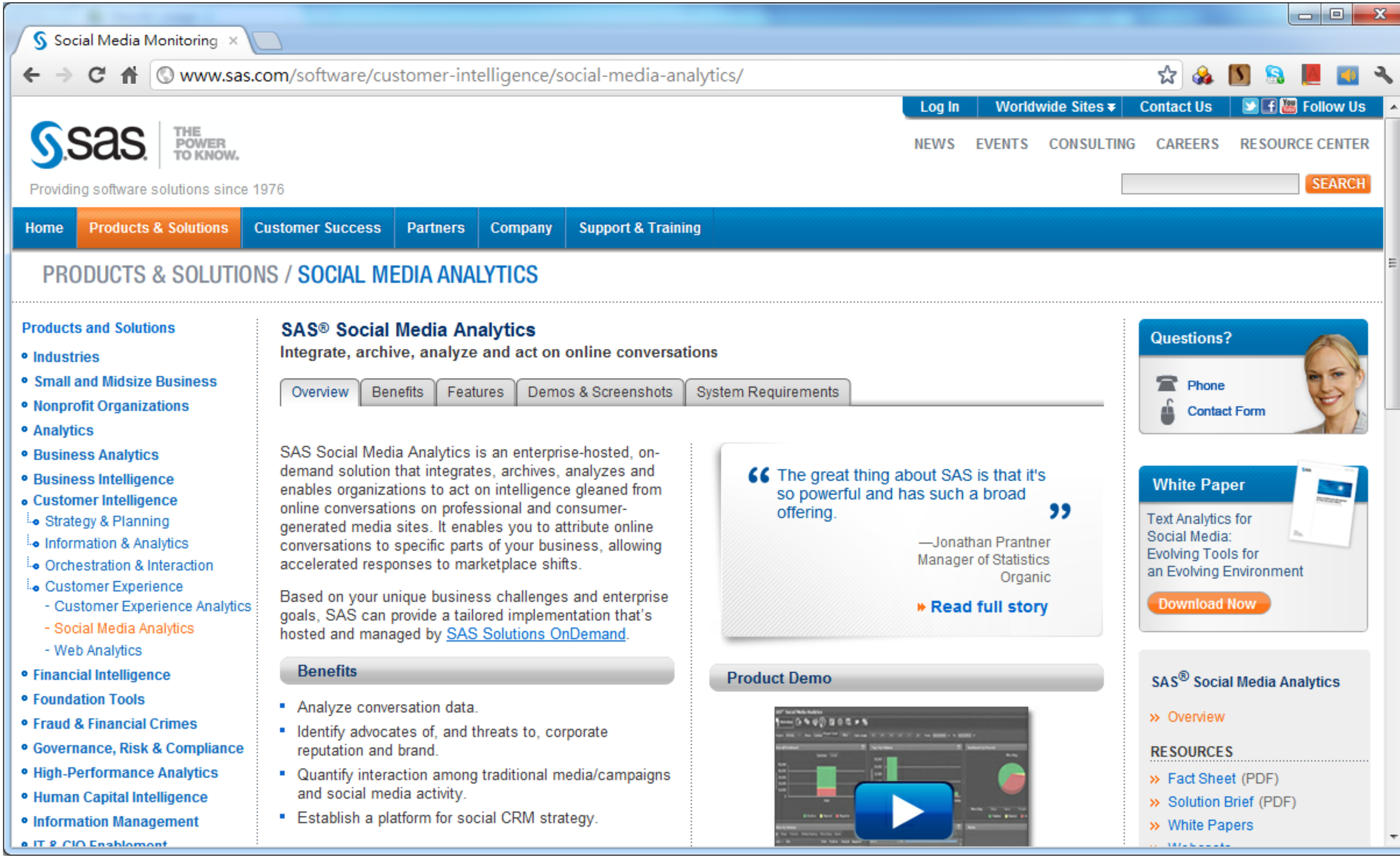


http://www.youtube.com/watch?v=IDHudt8M9P0 http://www.radian6.com/



http://www.youtube.com/watch?feature=player_embedded&v=8i6Exg3Urg0

http://www.sas.com/software/customer-intelligence/social-media-analytics/



http://www.tweetfeel.com

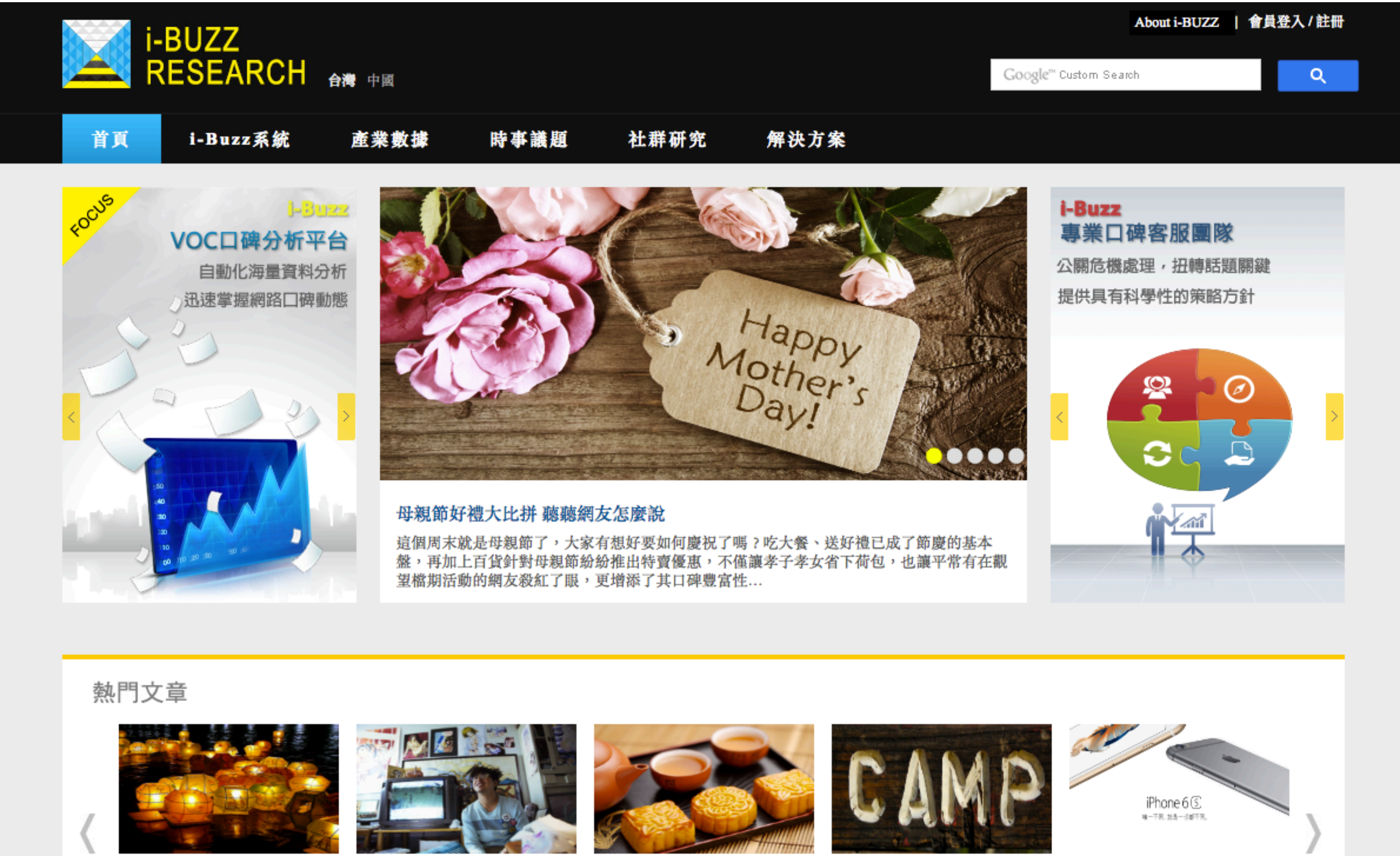
eLand

http://www.eland.com.tw/

OpView

http://www.opview.com.tw/

http://www.i-buzz.com.tw/



Resources of Opinion Mining

Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
  - http://ir.dcs.gla.ac.uk/test collections/access to data.html
- Cornell movie-review datasets
  - http://www.cs.cornell.edu/people/pabo/movie-review-data/
- Customer review datasets
  - http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip
- Multiple-aspect restaurant reviews
  - http://people.csail.mit.edu/bsnyder/naacl07
- NTCIR multilingual corpus
  - NTCIR Multilingual Opinion-Analysis Task (MOAT)

Source: Bo Pang and Lillian Lee (2008), "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval

Lexical Resources of Opinion Mining

- SentiWordnet
  - http://sentiwordnet.isti.cnr.it/
- General Inquirer
  - http://www.wjh.harvard.edu/~inquirer/
- OpinionFinder's Subjectivity Lexicon
  - http://www.cs.pitt.edu/mpqa/
- NTU Sentiment Dictionary (NTUSD)
  - http://nlg18.csie.ntu.edu.tw:8080/opinion/
- Hownet Sentiment
  - http://www.keenage.com/html/c_bulletin_2007.htm

Sentiment Analysis Resources

- Roget's Thesaurus:
  - http://thesaurus.com/Roget-alpha-index.html
- Suggested Upper Merged Ontology (SUMO)
  - http://www.adampease.org/OP/index.html

Source: Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.

NLP Toolkits

Toolkit	Language	Description
NLTK	Python	http://www.nltk.org/
OpenNLP	JAVA	https://opennlp.apache.org
CoreNLP	JAVA	http://stanfordnlp.github.io/CoreNLP/
Gensim	Python	http://radimrehurek.com/gensim/
FudanNLP	JAVA	https://code.google.com/archive/p/fudannlp/
LTP	C++/Python	http://www.ltp-cloud.com/intro/en/
NiuParser	C++	http://www.niuparser.com/index.en.html

Source: Sun, Shiliang, Chen Luo, and Junyu Chen. "A review of natural language processing techniques for opinion mining systems." Information Fusion 36 (2017): 10-25.

Annotated corpora for opinion mining

Corpora	Language	Description
MPQA opinion corpora	English	This corpus contains news articles manually annotated using an annotation scheme for opinions. Several versions annotated in different levels are provided. http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/
Movie review polarity dataset	English	The latest version of this dataset contains 1000 positive and 1000 negative processed reviews. http://www.cs.cornell.edu/people/pabo/movie-review-data/review_polarity.tar.gz
Movie review subjectivity dataset	English	This dataset includes 5000 subjective and 5000 objective processed sentences. http://www.cs.cornell.edu/people/pabo/movie-review-data/rotten_imdb.tar.gz
Multi-domain sentiment dataset	English	The dataset is constructed by Amazon product reviews for books, DVDs, electronics and kitchen appliances. Two kinds of datasets are available, one with the number of stars, the other with positive or negative labels. https://www.cs.jhu.edu/~mdredze/datasets/sentiment/

Source: Sun, Shiliang, Chen Luo, and Junyu Chen. "A review of natural language processing techniques for opinion mining systems." Information Fusion 36 (2017): 10-25.

Sentiment lexicons for opinion mining

Lexicon	Language	Description
Bing Liu's Opinion Lexicon	English	The latest version of this lexicon includes 4,783 negative words and 2,006 positive ones. http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
MPQA Subjectivity Lexicon	English	This lexicon includes 8,222 words with their subjectivities (strong or weak), POS tags and polarities. http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
SentiWordNet	English	SentiWordNet associates words to numerical scores ranging in [0.0, 1.0] which indicate the positivity, negativity and neutrality. For each word, the three scores sum up to 1.0. http://sentiwordnet.isti.cnr.it/
Harvard General Inquirer	English	Harvard General Inquirer contains 182 categories including positive and negative indicators. 1915 positive words and 2291 negative words are marked. http://www.wjh.harvard.edu/~inquirer/
LIWC	English	Linguistic Inquiry and Word Counts (LIWC) provides a lot of categorized regular expressions including some sentiment related categories such as "Negate" and "Anger". http://liwc.wpengine.com
HowNet	Chinese & English	HowNet provides a Chinese/English vocabulary for sentiment analysis, including 8942 Chinese entries and 8945 English entries. http://www.keenage.com/html/e_index.html
NTUSD	Chinese	NTU Sentiment Dictionary provides 2812 positive words and 8276 negative words in both simplified and traditional Chinese. http://academiasinicanlplab.github.io/

Source: Sun, Shiliang, Chen Luo, and Junyu Chen. "A review of natural language processing techniques for opinion mining systems." Information Fusion 36 (2017): 10-25.

Example of SentiWordNet

POS ID PosScore NegScore SynsetTerms Gloss

a 00217728 0.75 0 beautiful#1 delighting the senses or exciting intellectual or emotional admiration; "a beautiful child"; "beautiful country"; "a beautiful painting"; "a beautiful theory"; "a beautiful party“

a 00227507 0.75 0 best#1 (superlative of `good') having the most positive qualities; "the best film of the year"; "the best solution"; "the best time for planting"; "wore his best suit“

r 00042614 0 0.625 unhappily#2 sadly#1 in an unfortunate way; "sadly he died before he could see his grandchild“

r 00093270 0 0.875 woefully#1 sadly#3 lamentably#1 deplorably#1 in an unfortunate or deplorable manner; "he was sadly neglected"; "it was woefully inadequate“

r 00404501 0 0.25 sadly#2 with sadness; in a sad manner; "`She died last night,' he said sadly"

《知網》情感分析用詞語集（beta版）

- "中英文情感分析用詞語集"
  - 包含詞語約 17887
- "中文情感分析用詞語集"
  - 包含詞語約 9193
- "英文情感分析用詞語集"
  - 包含詞語 8945

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

中文正面情感詞語	836
中文負面情感詞語	1254
中文正面評價詞語	3730
中文負面評價詞語	3116
中文程度級別詞語	219
中文主張詞語	38
Total	9193

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “正面情感”詞語
  - 如：
    愛，讚賞，快樂，感同身受，好奇，喝彩，魂牽夢縈，嘉許...
- “負面情感”詞語
  - 如：
    哀傷，半信半疑，鄙視，不滿意，不是滋味兒，後悔，大失所望...

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “正面評價”詞語
  - 如：
    不可或缺，部優，才高八斗，沉魚落雁，催人奮進，動聽，對勁兒 ...
- “負面評價”詞語
  - 如：
    醜，苦，超標，華而不實，荒涼，混濁，畸輕畸重，價高，空洞無物 ...

Source: http://www.keenage.com/html/c_bulletin_2007.htm

中文情感分析用詞語集

- “程度級別”詞語
  - 1. “極其|extreme / 最|most”
    - 非常，極，極度，無以倫比，最為
  - 2. “很|very”
    - 多麼，分外，格外，著實
  - …
- “主張”詞語
  - 1. {perception|感知}
    - 感覺，覺得，預感
  - 2. {regard|認為}
    - 認為，以為，主張

Source: http://www.keenage.com/html/c_bulletin_2007.htm

Opinion Spam Detection

Opinion Spam Detection

- Opinion Spam Detection: Detecting Fake Reviews and Reviewers
  - Spam Review
  - Fake Review
  - Bogus Review
  - Deceptive review
  - Opinion Spammer
  - Review Spammer
  - Fake Reviewer
  - Shill (Stooge or Plant)

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Opinion Spamming

- Opinion Spamming
  - "illegal" activities
    - e.g., writing fake reviews, also called shilling
  - try to mislead readers or automated opinion mining and sentiment analysis systems by giving undeserving positive opinions to some target entities in order to promote the entities and/or by giving false negative opinions to some other entities in order to damage their reputations.

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Forms of Opinion spam

- fake reviews (also called bogus reviews)
- fake comments
- fake blogs
- fake social network postings
- deceptions
- deceptive messages

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Fake Review Detection

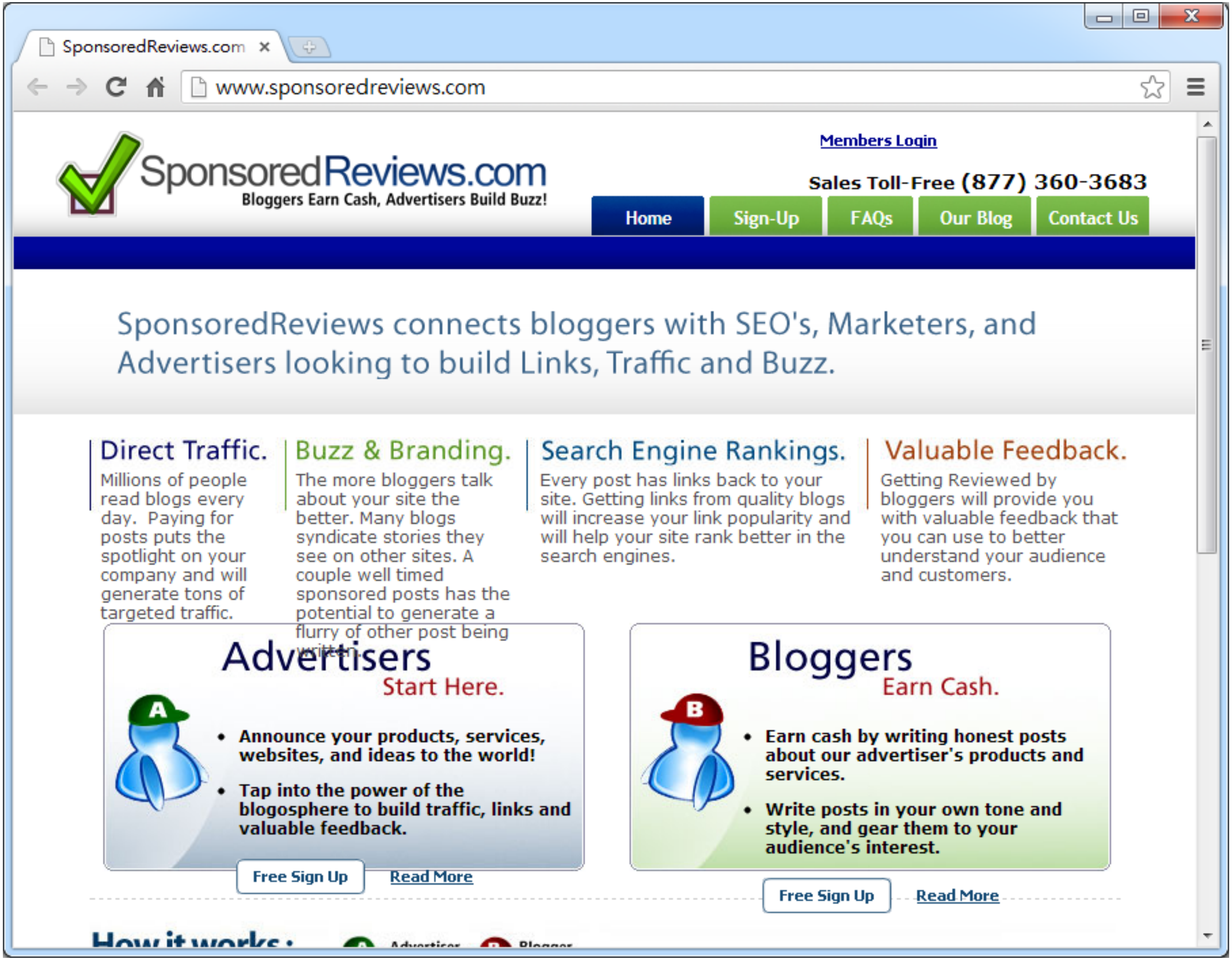
- Methods
  - supervised learning
  - pattern discovery
  - graph-based methods
  - relational modeling
- Signals
  - Review content
  - Reviewer abnormal behaviors
  - Product related features
  - Relationships

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Professional Fake Review Writing Services (some Reputation Management companies)

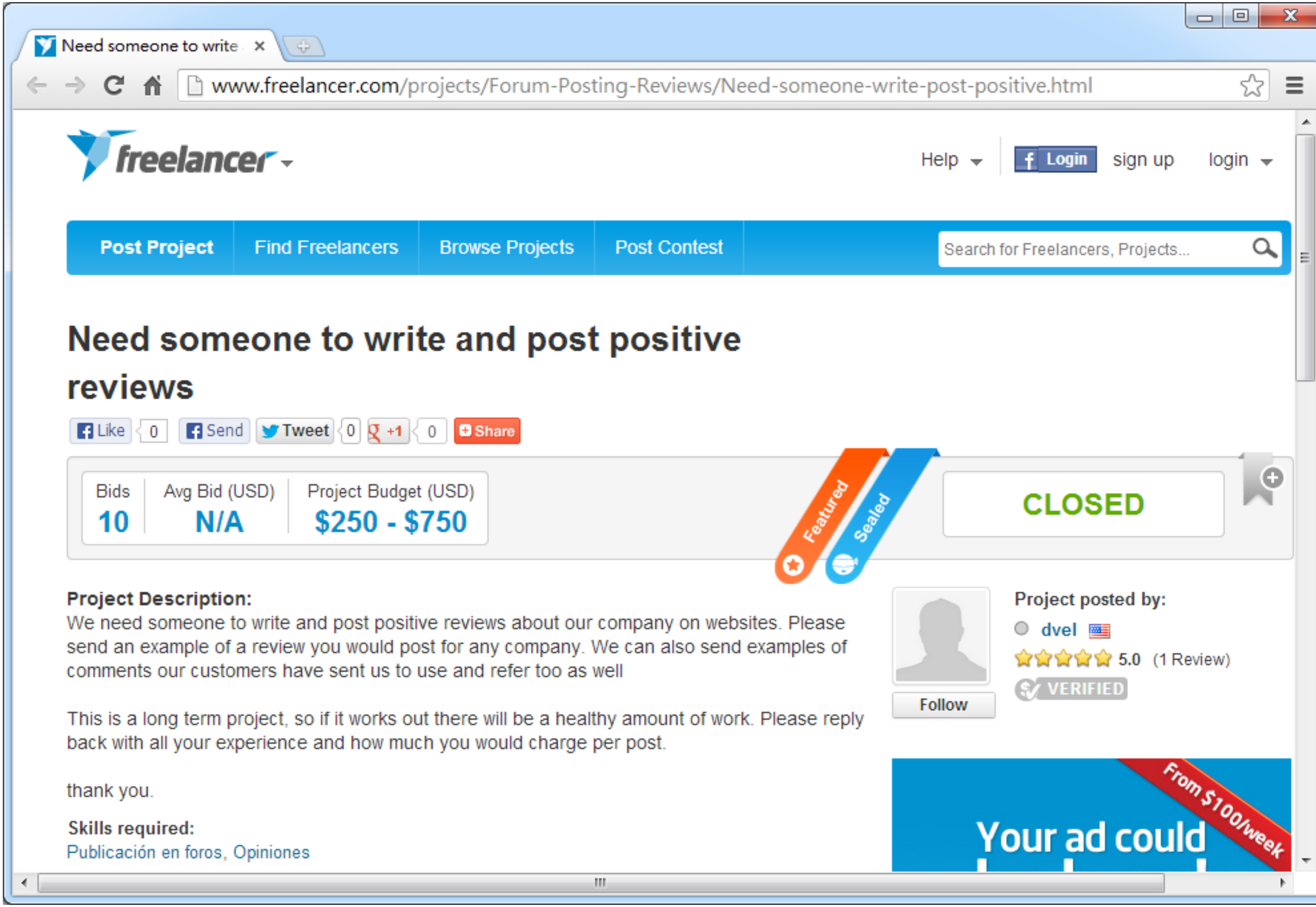
- Post positive reviews
- Sponsored reviews
- Pay per post
- Need someone to write positive reviews about our company (budget: $250-$750 USD)
- Fake review writer
- Product review writer for hire
- Hire a content writer
- Fake Amazon book reviews (hiring book reviewers)
- People are just having fun (not serious)

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html



Source: http://www.sponsoredreviews.com/

Source: https://payperpost.com/



Source:http://www.freelancer.com/projects/Forum-Posting-Reviews/Need-someone-write-post-positive.html

Papers on Opinion Spam Detection

1. Jing Wang, Clement. T. Yu, Philip S. Yu, Bing Liu, Weiyi Meng. “Diversionary comments under blog posts." Accepted. ACM Transactions on the Web (TWEB), 2015.
2. Huayi Li, Zhiyuan Chen, Arjun Mukherjee, Bing Liu and Jidong Shao. "Analyzing and Detecting Opinion Spam on a Large-scale Dataset via Temporal and Spatial Patterns." Short paper at ICWSM-2015, 2015.
3. Huayi Li, Arjun Mukherjee, Bing Liu, Rachel Kornfieldz and Sherry Emery. Detecting Campaign Promoters on Twitter using Markov Random Fields. to appear in Proceedings of IEEE International Conference on Data Mining (ICDM-2014), December 14-17, 2014.
4. Huayi Li, Zhiyuan Chen, Bing Liu, Xiaokai Wei and Jidong Shao. Spotting Fake Reviews via Collective Positive-Unlabeled Learning. to appear in Proceedings of IEEE International Conference on Data Mining (ICDM-2014, short paper), December 14-17, 2014.
5. Tieyun Qian, Bing Liu. Identifying Multiple Userids of the Same Author. To appear in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2013), October 18-21, 2013, Seattle, USA.
6. Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. Spotting Opinion Spammers using Behavioral Footprints. To appear in Proceedings of SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2013), August 11-14 2013 in Chicago, USA.
7. Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie Glance. What Yelp Fake Review Filter Might Be Doing. Proceedings of The International AAAI Conference on Weblogs and Social Media (ICWSM-2013), July 8-10, 2013, Boston, USA.
8. Geli Fei, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. Exploiting Burstiness in Reviews for Review Spammer Detection. Proceedings of The International AAAI Conference on Weblogs and Social Media (ICWSM-2013), July 8-10, 2013, Boston, USA.

Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Papers on Opinion Spam Detection

9. Arjun Mukherjee, Bing Liu, and Natalie Glance. Spotting Fake Reviewer Groups in Consumer Reviews. International World Wide Web Conference (WWW-2012), Lyon, France, April 16-20, 2012.
10. Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu. Identify Online Store Review Spammers via Social Review Graph. ACM Transactions on Intelligent Systems and Technology, accepted for publication, 2011.
11. Guan Wang, Sihong Xie, Bing Liu, Philip S. Yu. Review Graph based Online Store Review Spammer Detection. ICDM-2011, 2011.
12. Arjun Mukherjee, Bing Liu, Junhui Wang, Natalie Glance, Nitin Jindal. Detecting Group Review Spam. WWW-2011 poster paper, 2011.
13. Nitin Jindal, Bing Liu and Ee-Peng Lim. "Finding Unusual Review Patterns Using Unexpected Rules" Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM-2010, short paper), Toronto, Canada, Oct 26 - 30, 2010.
14. Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu and Hady Lauw. "Detecting Product Review Spammers using Rating Behaviors." Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM-2010, full paper), Toronto, Canada, Oct 26 - 30, 2010.
15. Nitin Jindal and Bing Liu. "Opinion Spam and Analysis." Proceedings of First ACM International Conference on Web Search and Data Mining (WSDM-2008), Feb 11-12, 2008, Stanford University, Stanford, California, USA.
16. Nitin Jindal and Bing Liu. "Review Spam Detection." Proceedings of WWW-2007 (poster paper), May 8-12, Banff, Canada.

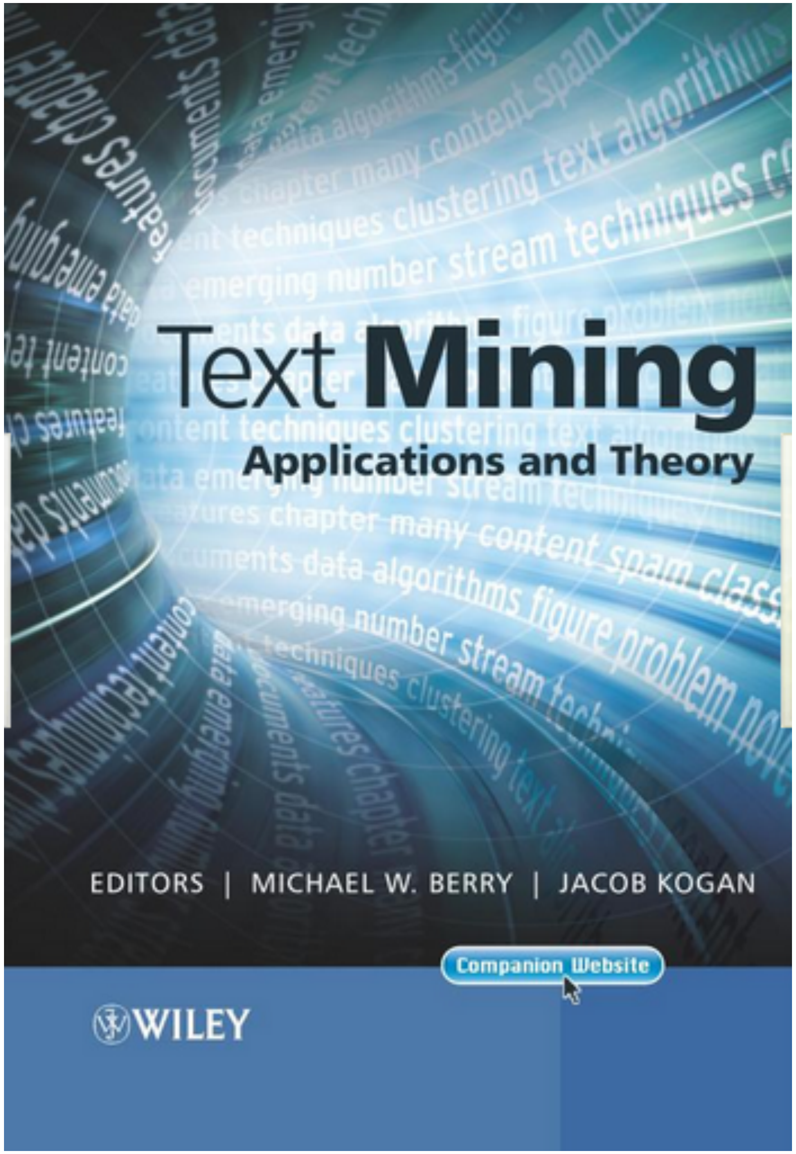
Source: http://www.cs.uic.edu/~liub/FBS/fake-reviews.html

Text Mining and Analytics Technology

Text Mining Techniques

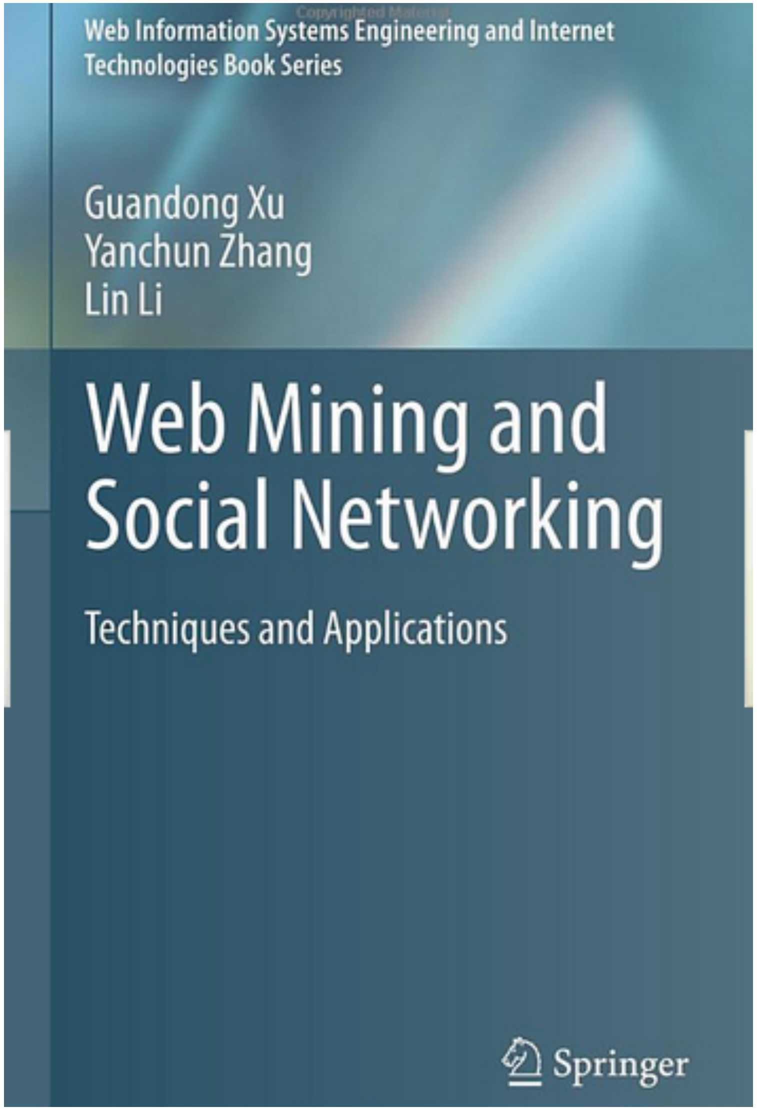
Natural Language Processing (NLP)

Text Mining



http://www.amazon.com/Text-Mining-Applications-Michael-Berry/dp/0470749822/

Web Mining and Social Networking



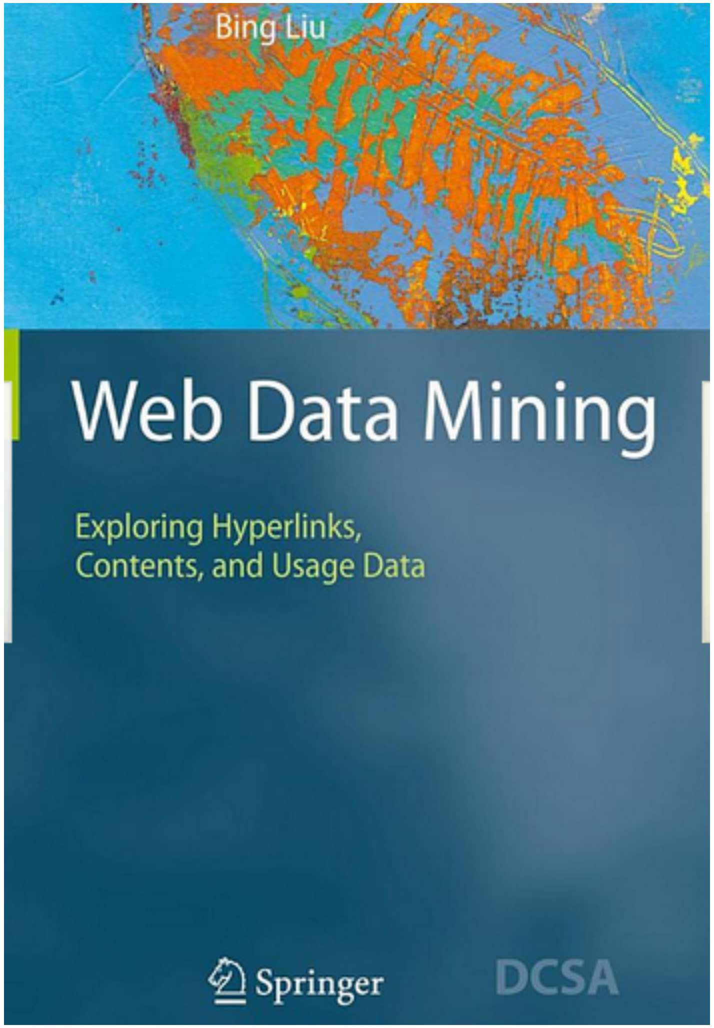
http://www.amazon.com/Web-Mining-Social-Networking-Applications/dp/1441977341

Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

http://www.amazon.com/Mining-Social-Web-Analyzing-Facebook/dp/1449388345

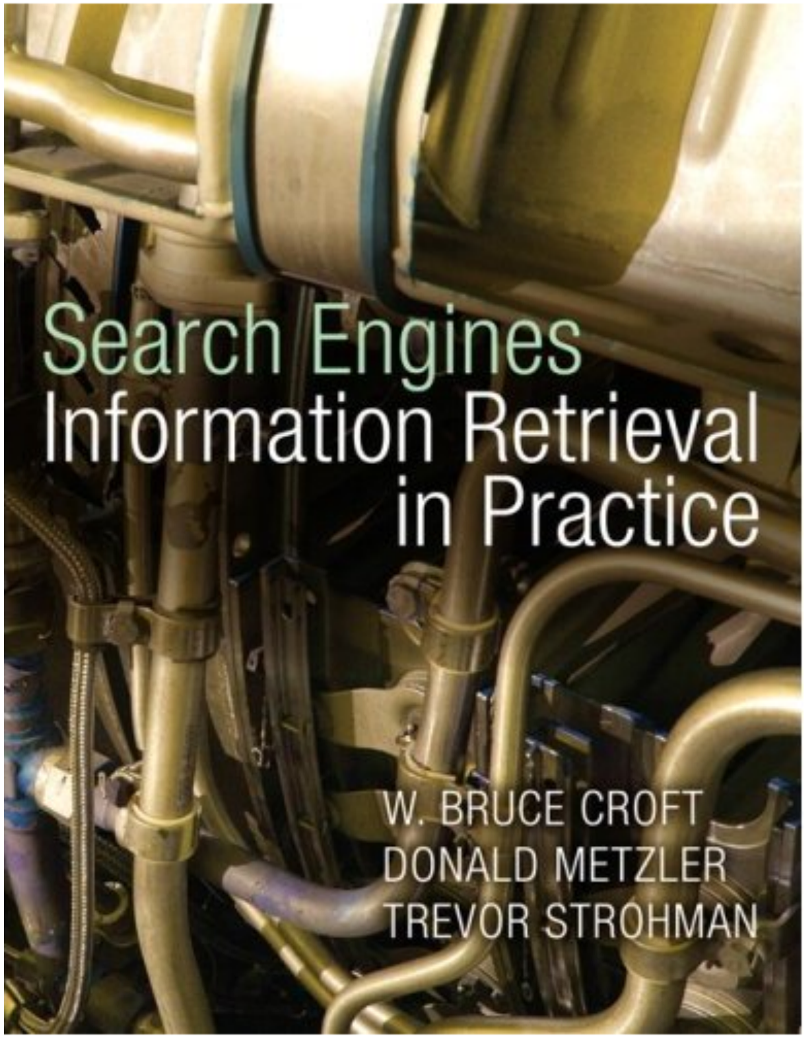
Web Data Mining:

Exploring Hyperlinks, Contents, and Usage Data



http://www.amazon.com/Web-Data-Mining-Data-Centric-Applications/dp/3540378812

Search Engines: Information Retrieval in Practice

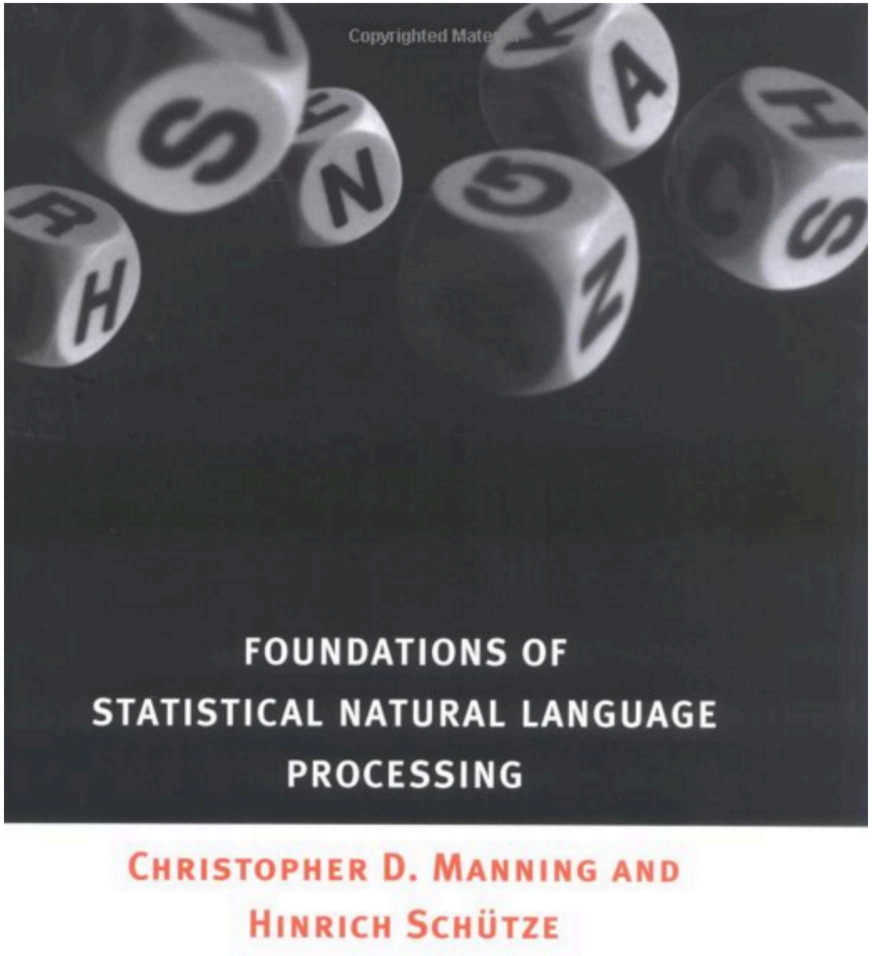


http://www.amazon.com/Search-Engines-Information-Retrieval-Practice/dp/0136072240

Christopher D. Manning and Hinrich Schütze (1999),

Foundations of
Statistical Natural Language Processing,

The MIT Press

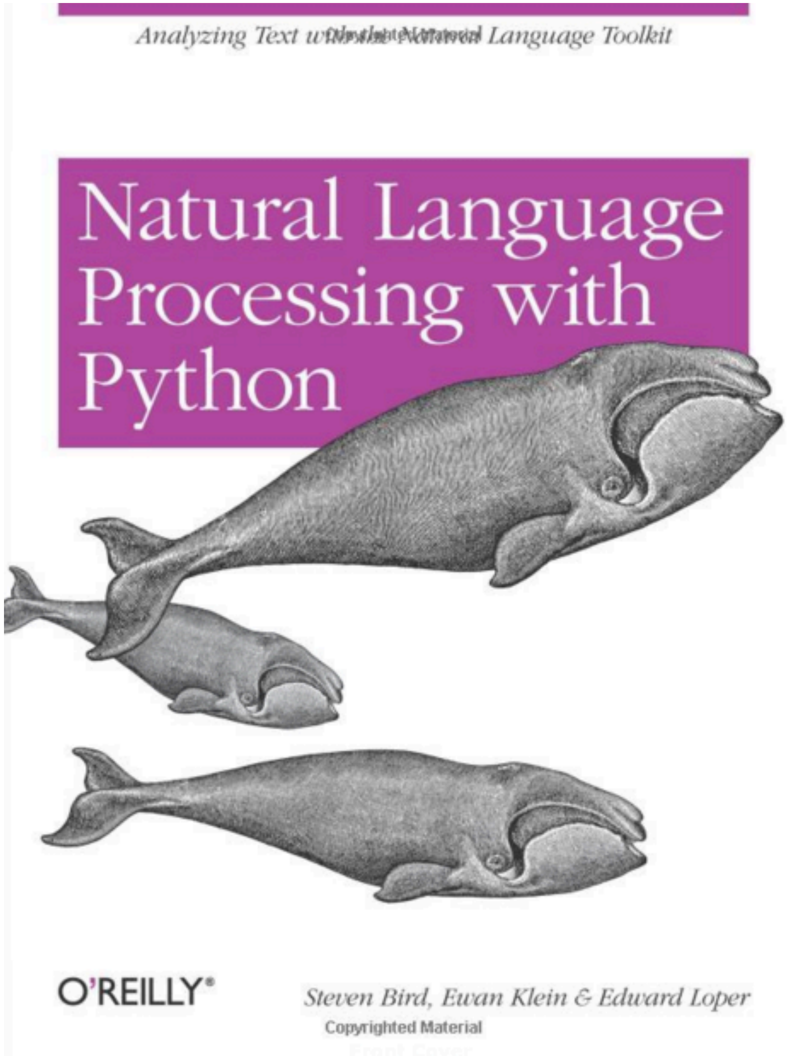


http://www.amazon.com/Foundations-Statistical-Natural-Language-Processing/dp/0262133601

Steven Bird, Ewan Klein and Edward Loper (2009),

Natural Language Processing with Python,

O'Reilly Media



http://www.amazon.com/Natural-Language-Processing-Python-Steven/dp/0596516495

Natural Language Processing with Python
– Analyzing Text with the Natural Language Toolkit

www.nltk.org/book/

Natural Language Processing with Python

– Analyzing Text with the Natural Language Toolkit

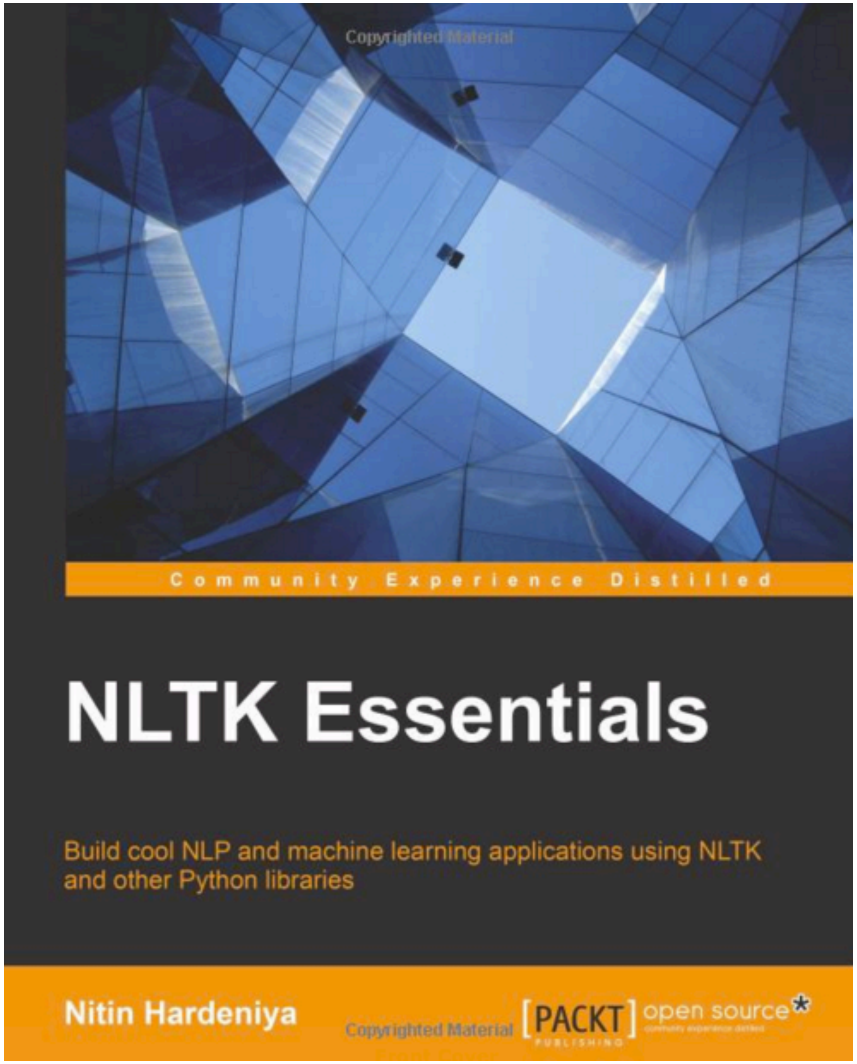
Steven Bird, Ewan Klein, and Edward Loper

The NLTK book is currently being updated for Python 3 and NLTK 3. This is work in progress; chapters that still need to be updated are indicated. The first edition of the book, published by O'Reilly, is available at http://nltk.org/book_1ed/. A second edition of the book is anticipated in early 2016.

Please post any questions about the materials to the nltk-users mailing list. Please report any errors on the issue tracker.

http://www.nltk.org/book/

Nitin Hardeniya (2015), NLTK Essentials, Packt Publishing



http://www.amazon.com/NLTK-Essentials-Nitin-Hardeniya/dp/1784396907

Text Mining
(text data mining)

the process of
deriving
high-quality information
from text

http://en.wikipedia.org/wiki/Text_mining

Typical Text Mining Tasks

- Text categorization
- Text clustering
- Concept/entity extraction
- Production of granular taxonomies
- Sentiment analysis
- Document summarization
- Entity relation modeling
  - i.e., learning relations between named entities.

http://en.wikipedia.org/wiki/Text_mining

Web Mining

- Web mining
  - discover useful information or knowledge from the **Web hyperlink structure, page content, and usage data.**
- Three types of web mining tasks
  - Web structure mining
  - Web content mining
  - Web usage mining

Source: Bing Liu (2009) Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data

Text Mining Concepts

- 85-90 percent of all corporate data is in some kind of unstructured form (e.g., text)
- Unstructured corporate data is doubling in size every 18 months
- Tapping into these information sources is not an option, but a need to stay competitive
- Answer: text mining
  - A semi-automated process of extracting knowledge from unstructured data sources
  - a.k.a. text data mining or knowledge discovery in textual databases

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Data Mining versus Text Mining

- Both seek for novel and useful patterns
- Both are semi-automated processes
- Difference is the nature of the data:
  - Structured versus unstructured data
  - Structured data: in databases
  - Unstructured data: Word documents, PDF files, text excerpts, XML files, and so on
- Text mining – first, impose structure to the data, then mine the structured data

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Text Mining Concepts

- Benefits of text mining are obvious especially in text-rich data environments
  - e.g., law (court orders), academic research (research articles), finance (quarterly reports), medicine (discharge summaries), biology (molecular interactions), technology (patent files), marketing (customer comments), etc.
- Electronic communization records (e.g., Email)
  - Spam filtering
  - Email prioritization and categorization
  - Automatic response generation

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Text Mining Application Area

- Information extraction
- Topic tracking
- Summarization
- Categorization
- Clustering
- Concept linking
- Question answering

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Text Mining Terminology

- Unstructured or semistructured data
- Corpus (and corpora)
- Terms
- Concepts
- Stemming
- Stop words (and include words)
- Synonyms (and polysemes)
- Tokenizing

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Text Mining Terminology

- Term dictionary
- Word frequency
- Part-of-speech tagging (POS)
- Morphology
- Term-by-document matrix (TDM)
  – Occurrence matrix
- Singular Value Decomposition (SVD)
  – Latent Semantic Indexing (LSI)

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Natural Language Processing (NLP)

- Structuring a collection of text
  - Old approach: bag-of-words
  - New approach: natural language processing
- NLP is …
  - a very important concept in text mining
  - a subfield of artificial intelligence and computational linguistics
  - the studies of "understanding" the natural human language
- Syntax versus semantics based text mining

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Natural Language Processing (NLP)

- What is “Understanding” ?
  - Human understands, what about computers?
  - Natural language is vague, context driven
  - True understanding requires extensive knowledge of a topic

  - Can/will computers ever understand natural language the same/accurate way we do?

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Natural Language Processing (NLP)

- Challenges in NLP
  - Part-of-speech tagging
  - Text segmentation
  - Word sense disambiguation
  - Syntax ambiguity
  - Imperfect or irregular input
  - Speech acts
- Dream of AI community
  - to have algorithms that are capable of automatically reading and obtaining knowledge from text

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

Natural Language Processing (NLP)

- WordNet
  - A laboriously hand-coded database of English words, their definitions, sets of synonyms, and various semantic relations between synonym sets
  - A major resource for NLP
  - Need automation to be completed
- Sentiment Analysis
  - A technique used to detect favorable and unfavorable opinions toward specific products and services
  - CRM application

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

NLP Task Categories

- Information retrieval (IR)
- Information extraction (IE)
- Named-entity recognition (NER)
- Question answering (QA)
- Automatic summarization
- Natural language generation and understanding (NLU)
- Machine translation (ML)
- Foreign language reading and writing
- Speech recognition
- Text proofing
- Optical character recognition (OCR)

Source: Turban et al. (2011), Decision Support and Business Intelligence Systems

CKIP 中研院中文斷詞系統

http://ckipsvr.iis.sinica.edu.tw/

中文斷詞系統

相關系統： 斷詞系統 | 剖析系統 | 詞首詞尾 | 平衡語料庫 | 廣義知網 | 句結構樹庫 | 錯字偵測

- 簡介
- 未知詞擷取做法
- 詞類標記列表
- 線上展示
- 線上服務申請
- 線上資源
- 公告
- 聯絡我們

隱私權聲明 | 版權聲明

數位典藏國家型科技計畫 National Digital Archives Program

線上展示使用簡化詞類進行斷詞標記，僅供參考並且系統不再進行更新。線上服務斷詞和授權mirror site僅提供精簡詞類，結果也與舊版的展示系統不同。

自 2014/01/06 起，本斷詞系統已經處理過 929135 篇文章

送出 清除

歐巴馬是美國的一位總統

歐巴馬是美國的一位總統

文章的文字檔
擷取未知詞過程
包含未知詞的斷詞標記結果
未知詞列表

歐巴馬(Nb) 是(SHI) 美國(Nc) 的(DE) 一(Neu) 位(Nf) 總統(Na)

中文文字處理: 中文斷詞

抗氣候變遷 白宮籲採緊急行動

中央通訊社 THE CENTRAL NEWS AGENCY 中央社 – 2014年5月6日 下午10:58

（中央社華盛頓6日綜合外電報導）白宮今天公布全球暖化對全美及美國經濟關鍵產業造成何種衝擊的新報告，呼籲採取緊急行動對抗氣候變遷。

這份為期4年的調查警告，極端氣候事件將對住家、基礎設施及產業帶來嚴重威脅。

美國總統歐巴馬2008年當選總統時曾在競選造勢時誓言，要讓美國成為對抗氣候變遷與相關「安全威脅」的領頭羊。

但歐巴馬在任上一直未能說服美國國會採取重大行動。

在本週對這項議題採取的新作為中，歐巴馬今天將與數名氣象學家接受電視訪問，討論美國全國氣候評估第3版調查結果。

美國數百名來自政府與民間的頂尖氣候科學家及技術專家，共同投入這項研究，檢視氣候變遷對當今帶來的衝擊並預測將對下個世紀帶來何種影響。

研究人員警告，加州可能發生旱災、奧克拉荷馬州發生草原大火，東岸則可能遭遇海平面上升，尤其佛羅里達，而這些事件多為人類造成。

海平面上升也將吞噬密西西比等低窪地區。

至於超過8000萬人居住且擁有全美部分成長最快都會區的東南部與加勒比海區，「海平面上升加上其他與氣候變遷有關的衝擊，以及地層下陷等既有問題，將對經濟和生態帶來重大影響」。

抗氣候變遷 白宮籲採緊急行動
中央社中央社 – 2014年5月6日 下午10:58
（中央社華盛頓6日綜合外電報導）白宮今天公布全球暖化對全美及美國經濟關鍵產業造成何種衝擊的新報告，呼籲採取緊急行動對抗氣候變遷。
這份為期4年的調查警告，極端氣候事件將對住家、基礎設施及產業帶來嚴重威脅。
美國總統歐巴馬2008年當選總統時曾在競選造勢時誓言，要讓美國成為對抗氣候變遷與相關「安全威脅」的領頭羊。
但歐巴馬在任上一直未能說服美國國會採取重大行動。
在本週對這項議題採取的新作為中，歐巴馬今天將與數名氣象學家接受電視訪問，討論美國全國氣候評估第3版調查結果。
美國數百名來自政府與民間的頂尖氣候科學家及技術專家，共同投入這項研究，檢視氣候變遷對當今帶來的衝擊並預測將對下個世紀帶來何種影響。
研究人員警告，加州可能發生旱災、奧克拉荷馬州發生草原大火，東岸則可能遭遇海平面上升，尤其佛羅里達，而這些事件多為人類造成。
海平面上升也將吞噬密西西比等低窪地區。
至於超過8000萬人居住且擁有全美部分成長最快都會區的東南部與加勒比海區，「海平面上升加上其他與氣候變遷有關的衝擊，以及地層下陷等既有問題，將對經濟和生態帶來重大影響」。
報告並說：「過去被認為是遙遠未來議題的氣候變遷，已著實成為當前議題。」（譯者：中央社蔡佳伶）1030506

https://tw.news.yahoo.com/%E6%8A%97%E6%B0%A3%E5%80%99%E8%AE%8A%E9%81%B7-%E7%99%BD%E5%AE%AE%E7%B1%B2%E6%8E%A1%E7%B7%8A%E6%80%A5%E8%A1%8C%E5%8B%95-145804493.html

CKIP 中研院中文斷詞系統

http://ckipsvr.iis.sinica.edu.tw/

中文斷詞系統

相關系統： 斷詞系統 | 剖析系統 | 詞首詞尾 | 平衡語料庫 | 廣義知網 | 句結構樹庫 | 錯字偵測

- 簡介
- 未知詞擷取做法
- 詞類標記列表
- 線上展示
- 線上服務申請
- 線上資源
- 公告
- 聯絡我們

隱私權聲明 | 版權聲明

數位典藏國家型科技計畫
National Digital Archives Program

線上展示使用簡化詞類進行斷詞標記，僅供參考並且系統不再進行更新。線上服務斷詞和授權mirror site僅提供精簡詞類，結果也與舊版的展示系統不同。

自 2014/01/06 起，本斷詞系統已經處理過 929136 篇文章

送出 清除

抗氣候變遷 白宮籲採緊急行動
中央社中央社 – 2014年5月6日 下午10:58
（中央社華盛頓6日綜合外電報導）白宮今天公布全球暖化對全美及美國經濟關鍵產業造成何種衝擊的新報告，呼籲採取緊急行動對抗氣候變遷。
這份為期4年的調查警告，極端氣候事件將對住家、基礎設施及產業帶來嚴重威脅。
美國總統歐巴馬2008年當選總統時曾在競選造勢時誓言，要讓美國成為對抗氣候變遷與相關「安全威脅」的領頭羊。
但歐巴馬在任上一直未能說服美國國會採取重大行動。
在本週對這項議題採取的新作為中，歐巴馬今天將與數名氣象學家接受電視訪問，討論美國全國氣候評估第3版調查結果。
美國數百名來自政府與民間的頂尖氣候科學家及技術專家，共同投入這項研究，檢視氣候變遷對當今帶來的衝擊並預測將對下個世紀帶來何種影響。
研究人員警告，加州可能發生旱災、奧克拉荷馬州發生草原大火，東岸則可能遭遇海平面上升，尤其佛羅里達，而這些事件多為人類造成。
海平面上升也將吞噬密西西比等低窪地區。
至於超過8000萬人居住且擁有全美部分成長最快都會區的東南部與加勒比海區，「海平面上升加上其他與氣候變遷有關的衝擊，以及地層下陷等既有問題，將對經濟和生態帶來重大影響」。
報告並說：「過去被認為是遙遠未來議題的氣候變遷，已著實成為當前議題。」（譯者：中央社蔡佳伶）1030506

CKIP 中研院中文斷詞系統

http://ckipsvr.iis.sinica.edu.tw/

中文斷詞系統

相關系統： 斷詞系統 | 剖析系統 | 詞首詞尾 | 平衡語料庫 | 廣義知網 | 句結構樹庫 | 錯字偵測

- 簡介
- 未知詞擷取做法
- 詞類標記列表
- 線上展示
- 線上服務申請
- 線上資源
- 公告
- 聯絡我們

隱私權聲明 | 版權聲明

抗(VJ) 氣候(Na) 變遷(VH) 白宮(Nc) 籲(VE) 採(VC) 緊急(VH) 行動(Na) 中央社(Nc) 中央社(Nc) ２０１４年(Nd) ５月(Nd) ６日(Nd) 下午(Nd) １

５８(Neu) （(PARENTHESISCATEGORY) 中央社(Nc) 華盛頓(Nc) ６日(Nd) 綜合(A) 外電(Na) 報導(VE) ）(PARENTHESISCATEGORY) 白宮(Nc) 今天(Nd

呼籲(VE) 採取(VC) 緊急(VH) 行動(Na) 對抗(VC) 氣候(Na) 變遷(VH) 。(PERIODCATEGORY)

這(Nep) 份(Nf) 為期(VH) ４年(Nd) 的(DE) 調查(VE) 警告(VE) ，(COMMACATEGORY)

極端(VH) 氣候(Na) 事件(Na) 將(D) 對(P) 住家(Na) 、(PAUSECATEGORY) 基礎(VH) 設施(Na) 及(Caa) 產業(Na) 帶來(VC) 嚴重(VH) 威脅(Na) 。

美國(Nc) 總統(Na) 歐巴馬(Nb) ２００８年(Nd) 當選(VG) 總統(Na) 時(Ng) 曾(D) 在(P) 競選(VC) 造勢(VB) 時(Ng) 誓言(VE) ，(COMMACATEGORY

要(D) 讓(VL) 美國(Nc) 成為(VG) 對抗(VC) 氣候(Na) 變遷(VH) 與(Caa) 相關(VH) 「(PARENTHESISCATEGORY) 安全(VH) 威脅(Na) 」(PARENTHES

但(Cbb) 歐巴馬(Nb) 在任(VH) 上(Ng) 一直(D) 未(D) 能(D) 說服(VF) 美國(Nc) 國會(Nc) 採取(VC) 重大(VH) 行動(Na) 。(PERIODCATEGORY)

在(P) 本(Nes) 週(Nf) 對(P) 這(Nep) 項(Nf) 議題(Na) 採取(VC) 的(DE) 新作(Na) 為(P) 中(Ncd) ，(COMMACATEGORY)

歐巴馬(Nb) 今天(Nd) 將(D) 與(P) 數(Neu) 名(Nf) 氣象學家(Na) 接受(VC) 電視(Na) 訪問(VC) ，(COMMACATEGORY)

討論(VE) 美國(Nc) 全國(Nc) 氣候(Na) 評估(VE) 第３(Neu) 版(Na) 調查(VE) 結果(Dk) 。(PERIODCATEGORY)

美國(Nc) 數百(Neu) 名(Nf) 來自(VJ) 政府(Na) 與(Caa) 民間(Nc) 的(DE) 頂尖(VH) 氣候(Na) 科學家(Na) 及(Caa) 技術(Na) 專家(Na) ，(COMM

共同(A) 投入(VC) 這(Nep) 項(Nf) 研究(Na) ，(COMMACATEGORY)

檢視(VC) 氣候(Na) 變遷(VH) 對(P) 當今(Nd) 帶來(VC) 的(DE) 衝擊(Na) 並(D) 預測(VE) 將(D) 對(P) 下(Nes) 個(Nf) 世紀(Na) 帶來(VC) 何

研究(Na) 人員(Na) 警告(VE) ，(COMMACATEGORY)

加州(Nc) 可能(D) 發生(VJ) 旱災(Na) 、(PAUSECATEGORY) 奧克拉荷馬州(Nc) 發生(VJ) 草原(Na) 大火(Na) ，(COMMACATEGORY)

東岸(Nc) 則(D) 可能(D) 遭遇(VJ) 海平面(Na) 上升(VA) ，(COMMACATEGORY)

尤其(D) 佛羅里達(Nc) ，(COMMACATEGORY)

而(Cbb) 這些(Neqa) 事件(Na) 多(D) 為(VG) 人類(Na) 造成(VK) 。(PERIODCATEGORY)

海平面(Na) 上升(VA) 也(D) 將(D) 吞噬(VC) 密西西比(Nb) 等(Cab) 低窪(VH) 地區(Nc) 。(PERIODCATEGORY)

至於(P) 超過(VJ) ８０００萬(Neu) 人(Na) 居住(VA) 且(Cbb) 擁有(VJ) 全美(Nb) 部分(Neqa) 成長(VH) 最(Dfa) 快(VH) 都會區(Nc) 的(DE) 東

「(PARENTHESISCATEGORY) 海平面(Na) 上升(VA) 加上(VC) 其他(Neqa) 與(Caa) 氣候(Na) 變遷(VH) 有關(VJ) 的(DE) 衝擊(Na) ，(COMMACATEGOR http://nlp.stanford.edu/software/index.shtml

The Stanford Natural Language Processing Group

home · people · teaching · research · publications · software · events · local

Stanford NLP Software

The Stanford NLP Group makes parts of our Natural Language Processing software available to everyone. These are statistical NLP toolkits for various major computational linguistics problems. They can be incorporated into applications with human language technology needs.

All the software we distribute here is written in Java. All recent distributions require Oracle Java 6+ or OpenJDK 7+. Distribution packages include components for command-line invocation, jar files, a Java API, and source code. A number of helpful people have extended our work with bindings or translations for other languages. As a result, much of this software can also easily be used from Python (or Jython), Ruby, Perl, Javascript, and F# or other .NET languages.

Supported software distributions

This code is being developed, and we try to answer questions and fix bugs on a best-effort basis.

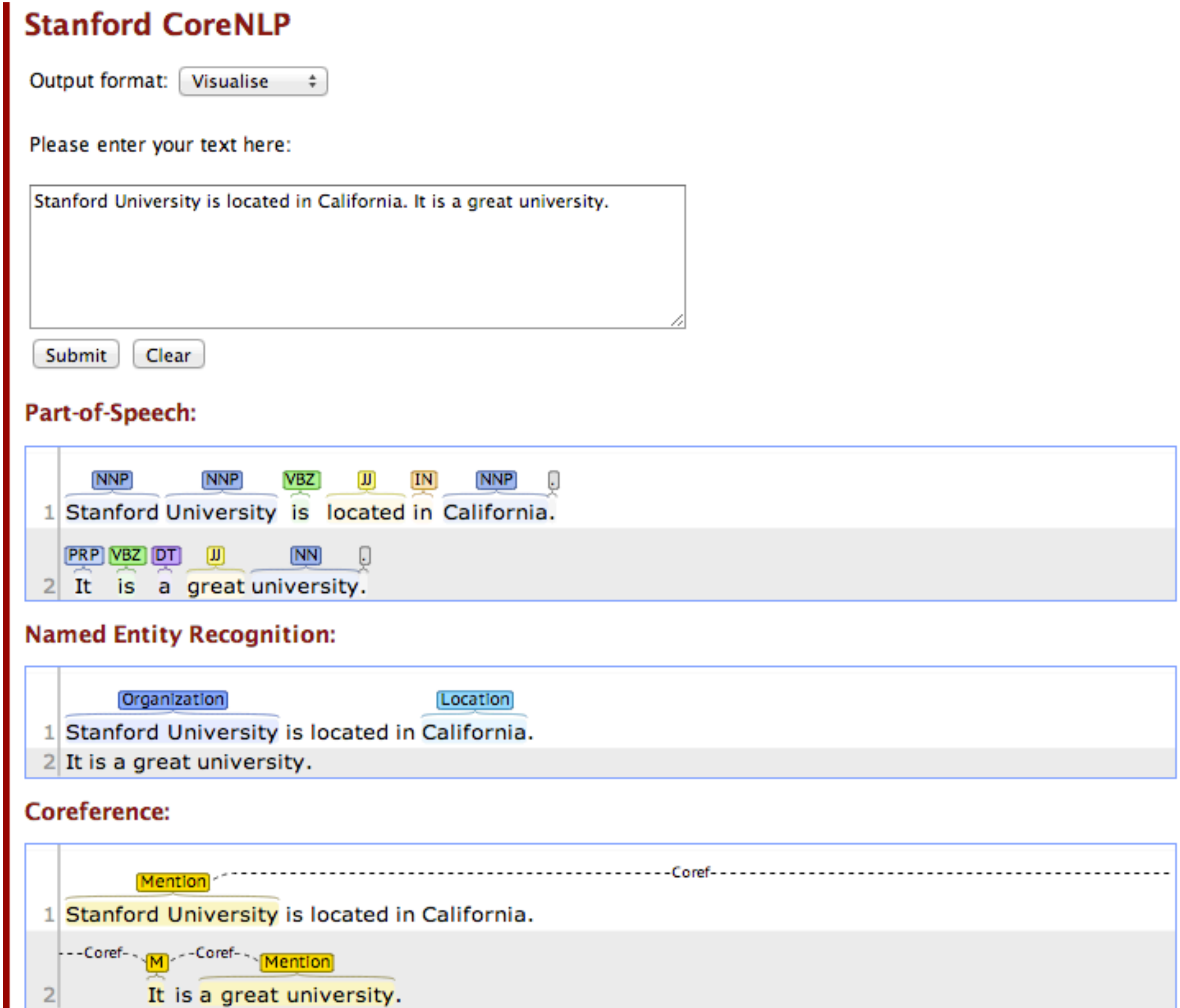
All these software distributions are open source, **licensed under the GNU General Public License** (v2 or later). Note that this is the *full* GPL, which allows many free uses, but *does not allow* its incorporation into any type of distributed proprietary software, even in part or in translation. **Commercial licensing** is also available; please contact us if you are interested.

Stanford CoreNLP
: An integrated suite of natural language processing tools for English and (mainland) Chinese in Java, including tokenization, part-of-speech tagging, named entity recognition, parsing, and coreference. See also: Stanford Deterministic Coreference Resolution, and the online CoreNLP demo, and the CoreNLP FAQ.

Stanford Parser
: Implementations of probabilistic natural language parsers in Java: highly optimized PCFG and dependency parsers, a lexicalized PCFG parser, and a deep learning reranker. See also: Online parser demo, the Stanford Dependencies page, and Parser FAQ.

Stanford POS Tagger
: A maximum-entropy (CMM) part-of-speech (POS) tagger for English,

Stanford CoreNLP http://nlp.stanford.edu:8080/corenlp/process

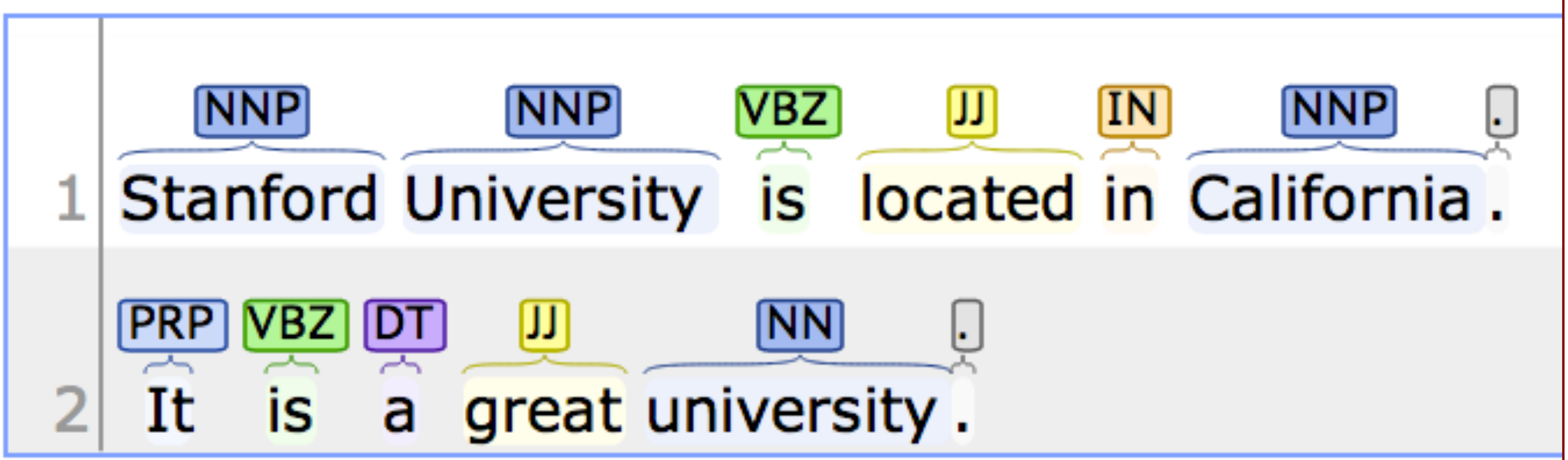


Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California.
It is a great university.

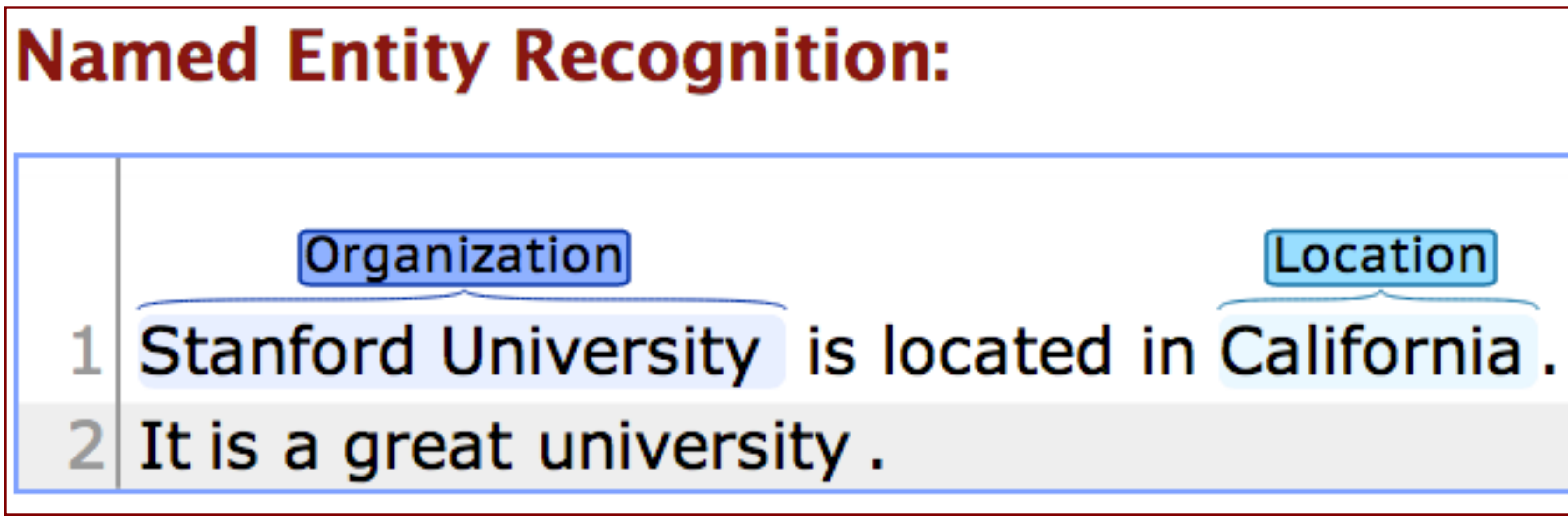
Part-of-Speech:



Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

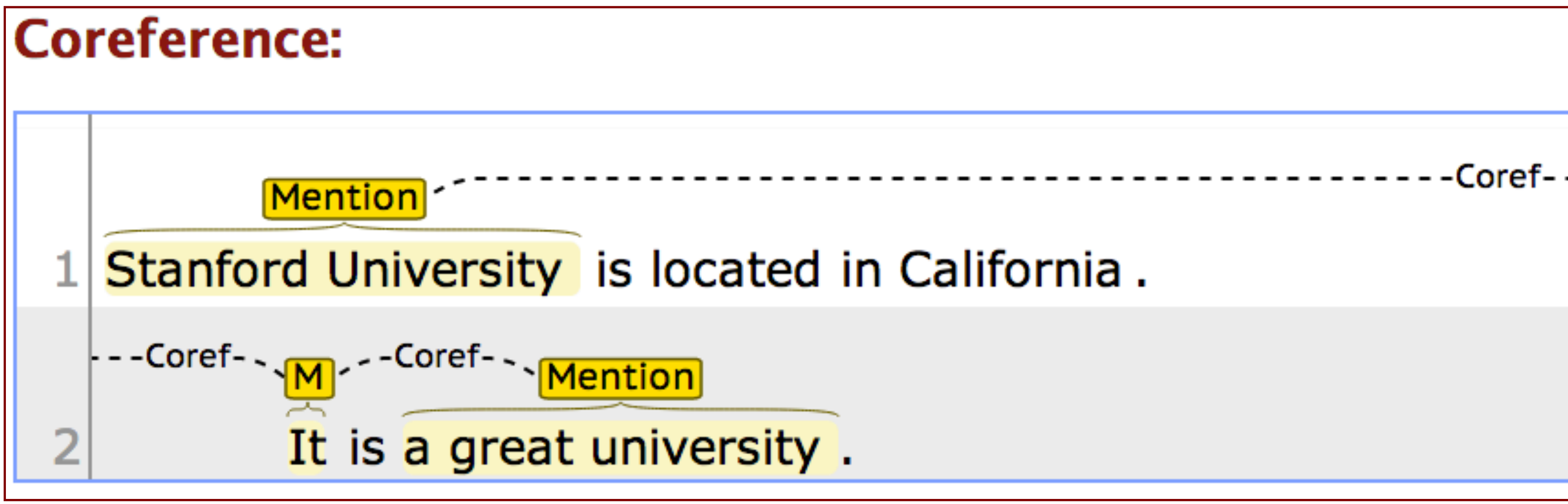
Stanford University is located in California.
It is a great university.



Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California.
It is a great university.

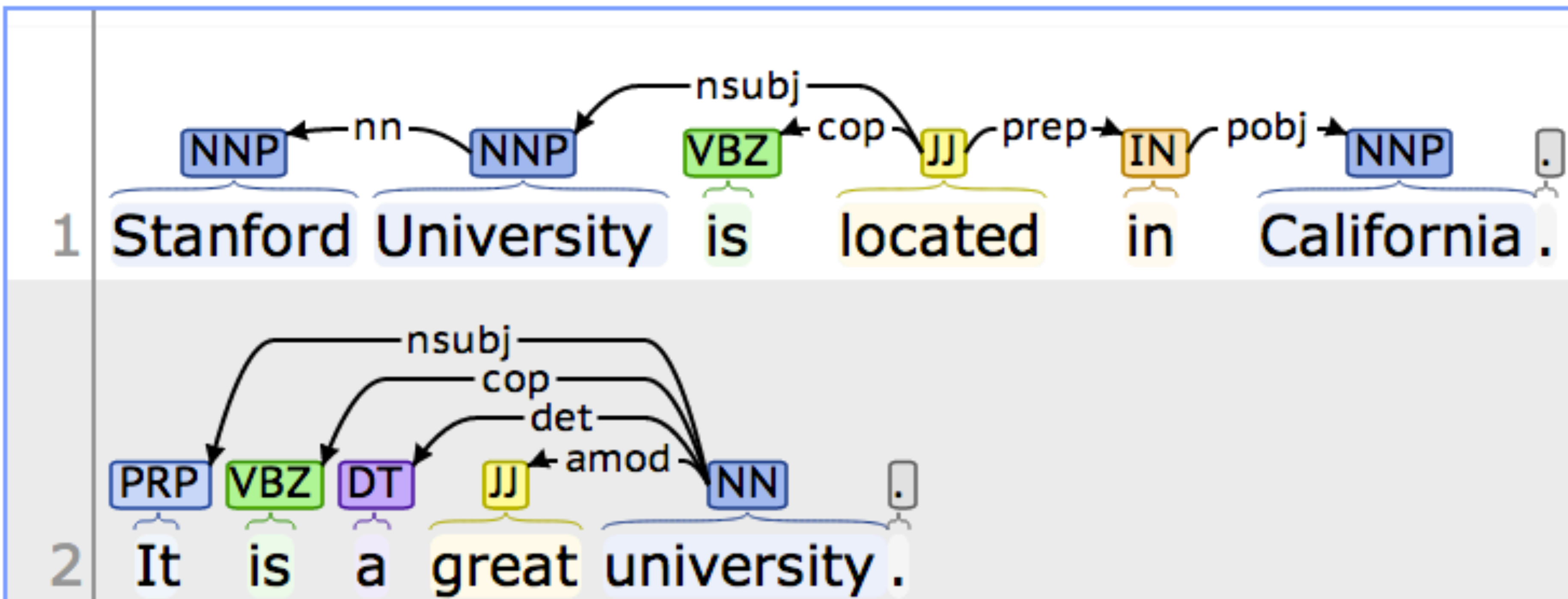


Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California.
It is a great university.

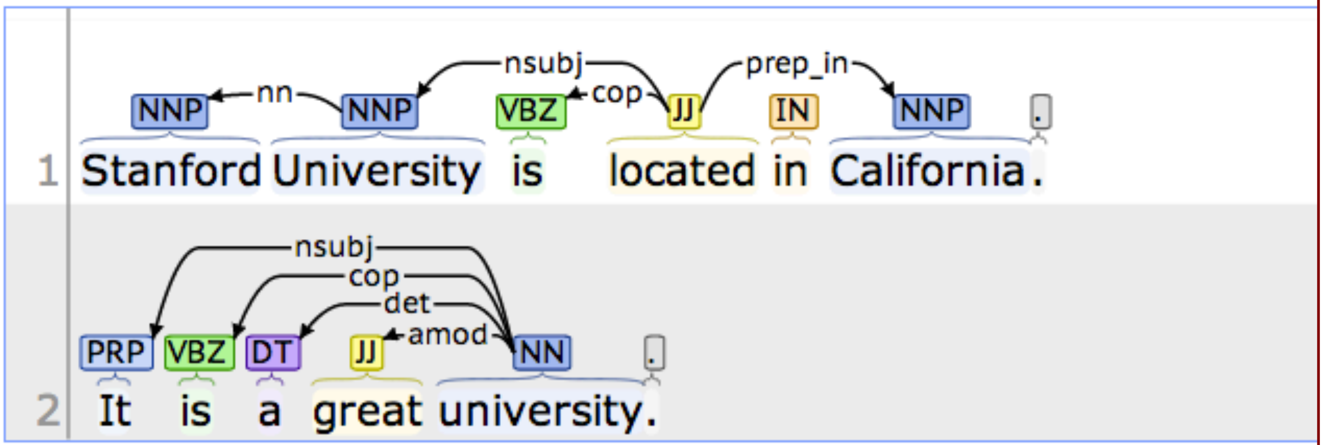
Basic dependencies:



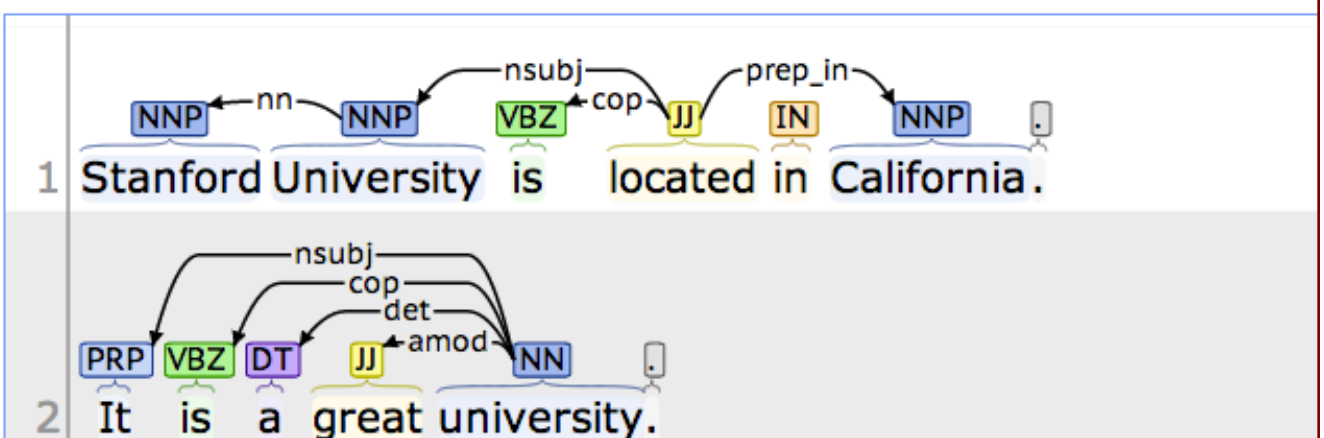
Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Collapsed dependencies:



Collapsed CC-processed dependencies:



Visualisation provided using the brat visualisation/annotation software.

Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Output format: Pretty print

Please enter your text here:

Stanford University is located in California. It is a great university.

Submit Clear

Stanford CoreNLP XML Output

Document

Document Info

Sentences

Sentence #1

Tokens

Id	Word	Lemma	Char begin	Char end	POS	NER	Normalized NER	Speaker
1	Stanford	Stanford	0	8	NNP	ORGANIZATION		PER0
2	University	University	9	19	NNP	ORGANIZATION		PER0
3	is	be	20	22	VBZ	O		PER0
4	located	located	23	30	JJ	O		PER0
5	in	in	31	33	IN	O		PER0
6	California	California	34	44	NNP	LOCATION		PER0
7	.	.	44	45	.	O		PER0

Parse tree

(ROOT (S (NP (NNP Stanford) (NNP University)) (VP (VBZ is) (ADJP (JJ located) (PP (IN in) (NP (NNP California))))) (. .)))

Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California.
It is a great university.

Sentence #1

Tokens

Id	Word	Lemma	Char begin	Char end	POS	NER	Normalized NER	Speaker
1	Stanford	Stanford	0	8	NNP	ORGANIZATION		PER0
2	University	University	9	19	NNP	ORGANIZATION		PER0
3	is	be	20	22	VBZ	O		PER0
4	located	located	23	30	JJ	O		PER0
5	in	in	31	33	IN	O		PER0
6	California	California	34	44	NNP	LOCATION		PER0
7	.	.	44	45	.	O		PER0

Parse tree

(ROOT (S (NP (NNP Stanford) (NNP University)) (VP (VBZ is) (ADJP (JJ located) (PP (IN in) (NP (NNP California))))) (. .)))

Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California. It is a great university.

Sentence #2

Tokens

Id	Word	Lemma	Char begin	Char end	POS	NER	Normalized NER	Speaker
1	It	it	46	48	PRP	O		PER0
2	is	be	49	51	VBZ	O		PER0
3	a	a	52	53	DT	O		PER0
4	great	great	54	59	JJ	O		PER0
5	university	university	60	70	NN	O		PER0
6	.	.	70	71	.	O		PER0

Parse tree

(ROOT (S (NP (PRP It)) (VP (VBZ is) (NP (DT a) (JJ great) (NN university))) (. .)))

Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California.
It is a great university.

Coreference resolution graph

1.

Sentence	Head	Text	Context
1	2 (gov)	Stanford University	
2	1	It	
2	5	a great university	

Tokens

Id	Word	Lemma	Char begin	Char end	POS	NER	Normalized NER		Speaker
1	Stanford	Stanford	0	8	NNP	ORGANIZATION			PER0
2	University	University	9	19	NNP	ORGANIZATION			PER0
3	is	be	20	22	VBZ	O		PER0	
4	located	located	23	30	JJ	O		PER0	
5	in	in	31	33	IN	O		PER0	
6	California	California	34	44	NNP	LOCATION		PER0	
7	.	.	44	45	.	O		PER0	

Parse tree
(ROOT (S (NP (NNP Stanford) (NNP University)) (VP (VBZ is) (ADJP (JJ located) (PP (IN in) (NP (NNP California))))) (. .)))

Uncollapsed dependencies

root (ROOT-0 , located-4)
nn (University-2 , Stanford-1)
nsubj (located-4 , University-2)
cop (located-4 , is-3)
prep (located-4 , in-5)
pobj (in-5 , California-6)

Collapsed dependencies

root (ROOT-0 , located-4)
nn (University-2 , Stanford-1)
nsubj (located-4 , University-2)
cop (located-4 , is-3)
prep_in (located-4 , California-6)

Collapsed dependencies with CC processed

root (ROOT-0 , located-4)
nn (University-2 , Stanford-1)
nsubj (located-4 , University-2)
cop (located-4 , is-3)
prep_in (located-4 , California-6)

Stanford CoreNLP

http://nlp.stanford.edu:8080/corenlp/process

Stanford University is located in California.
It is a great university.

Stanford CoreNLP http://nlp.stanford.edu:8080/corenlp/process

Output format: XML

Please enter your text here:

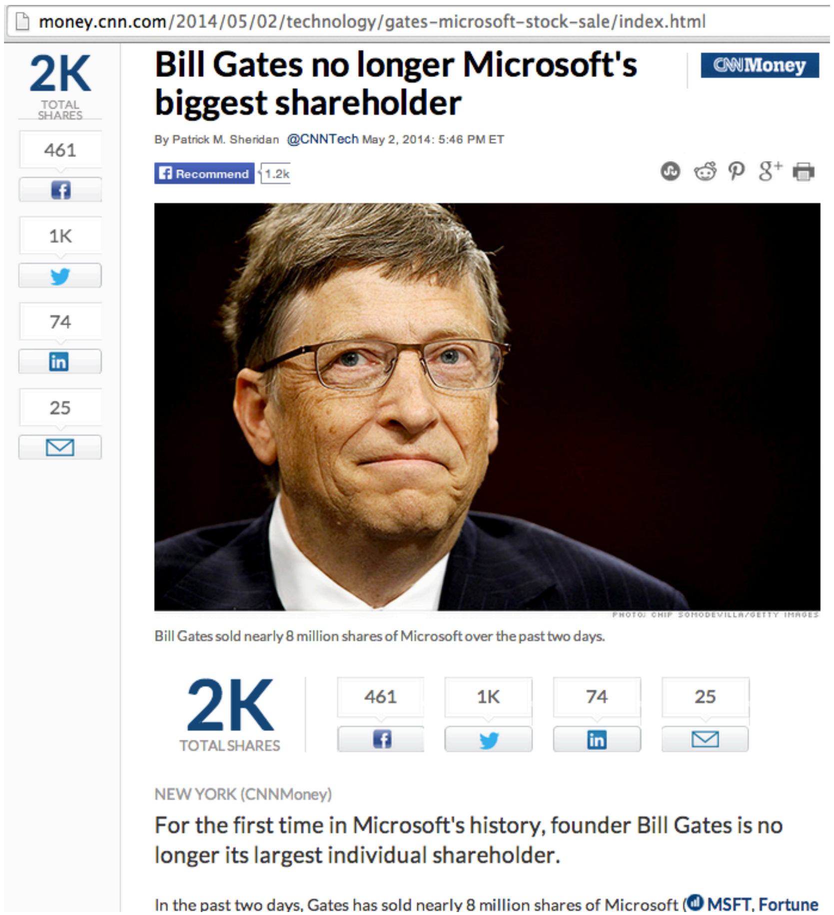
Stanford University is located in California. It is a great university.

Submit Clear


```xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="CoreNLP-to-HTML.xsl" type="text/xsl"?>
<root>
  <document>
    <sentences>
      <sentence id="1">
        <tokens>
          <token id="1">
            <word>Stanford</word>
            <lemma>Stanford</lemma>
            <CharacterOffsetBegin>0</CharacterOffsetBegin>
            <CharacterOffsetEnd>8</CharacterOffsetEnd>
            <POS>NNP</POS>
            <NER>ORGANIZATION</NER>
            <Speaker>PER0</Speaker>
          </token>
          <token id="2">
            <word>University</word>
            <lemma>University</lemma>
            <CharacterOffsetBegin>9</CharacterOffsetBegin>
            <CharacterOffsetEnd>19</CharacterOffsetEnd>
            <POS>NNP</POS>
            <NER>ORGANIZATION</NER>
            <Speaker>PER0</Speaker>
          </token>
```

NER for News Article

http://money.cnn.com/2014/05/02/technology/gates-microsoft-stock-sale/index.html


money.cnn.com/2014/05/02/technology/gates-microsoft-stock-sale/index.html

2K TOTAL SHARES

461 1K 74 25

Bill Gates no longer Microsoft's biggest shareholder

CNNMoney

By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Recommend 1.2k

PHOTO: CHIP SOMODEVILLA/GETTY IMAGES

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

2K TOTAL SHARES 461 1K 74 25

NEW YORK (CNNMoney)

For the first time in Microsoft's history, founder Bill Gates is no longer its largest individual shareholder.

In the past two days, Gates has sold nearly 8 million shares of Microsoft (MSFT, Fortune

Bill Gates no longer Microsoft's biggest shareholder
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

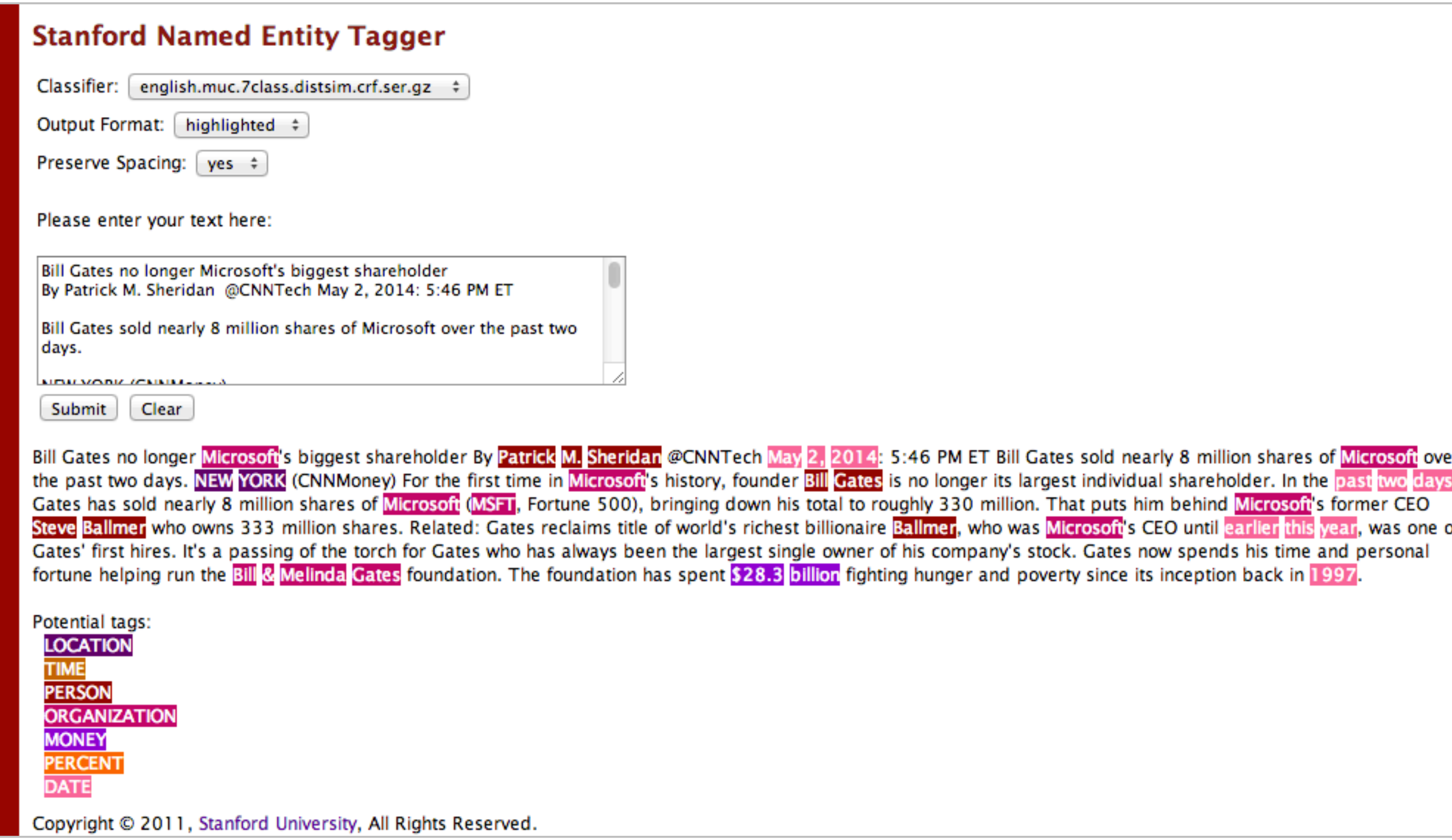
NEW YORK (CNNMoney)

For the first time in Microsoft's history, founder Bill Gates is no longer its largest individual shareholder.
In the past two days, Gates has sold nearly 8 million shares of Microsoft (MSFT, Fortune 500), bringing down his total to roughly 330 million.

That puts him behind Microsoft's former CEO Steve Ballmer who owns 333 million shares.
Related: Gates reclaims title of world's richest billionaire
Ballmer, who was Microsoft's CEO until earlier this year, was one of Gates' first hires.
It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the Bill & Melinda Gates foundation.
The foundation has spent $28.3 billion fighting hunger and poverty since its inception back in 1997.

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process



Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford Named Entity Tagger

Classifier: english.muc.7class.distsim.crf.ser.gz

Output Format: inlineXML

Preserve Spacing: yes

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

Submit Clear

Bill Gates no longer <ORGANIZATION>Microsoft</ORGANIZATION>'s biggest shareholder By <PERSON>Patrick M. Sheridan</PERSON> @CNNTech <DATE>May 2, 2014</DATE>: 5:46 PM ET Bill Gates sold nearly 8 million shares of <ORGANIZATION>Microsoft</ORGANIZATION> over the past two days. <LOCATION>NEW YORK</LOCATION> (CNNMoney) For the first time in <ORGANIZATION>Microsoft</ORGANIZATION>'s history, founder <PERSON>Bill Gates</PERSON> is no longer its largest individual shareholder. In the <DATE>past two days</DATE>, Gates has sold nearly 8 million shares of <ORGANIZATION>Microsoft</ORGANIZATION> (<ORGANIZATION>MSFT</ORGANIZATION>, Fortune 500), bringing down his total to roughly 330 million. That puts him behind <ORGANIZATION>Microsoft</ORGANIZATION>'s former CEO <PERSON>Steve Ballmer</PERSON> who owns 333 million shares. Related: Gates reclaims title of world's richest billionaire <PERSON>Ballmer</PERSON>, who was <ORGANIZATION>Microsoft</ORGANIZATION>'s CEO until <DATE>earlier this year</DATE>, was one of Gates' first hires. It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the <ORGANIZATION>Bill & Melinda Gates</ORGANIZATION> foundation. The foundation has spent <MONEY>$28.3 billion</MONEY> fighting hunger and poverty since its inception back in <DATE>1997</DATE>.

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford Named Entity Tagger

Classifier: english.muc.7class.distsim.crf.ser.gz

Output Format: xml

Preserve Spacing: yes

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

Submit Clear

<wi num="0" entity="O">Bill</wi> <wi num="1" entity="O">Gates</wi> <wi num="2" entity="O">no</wi> <wi num="3" entity="O">longer</wi> <wi num="4" entity="ORGANIZATION">Microsoft</wi><wi num="5" entity="O">'s</wi> <wi num="6" entity="O">biggest</wi> <wi num="7" entity="O">shareholder</wi> <wi num="8" entity="O">By</wi> <wi num="9" entity="PERSON">Patrick</wi> <wi num="10" entity="PERSON">M.</wi> <wi num="11" entity="PERSON">Sheridan</wi> <wi num="12" entity="O">@CNNTech</wi> <wi num="13" entity="DATE">May</wi> <wi num="14" entity="DATE">2</wi><wi num="15" entity="DATE">,</wi> <wi num="16" entity="DATE">2014</wi><wi num="17" entity="O">:</wi> <wi num="18" entity="O">5:46</wi> <wi num="19" entity="O">PM</wi> <wi num="20" entity="O">ET</wi> <wi num="21" entity="O">Bill</wi> <wi num="22" entity="O">Gates</wi> <wi num="23" entity="O">sold</wi> <wi num="24" entity="O">nearly</wi> <wi num="25" entity="O">8</wi> <wi num="26" entity="O">million</wi> <wi num="27" entity="O">shares</wi> <wi num="28" entity="O">of</wi> <wi num="29" entity="ORGANIZATION">Microsoft</wi> <wi num="30" entity="O">over</wi> <wi num="31" entity="O">the</wi> <wi num="32" entity="O">past</wi> <wi num="33" entity="O">two</wi> <wi num="34" entity="O">days</wi><wi num="35" entity="O">.</wi> <wi num="0" entity="LOCATION">NEW</wi> <wi num="1" entity="LOCATION">YORK</wi> <wi num="2" entity="O">-LRB-</wi><wi num="3" entity="O">CNNMoney</wi><wi num="4" entity="O">-RRB-</wi> <wi num="5" entity="O">For</wi> <wi num="6" entity="O">the</wi> <wi num="7" entity="O">first</wi> <wi num="8" entity="O">time</wi> <wi num="9" entity="O">in</wi> <wi num="10" entity="ORGANIZATION">Microsoft</wi><wi num="11" entity="O">'s</wi> <wi num="12" entity="O">history</wi><wi num="13" entity="O">,</wi> <wi num="14" entity="O">founder</wi> <wi num="15" entity="PERSON">Bill</wi> <wi num="16" entity="PERSON">Gates</wi> <wi num="17" entity="O">is</wi> <wi num="18" entity="O">no</wi> <wi num="19" entity="O">longer</wi> <wi num="20" entity="O">its</wi> <wi num="21" entity="O">largest</wi> <wi num="22" entity="O">individual</wi> <wi num="23" entity="O">shareholder</wi><wi num="24" entity="O">.</wi> <wi num="0" entity="O">In</wi> <wi num="1" entity="O">the</wi> <wi num="2" entity="DATE">past</wi> <wi num="3" entity="DATE">two</wi> <wi num="4"

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford Named Entity Tagger

Classifier: english.muc.7class.distsim.crf.ser.gz

Output Format: slashTags

Preserve Spacing: yes

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

Submit Clear

Bill/O Gates/O no/O longer/O Microsoft/ORGANIZATION's/O biggest/O shareholder/O By/O Patrick/PERSON M./PERSON Sheridan/PERSON @CNNTech/O May/DATE 2/DATE,/DATE 2014/DATE:/O 5:46/O PM/O ET/O Bill/O Gates/O sold/O nearly/O 8/O million/O shares/O of/O Microsoft/ORGANIZATION over/O the/O past/O two/O days/O./O NEW/LOCATION YORK/LOCATION -LRB-/OCNNMoney/O-RRB-/O For/O the/O first/O time/O in/O Microsoft/ORGANIZATION's/O history/O,/O founder/O Bill/PERSON Gates/PERSON is/O no/O longer/O its/O largest/O individual/O shareholder/O./O In/O the/O past/DATE two/DATE days/DATE,/O Gates/O has/O sold/O nearly/O 8/O million/O shares/O of/O Microsoft/ORGANIZATION -LRB-/OMSFT/ORGANIZATION,/O Fortune/O 500/O-RRB-/O,/O bringing/O down/O his/O total/O to/O roughly/O 330/O million/O./O That/O puts/O him/O behind/O Microsoft/ORGANIZATION's/O former/O CEO/O Steve/PERSON Ballmer/PERSON who/O owns/O 333/O million/O shares/O./O Related/O:/O Gates/O reclaims/O title/O of/O world/O's/O richest/O billionaire/O Ballmer/PERSON,/O who/O was/O Microsoft/ORGANIZATION's/O CEO/O until/O earlier/DATE this/DATE year/DATE,/O was/O one/O of/O Gates/O'/O first/O hires/O./O It/O's/O a/O passing/O of/O the/O torch/O for/O Gates/O who/O has/O always/O been/O the/O largest/O single/O owner/O of/O his/O company/O's/O stock/O./O Gates/O now/O spends/O his/O time/O and/O personal/O fortune/O helping/O run/O the/O Bill/ORGANIZATION &/ORGANIZATION Melinda/ORGANIZATION Gates/ORGANIZATION foundation/O./O The/O foundation/O has/O spent/O $/MONEY28.3/MONEY billion/MONEY fighting/O hunger/O and/O poverty/O since/O its/O inception/O back/O in/O 1997/DATE./O

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford Named Entity Tagger

Classifier: english.conll.4class.distsim.crf.ser.gz

Output Format: highlighted

Preserve Spacing: yes

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

Submit Clear

Bill Gates no longer Microsoft's biggest shareholder By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET Bill Gates sold nearly 8 million shares of Microsoft over the past two days. NEW YORK (CNNMoney) For the first time in Microsoft's history, founder Bill Gates is no longer its largest individual shareholder. In the past two days, Gates has sold nearly 8 million shares of Microsoft (MSFT, Fortune 500), bringing down his total to roughly 330 million. That puts him behind Microsoft's former CEO Steve Ballmer who owns 333 million shares. Related: Gates reclaims title of world's richest billionaire Ballmer, who was Microsoft's CEO until earlier this year, was one of Gates' first hires. It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the Bill & Melinda Gates foundation. The foundation has spent $28.3 billion fighting hunger and poverty since its inception back in 1997.

Potential tags:

- LOCATION
- ORGANIZATION
- PERSON
- MISC

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford Named Entity Tagger

Classifier: english.all.3class.distsim.crf.ser.gz

Output Format: highlighted

Preserve Spacing: yes

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET

Bill Gates sold nearly 8 million shares of Microsoft over the past two days.

Submit Clear

Bill Gates no longer Microsoft's biggest shareholder By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET Bill Gates sold nearly 8 million shares of Microsoft over the past two days. NEW YORK (CNNMoney) For the first time in Microsoft's history, founder Bill Gates is no longer its largest individual shareholder. In the past two days, Gates has sold nearly 8 million shares of Microsoft (MSFT, Fortune 500), bringing down his total to roughly 330 million. That puts him behind Microsoft's former CEO Steve Ballmer who owns 333 million shares. Related: Gates reclaims title of world's richest billionaire Ballmer, who was Microsoft's CEO until earlier this year, was one of Gates' first hires. It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the Bill & Melinda Gates foundation. The foundation has spent $28.3 billion fighting hunger and poverty since its inception back in 1997.

Potential tags:
- LOCATION
- ORGANIZATION
- PERSON

Classifier: english.muc.**7class**.distsim.crf.ser.gz

Bill Gates no longer Microsoft's biggest shareholder By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET Bill Gates sold nearly 8 million shares of Microsoft over the past two days. NEW YORK (CNNMoney) For the first time in Microsoft's history, founder Bill Gates is no longer its largest individual shareholder. In the past two days, Gates has sold nearly 8 million shares of Microsoft (MSFT, Fortune 500), bringing down his total to roughly 330 million. That puts him behind Microsoft's former CEO Steve Ballmer who owns 333 million shares. Related: Gates reclaims title of world's richest billionaire Ballmer, who was Microsoft's CEO until earlier this year, was one of Gates' first hires. It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the Bill & Melinda Gates foundation. The foundation has spent $28.3 billion fighting hunger and poverty since its inception back in 1997.

Potential tags:
- LOCATION
- TIME
- PERSON
- ORGANIZATION
- MONEY
- PERCENT
- DATE

Classifier: english.all.**3class**.distsim.crf.ser.gz

Bill Gates no longer Microsoft's biggest shareholder By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET Bill Gates sold nearly 8 million shares of Microsoft over the past two days. NEW YORK (CNNMoney) For the first time in Microsoft's history, founder Bill Gates is no longer its largest individual shareholder. In the past two days, Gates has sold nearly 8 million shares of Microsoft (MSFT, Fortune 500), bringing down his total to roughly 330 million. That puts him behind Microsoft's former CEO Steve Ballmer who owns 333 million shares. Related: Gates reclaims title of world's richest billionaire Ballmer, who was Microsoft's CEO until earlier this year, was one of Gates' first hires. It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the Bill & Melinda Gates foundation. The foundation has spent $28.3 billion fighting hunger and poverty since its inception back in 1997.

Potential tags:
- LOCATION
- ORGANIZATION
- PERSON

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford NER Output Format: inlineXML

Bill Gates no longer <ORGANIZATION>Microsoft</ORGANIZATION>'s biggest shareholder By <PERSON>Patrick M. Sheridan</PERSON> @CNNTech <DATE>May 2, 2014</DATE>: 5:46 PM ET Bill Gates sold nearly 8 million shares of <ORGANIZATION>Microsoft</ORGANIZATION> over the past two days. <LOCATION>NEW YORK</LOCATION> (CNNMoney) For the first time in <ORGANIZATION>Microsoft</ORGANIZATION>'s history, founder <PERSON>Bill Gates</PERSON> is no longer its largest individual shareholder. In the <DATE>past two days</DATE>, Gates has sold nearly 8 million shares of <ORGANIZATION>Microsoft</ORGANIZATION> (<ORGANIZATION>MSFT</ORGANIZATION>, Fortune 500), bringing down his total to roughly 330 million. That puts him behind <ORGANIZATION>Microsoft</ORGANIZATION>'s former CEO <PERSON>Steve Ballmer</PERSON> who owns 333 million shares. Related: Gates reclaims title of world's richest billionaire <PERSON>Ballmer</PERSON>, who was <ORGANIZATION>Microsoft</ORGANIZATION>'s CEO until <DATE>earlier this year</DATE>, was one of Gates' first hires. It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the <ORGANIZATION>Bill & Melinda Gates</ORGANIZATION> foundation. The foundation has spent <MONEY>$28.3 billion</MONEY> fighting hunger and poverty since its inception back in <DATE>1997</DATE>.

Stanford Named Entity Tagger (NER)

http://nlp.stanford.edu:8080/ner/process

Stanford NER Output Format: slashTags

Bill/O Gates/O no/O longer/O Microsoft/ORGANIZATION's/O biggest/O shareholder/O By/O Patrick/PERSON M./PERSON Sheridan/PERSON @CNNTech/O May/DATE 2/DATE,/DATE 2014/DATE:/O 5:46/O PM/O ET/O Bill/O Gates/O sold/O nearly/O 8/O million/O shares/O of/O Microsoft/ORGANIZATION over/O the/O past/O two/O days/O./O NEW/LOCATION YORK/LOCATION -LRB-/OCNNMoney/O-RRB-/O For/O the/O first/O time/O in/O Microsoft/ORGANIZATION's/O history/O,/O founder/O Bill/PERSON Gates/PERSON is/O no/O longer/O its/O largest/O individual/O shareholder/O./O In/O the/O past/DATE two/DATE days/DATE,/O Gates/O has/O sold/O nearly/O 8/O million/O shares/O of/O Microsoft/ORGANIZATION -LRB-/OMSFT/ORGANIZATION,/O Fortune/O 500/O-RRB-/O,/O bringing/O down/O his/O total/O to/O roughly/O 330/O million/O./O That/O puts/O him/O behind/O Microsoft/ORGANIZATION's/O former/O CEO/O Steve/PERSON Ballmer/PERSON who/O owns/O 333/O million/O shares/O./O Related/O:/O Gates/O reclaims/O title/O of/O world/O's/O richest/O billionaire/O Ballmer/PERSON,/O who/O was/O Microsoft/ORGANIZATION's/O CEO/O until/O earlier/DATE this/DATE year/DATE,/O was/O one/O of/O Gates/O'/O first/O hires/O./O It/O's/O a/O passing/O of/O the/O torch/O for/O Gates/O who/O has/O always/O been/O the/O largest/O single/O owner/O of/O his/O company/O's/O stock/O./O Gates/O now/O spends/O his/O time/O and/O personal/O fortune/O helping/O run/O the/O Bill/ORGANIZATION &/ORGANIZATION Melinda/ORGANIZATION Gates/ORGANIZATION foundation/O./O The/O foundation/O has/O spent/O $/MONEY28.3/MONEY billion/MONEY fighting/O hunger/O and/O poverty/O since/O its/O inception/O back/O in/O 1997/DATE./O

Summary

- Sentiment Analysis
- Architectures of Sentiment Analytics
- Opinion Spam Detection
- Text Mining Techniques and Natural Language Processing

References

- Cambria, Erik, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. "SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives." In the 26th International Conference on Computational Linguistics (COLING), Osaka. 2016.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts (2013), "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," In Proceedings of the conference on empirical methods in natural language processing (EMNLP), vol. 1631, p. 1642 http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf
- Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.
- Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, vol 139, no. 11, 2016. pp.5-15.
- Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.
- Steven Struhl (2015), Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Kogan Page
- Bing Liu (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press

References

- Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” 2nd Edition, Springer. http://www.cs.uic.edu/~liub/WebMiningBook.html
- Bing Liu (2013), Opinion Spam Detection: Detecting Fake Reviews and Reviewers, http://www.cs.uic.edu/~liub/FBS/fake-reviews.html
- Bo Pang and Lillian Lee (2008), "Opinion mining and sentiment analysis,” Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, 2008.
- Wiltrud Kessler (2012), Introduction to Sentiment Analysis, http://www.ims.uni-stuttgart.de/~kesslewd/lehre/sentimentanalysis12s/introduction_sentimentanalysis.pdf
- Z. Zhang, X. Li, and Y. Chen (2012), "Deciphering word-of-mouth in social media: Text-based metrics of consumer reviews," ACM Trans. Manage. Inf. Syst. (3:1) 2012, pp 1-23.
- Efraim Turban, Ramesh Sharda, Dursun Delen (2011), Decision Support and Business Intelligence Systems, Ninth Edition, 2011, Pearson.
- Guandong Xu, Yanchun Zhang, Lin Li (2011), Web Mining and Social Networking: Techniques and Applications, 2011, Springer

References

- Efraim Turban, Ramesh Sharda, Dursun Delen, Decision Support and Business Intelligence Systems, Ninth Edition, 2011, Pearson.
- Steven Bird, Ewan Klein and Edward Loper, Natural Language Processing with Python, 2009, O'Reilly Media, http://www.nltk.org/book/ , http://www.nltk.org/book_1ed/
- Nitin Hardeniya, NLTK Essentials, 2015, Packt Publishing
- Michael W. Berry and Jacob Kogan, Text Mining: Applications and Theory, 2010, Wiley
- Guandong Xu, Yanchun Zhang, Lin Li, Web Mining and Social Networking: Techniques and Applications, 2011, Springer
- Matthew A. Russell, Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites, 2011, O'Reilly Media
- Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2009, Springer
- Bruce Croft, Donald Metzler, and Trevor Strohman, Search Engines: Information Retrieval in Practice, 2008, Addison Wesley, http://www.search-engines-book.com/
- Christopher D. Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, 1999, The MIT Press
- Text Mining, http://en.wikipedia.org/wiki/Text_mining