

Natural Language Processing  
SoSe 2015



## Sentiment Analysis

*Dr. Mariana Neves*

*June 8th, 2015*

(based on the slides of Dr. Saeedeh Momtazi)

# Outline

- Applications
- Task
- Machine Learning Approach
- Rule-based Approach

# Outline

- Applications
- Task
- Machine Learning Approach
- Rule-based Approach

# Product reviews

## Customer Reviews Speech and Language Processing, 2nd Edition



### The most helpful favorable review

4 of 4 people found the following review helpful

**★★★★★ Great introductions and reference book**  
 I read the first edition of that book and it is terrific. The second edition is much more adapted to current research. Statistical methods in NLP are more detailed and some syntax-based approaches are presented. My specific interest is in machine translation and dialogue systems. Both chapters are extensively rewritten and much more elaborated. I believe this book is...

[Read the full review >](#)

Published on August 9, 2008 by carheg

> See more [5 star](#), [4 star](#) reviews

vs.

### The most helpful critical review

37 of 37 people found the following review helpful

**★★★☆☆ Good description of the problems in the field, but look elsewhere for practical solutions**  
 The authors have the challenge of covering a vast area, and they do a good job of highlighting the hard problems within individual sub-fields, such as machine translation. The availability of an accompanying Web site is a strong plus, as is the extensive bibliography, which also includes links to freely available software and resources.

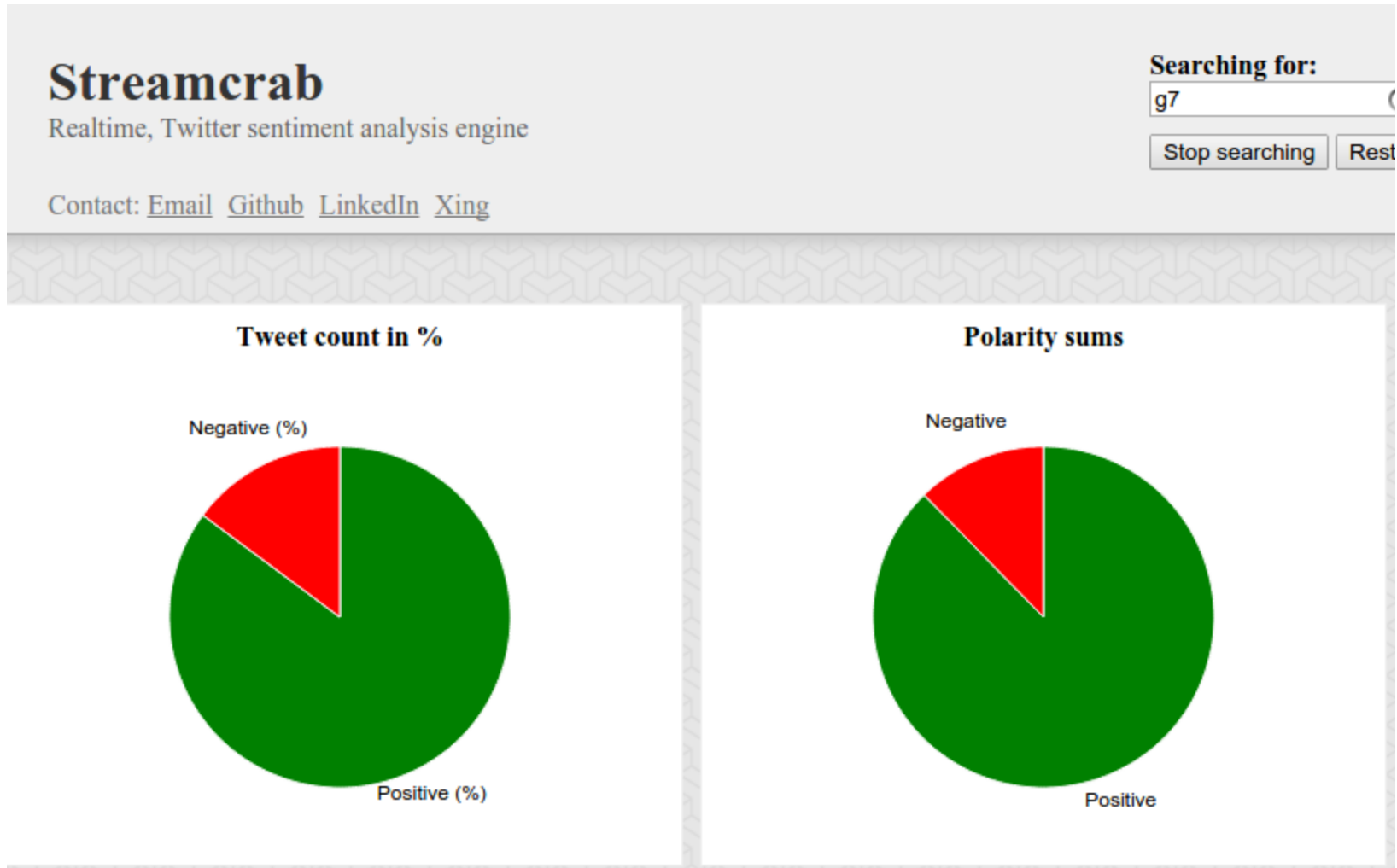
Now for the...

[Read the full review >](#)

Published on April 2, 2009 by P. Nadkarni

> See more [3 star](#), [2 star](#), [1 star](#) reviews

# Social Media



(<http://www.streamcrab.com/>)

# Event Analysis and Prediction

## Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis

Haruna Isah, Paul Trundle, Daniel Neagu  
Artificial Intelligence Research (AIRe) Group  
School of Electrical Engineering and Computer Science  
University of Bradford  
Bradford, UK

H.Isah@student.bradford.ac.uk, P.R.Trundle@bradford.ac.uk, D.Neagu@bradford.ac.uk

# Event Analysis and Prediction

## Predicting Elections from Social Networks Based on Sub-event Detection and Sentiment Analysis

Sayan Unankard<sup>1</sup>, Xue Li<sup>1</sup>, Mohamed Sharaf<sup>1</sup>, Jiang Zhong<sup>2</sup>, and Xueming Li<sup>2</sup>

<sup>1</sup> School of Information Technology and Electrical Engineering,  
The University of Queensland, Brisbane QLD 4072, Australia




<sup>2</sup> Key Laboratory of Dependable Service Computing in Cyber Physical Society  
Ministry of Education, Chongqing 400044, China  
{uqsunank,m.sharaf}@uq.edu.au, xueli@itee.uq.edu.au,  
{zhongjiang,lixuemin}@cqu.edu.cn

**Abstract.** Social networks are widely used by all kinds of people to express their opinions. Predicting election outcomes is now becoming a compelling research issue. People express themselves spontaneously with respect to the social events in their social networks. Real time prediction on ongoing election events can provide feedback and trend analysis for politicians and news analysts to make informed decisions. This paper proposes an approach to predicting election results by incorporating sub-event detection and sentiment analysis in social networks to analyse as well as visualise political preferences revealed by those social network users. Extensive experiments are conducted to evaluate the performance of our approach based on a real-world *Twitter* dataset. Our experiments show that the proposed approach can effectively predict the election results over the given baselines.

# Event Analysis and Prediction

## Impact Feed

Positive Negative

<p>+0.94</p> <p>-  +</p> <p>[ Neg:0   Pos:50 ]</p>	<p><b>Prologis Inc</b>            Prologis Raised to "Buy" at Vetr Inc.            4/17/2015   mideasttime.com</p>
<p>+0.90</p> <p>-  +</p> <p>[ Neg:0   Pos:30 ]</p>	<p><b>Cabot Oil &amp; Gas Corporation</b>            Cabot Oil &amp; Gas Co. Upgraded to "Overweight" at Simmons            4/17/2015   lulegacy.com</p>
<p>+0.90</p> <p>-  +</p> <p>[ Neg:0   Pos:30 ]</p>	<p><b>Actavis Inc</b>            Barclays Raises Actavis plc Price Target to \$300.00            4/17/2015   wkrb13.com</p>
<p>+0.90</p> <p>-  +</p> <p>[ Neg:0   Pos:30 ]</p>	<p><b>Agilent Technologies Inc</b>            Agilent Technologies Raised to Buy at Vetr Inc.            4/17/2015   lulegacy.com</p>
<p>+0.90</p> <p>-  +</p> <p>[ Neg:0   Pos:28 ]</p>	<p><b>Ameriprise Financial Inc</b>            Traders Sell Shares of Ameriprise Financial on Strength Following Insider ...            4/17/2015   lulegacy.com</p>
<p>+0.87</p> <p>-  +</p> <p>[ Neg:0   Pos:21 ]</p>	<p><b>Texas Instruments Incorporated</b>            Texas Instruments Incorporated Declares \$0.34 Quarterly Dividend            4/17/2015   wkrb13.com</p>

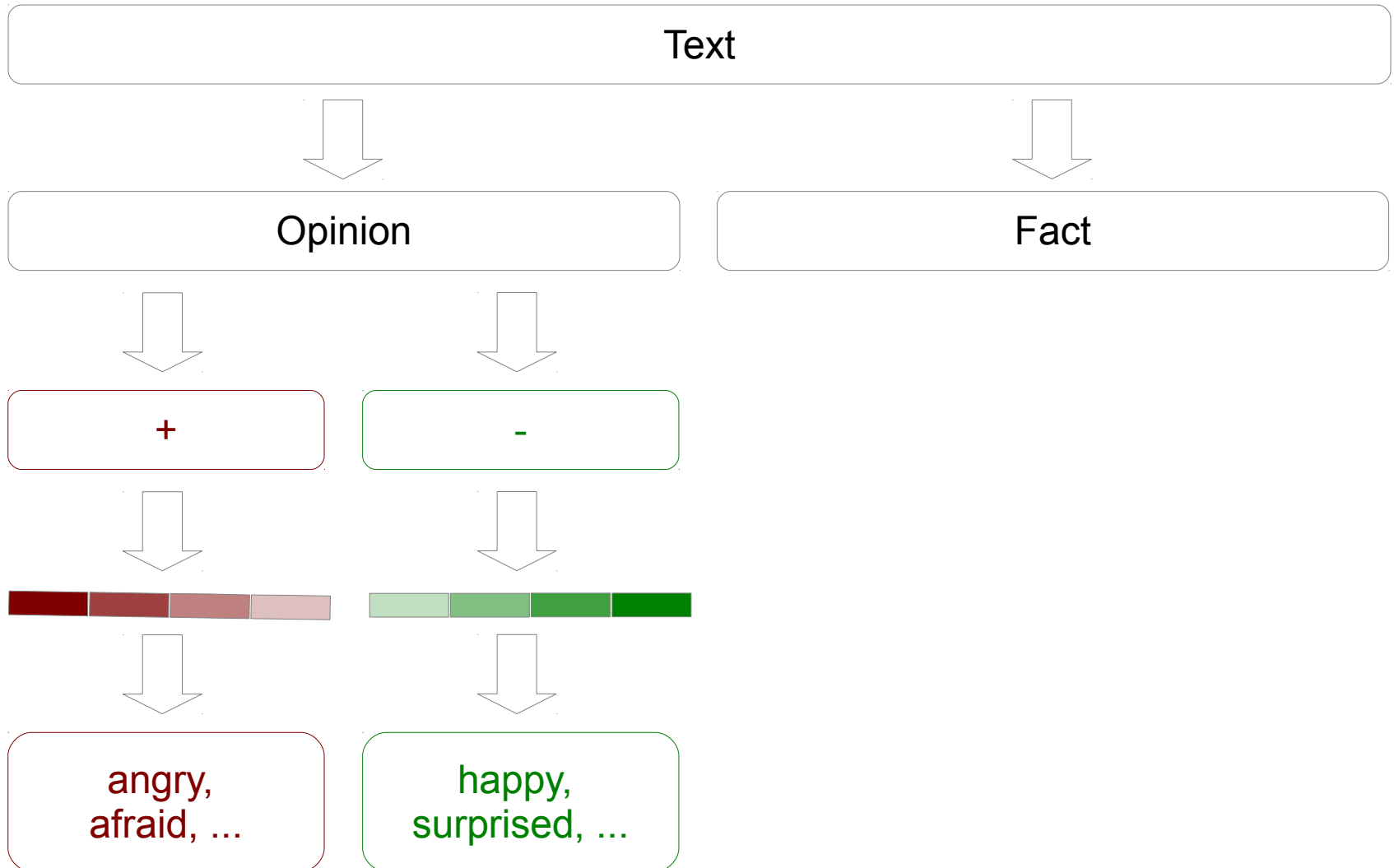
(<http://www.thestocksonar.com/Sentiment-Analysis>)



# Outline

- Applications
- **Task**
- Machine Learning Approach
- Rule-based Approach

# Sentiment Analysis Levels



# Advanced Sentiment Analysis

- Opinion holder and Opinion target/aspect
  - Students [OP HOLDER] like Wikipedia [TARGET] because it is easy to use and it sounds authoritative.
  - I had a nice stay in this hotel and the rooms [ASPECT] were very clean.

# Advanced Sentiment Analysis

- Mixed opinions
  - The restaurant has an amazing view but it is very dirty.

## Other names

- Opinion mining
- Opinion extraction
- Sentiment mining
- Subjectivity detection
- Subjectivity analysis

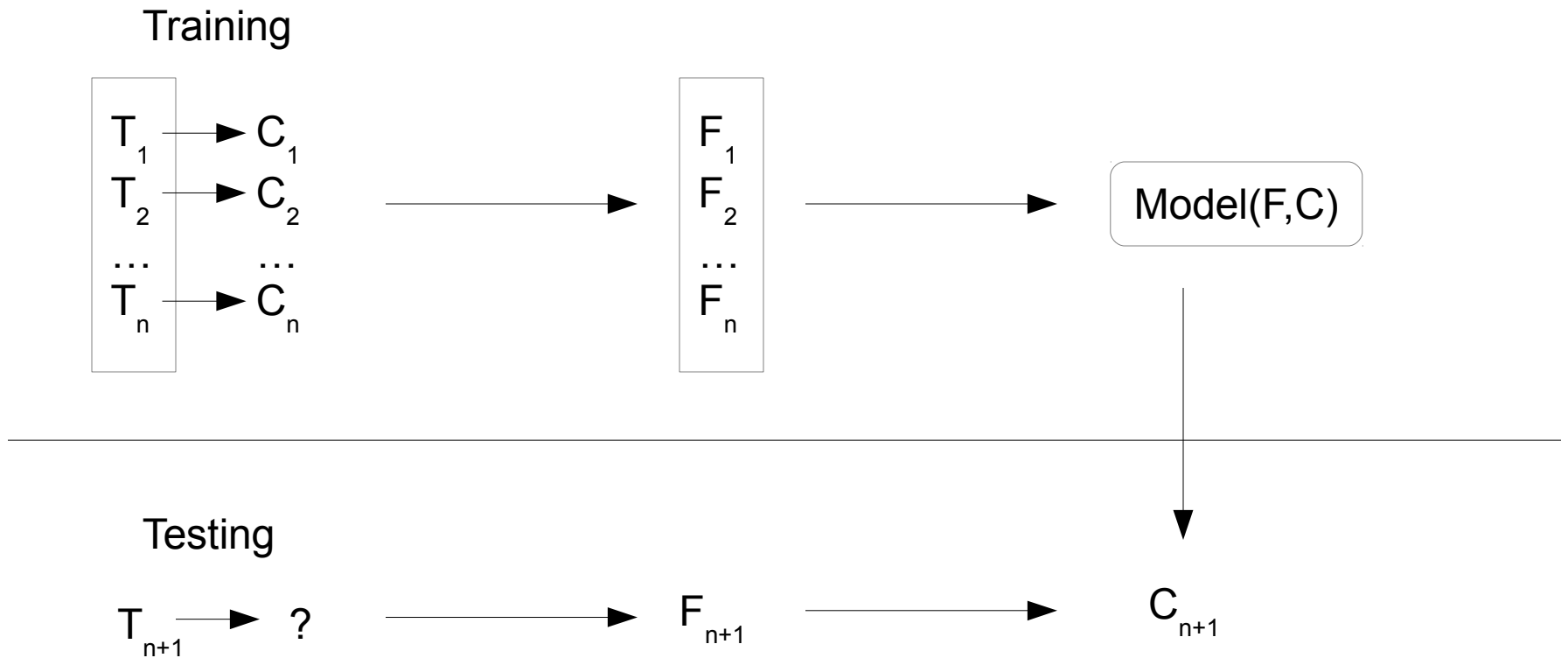
# Sentiment Analysis Approaches

- Machine learning methods
  - ⇒ classification
- Rule-based methods
  - ⇒ dictionary oriented

# Outline

- Applications
- Task
- Machine Learning Approach
- Rule-based Approach

# Machine Learning Approach





# Sentiment Classification

- Using any kinds of supervised classifiers
  - K Nearest Neighbor
  - Support Vector Machines
  - Naïve Bayes
  - Maximum Entropy
  - Logistic Regression
  - ...

# Features

- All words or adjectives?
  - All words works better than adjectives only

# Features

- Word occurrence or frequency?
  - Word occurrence is more useful than frequency
  - Using binary value for words
  - Replace all word counts higher than 0 in each text by 1

# Features

- Negation
  - Negation words change the text polarity
    - Adding prefix NOT– to every word between negation and next punctuation
- „I did **not** like the restaurant location, but the food ...“
- I did not **NOT**-like **NOT**-the **NOT**-restaurant **NOT**-location, but the food ...

# Features

- Other emotions
  - Considering emoticons as additional features
    - :)
    - :(
  - As well as smilies
    - 😊
    - 😐
    - ☹️

# Fine-grained analysis

- Dealing with finer classes of sentiment

-3,-2,-1,+1,+2,+3

TA_TOKEN	TA_TYPE
well established	WeakPositiveSentiment
most extensively used drug	StrongPositiveSentiment
capable	WeakPositiveSentiment
most severe	MajorProblem
not develop mature	WeakNegativeSentiment
important	WeakPositiveSentiment
lack	MinorProblem
deadliest	StrongNegativeSentiment
prompted	WeakPositiveSentiment
effort	WeakNegativeSentiment
most severe	MajorProblem
Lack	MinorProblem
most devastating	StrongNegativeSentiment
severe	MajorProblem
malignant	MinorProblem
most severe	MajorProblem
important	WeakPositiveSentiment
neglected	WeakNegativeSentiment
complete a complex	MajorProblem
important	WeakPositiveSentiment
inhibits	MinorProblem

(SAP HANA database)

# Fine-grained Analysis

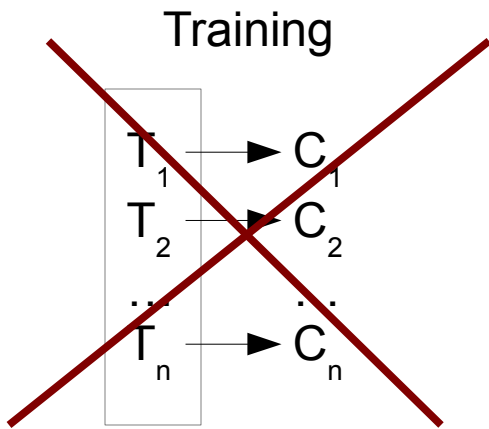
- Approaches
  - Using multiclass classifier (6 classes in this case)
  - Using two level classifier
    - First level: polarity classifier (positive or negative)
    - Second level: strength classifier (1 or 2 or 3)

# Outline

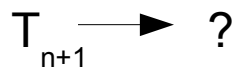
- Applications
- Task
- Machine Learning Approach
- Rule-based Approach



# Rule-based Approach



Testing



$C_{n+1}$





## Rule-based Approach

- Looking for opinionated words in each text
- Classifying the text based on the number of positive and negative words



## Rule-based Approach

- Considering different rules for classification
  - Fine-grained dictionary
  - Negation words
  - Booster words
  - Idioms
  - Emoticons
  - Mixed opinions
  - Linguistic features of the language




# Rule-based Approach

- Fine-grained Dictionary
  - „It was a **good** song.“ 
  - „The song was **excellent**.“ 


# Rule-based Approach

- Negation Words
  - „It was a good song.“ 
  - „The song was **not** good.“ 

# Rule-based Approach

- Booster Words
  - „The song was interesting.“ 
  - „The song was **very** interesting.“ 
  - „The song was **somewhat** interesting.“ 

# Rule-based Approach

- Idioms
  - „shock horror“ 

## Rule-based Approach

- Mixed Opinions

„The song was **good**, but I think its title was **strange**.“   



# Opinion Dictionary

- English
  - Subjectivity Clues (2005)
  - SentiSpin (2005)
  - SentiWordNet (2006)
  - Polarity Enhancement (2009)
  - SentiStrength (2010)

# Opinion Dictionary

- German
  - GermanPolarityClues (2010)
  - SentiWortSchatz (2010)
  - GermanSentiStrength (2012)

# Machine Learning with Opinion Dictionary

- Using opinion words as a feature in the algorithms
- Ignoring other words in the text