

# Web Mining (網路探勘)

## Opinion Mining and Sentiment Analysis (意見探勘與情感分析)

1011WM11

TLMXM1A

Wed 8,9 (15:10-17:00) U705

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/>

2012-12-12

# Outline

- Introduction to Opinion Mining and Sentiment Analysis
- Social Media Monitoring/Analysis
- Resources of Opinion Mining

# 課程大綱 (Syllabus)

週次	日期	內容 (Subject/Topics)
1	101/09/12	Introduction to Web Mining (網路探勘導論)
2	101/09/19	Association Rules and Sequential Patterns (關聯規則和序列模式)
3	101/09/26	Supervised Learning (監督式學習)
4	101/10/03	Unsupervised Learning (非監督式學習)
5	101/10/10	國慶紀念日(放假一天)
6	101/10/17	Paper Reading and Discussion (論文研讀與討論)
7	101/10/24	Partially Supervised Learning (部分監督式學習)
8	101/10/31	Information Retrieval and Web Search (資訊檢索與網路搜尋)
9	101/11/07	Social Network Analysis (社會網路分析)

# 課程大綱 (Syllabus)

週次    日期    內容 (Subject/Topics)

10    101/11/14    Midterm Presentation (期中報告)

11    101/11/21    Web Crawling (網路爬行)

12    101/11/28    Structured Data Extraction (結構化資料擷取)

13    101/12/05    Information Integration (資訊整合)

14    101/12/12    Opinion Mining and Sentiment Analysis  
(意見探勘與情感分析)

15    101/12/19    Paper Reading and Discussion (論文研讀與討論)

16    101/12/26    Web Usage Mining (網路使用挖掘)

17    102/01/02    Project Presentation 1 (期末報告1)

18    102/01/09    Project Presentation 2 (期末報告2)

# Social Media and the Voice of the Customer

- Listen to the **Voice of the Customer (VoC)**
  - Social media can give companies a torrent of highly valuable customer feedback.
  - Such input is largely free
  - Customer feedback issued through social media is qualitative data, just like the data that market researchers derive from focus group and in-depth interviews
  - Such qualitative data is in digital form – in text or digital video on a web site.

# Listen and Learn

## Text Mining for VoC

- Categorization
  - Understanding what topics people are talking or writing about in the unstructured portion of their feedback.
- Sentiment Analysis
  - Determining whether people have **positive**, **negative**, or **neutral** views on those topics.

# Opinion Mining and Sentiment Analysis

- Mining opinions which indicate **positive** or **negative** sentiments
- Analyzes people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics, and their attributes.

# Opinion Mining and Sentiment Analysis

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

# Terminology

- Sentiment Analysis  
is more widely used in industry
- Opinion mining / Sentiment Analysis  
are widely used in academia
- Opinion mining / Sentiment Analysis  
can be used interchangeably

# **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. ... ”

# 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**.

+Positive  
Opinion

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

(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

# Why are opinions important?

- “Opinions” are key influencers of our behaviors.
- Our beliefs and perceptions of reality are conditioned on how others see the world.
- Whenever we need to make a decision, we often seek out the opinion of others.

In the past,

## – Individuals

- Seek opinions from friends and family

## – Organizations

- Use surveys, focus groups, opinion pools, consultants

# Word-of-mouth on the Social media

- Personal experiences and opinions about anything in reviews, forums, blogs, micro-blog, **Twitter**.
- Posting at social networking sites, e.g., **Facebook**
- Comments about articles, issues, topics, reviews.

# Social media + beyond

- Global scale
  - No longer – one's circle of friends.
- Organization internal data
  - Customer feedback from emails, call center
- News and reports
  - Opinions in news articles and commentaries

# Applications of Opinion Mining

- **Businesses and organizations**
  - Benchmark products and services
  - Market intelligence
    - Business spend a huge amount of money to find consumer opinions using consultants, surveys, and focus groups, etc.
- **Individual**
  - Make decision to buy products or to use services
  - Find public opinions about political candidates and issues
- **Ads placements**: Place ads in the social media content
  - Place an ad if one praises a product
  - Place an ad from a competitor if one criticizes a product
- **Opinion retrieval**: provide general search for opinions.

# 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

# 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

# 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

# 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/>

The screenshot displays the Attensity website homepage. At the top, there's a navigation bar with the Attensity logo, a language selector set to 'English', and links for Contact, Resources, Support, Blog, and a search bar. Below this is a secondary navigation bar with links for Products, Solutions, Services, Customers, and Partners. The main content area features a large central banner with the headline 'Your real-time window into the social web.' and a quote from Yahoo! praising Attensity's analytics. To the left of the banner is a sidebar with links to Social Analytics, Social Response, Customer Analytics, Industry Solutions, and Why Attensity. To the right, there are several overlapping dashboard previews showing various charts, graphs, and social media feeds. At the bottom, there are four distinct sections: 'Attensity for Marketing' and 'Attensity for Customer Service' buttons, a 'Success Story' for JetBlue Airways with a 'Download Now' button, an 'About Attensity' section stating it's the leading provider of social analytics, and a 'Watch Video' section for the 'Command Center Video'.

Attensity Home Page | Attensity

www.attensity.com/home/

Select your language **English**

Contact Resources Support Blog Search

Products Solutions Services Customers Partners

## Your real-time window into the social web.

*"Teaming with a leading analytics provider like Attensity offers Yahoo! a great opportunity to deliver the key news and analysis that matter."*  
– Yahoo!

**Learn More**

**Social Analytics**

**Social Response**

**Customer Analytics**

**Industry Solutions**

**Why Attensity**

**Attensity for Marketing**

**Attensity for Customer Service**

**Attensity for IT**

**Effectiveness of your social marketing strategies:**

**Success Story**

JetBlue Airways

**Listen.**

**About Attensity**

Attensity is the leading provider of social analytics and engagement solutions.

**Watch Video**

Command Center Video

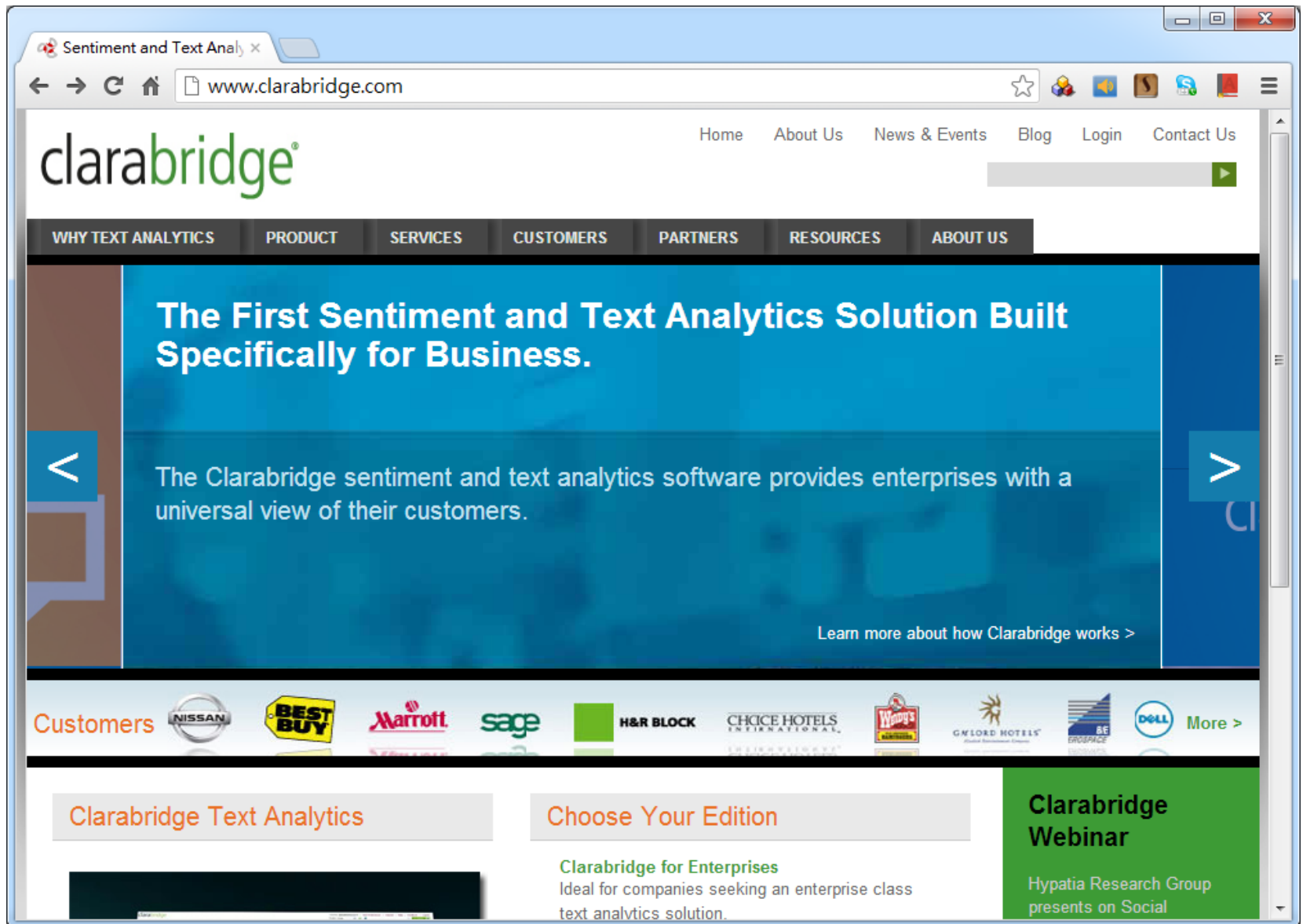
**ATTENSITY**

www.attensity.com/home/#fragment-1

<http://www.youtube.com/watch?v=4goxmBEg2lw#/>

# Clarabridge: Sentiment and Text Analytics Software

<http://www.clarabridge.com/>



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

<http://www.radian6.com/>

Social Media Monitoring x

www.radian6.com

Country 1 888 672 3426 About Radian6 Contact CUSTOMER LOGIN Search GO

salesforce radian6

How We Help What We Sell See Demo Free Resources Training & Support

The Social Enterprise.  
Get closer to your customer.  
Learn how >

Have Us Contact You

Live Demo

Free Trial

Chat & find out more.

Offline. Leave us a message

**Sales** The social web is a goldmine of untapped sales opportunities. Let us help you realize your potential. [Learn more >](#)

**Marketing** Brands are now the sum of the conversations about them. We can help you hear what's being said. [Learn more >](#)

**Customer Service** Take your customer service where your consumers are gathering. Respond to issues voiced on the social web. [Learn more >](#)

**Newsletter** Sign up and get the regular Radian6 goods. Enter email address [GO](#)

Mashable named Radian6's Co-founder Chris Ramsey one of five masterminds redefining social media

JUST Get the Skinny

WEBINAR / June 7th at 2pm est


CASE STUDY

radian6 Community

[http://www.youtube.com/watch?feature=player\\_embedded&v=8i6Exg3Urg0](http://www.youtube.com/watch?feature=player_embedded&v=8i6Exg3Urg0)

Social Media Monitoring x

← → ↻ ⬆ ⌚ www.sas.com/software/customer-intelligence/social-media-analytics/ ☆ 📱 📧 📄 📅 🔍

 **THE POWER TO KNOW.**

Providing software solutions since 1976

Log In Worldwide Sites ▼ Contact Us 📧 📱 📄 📅 Follow Us

NEWS EVENTS CONSULTING CAREERS RESOURCE CENTER

**SEARCH**

Home Products & Solutions Customer Success Partners Company Support & Training

**PRODUCTS & SOLUTIONS / SOCIAL MEDIA ANALYTICS**

**Products and Solutions**

- Industries
- Small and Midsize Business
- Nonprofit Organizations
- Analytics
  - Business Analytics
  - Business Intelligence
  - Customer Intelligence
    - Strategy & Planning
    - Information & Analytics
    - Orchestration & Interaction
    - Customer Experience
      - Customer Experience Analytics
      - Social Media Analytics**
      - Web Analytics

- Financial Intelligence
- Foundation Tools
- Fraud & Financial Crimes
- Governance, Risk & Compliance
- High-Performance Analytics
- Human Capital Intelligence
- Information Management
- IT & CIO Enablement

**SAS® Social Media Analytics**  
Integrate, archive, analyze and act on online conversations

Overview

Benefits

Features

Demos & Screenshots

System Requirements


SAS Social Media Analytics is an enterprise-hosted, on-demand solution that integrates, archives, analyzes and enables organizations to act on intelligence gleaned from online conversations on professional and consumer-generated media sites. It enables you to attribute online conversations to specific parts of your business, allowing accelerated responses to marketplace shifts.

Based on your unique business challenges and enterprise goals, SAS can provide a tailored implementation that's hosted and managed by [SAS Solutions OnDemand](#).


**Benefits**


- Analyze conversation data.
- Identify advocates of, and threats to, corporate reputation and brand.
- Quantify interaction among traditional media/campaigns and social media activity.
- Establish a platform for social CRM strategy.


**Product Demo**




**Questions?**

 Phone

 Contact Form



**White Paper**



Text Analytics for Social Media:  
Evolving Tools for an Evolving Environment

**Download Now**

**SAS® Social Media Analytics**

**» Overview**

**RESOURCES**

**» Fact Sheet (PDF)**

**» Solution Brief (PDF)**

**» White Papers**

23

What do tweeples think ab x

www.tweetfeel.com/index.php#iPhone4s

FAQ | Contact Us

# tweetfeel

|| iPhone4s Search

Try some Twitter trends: [Tomorrow is June](#) [H&M](#) [Defense of Marriage Act](#) [Diddy's](#) [Bloomberg](#) [UCLA](#) [ESPN](#)

 40  41 = 51%

Those are all the results available right now. Try again or try another term to see how people feel towards it.  
Got questions? [Read our FAQ.](#)

 RT @jigglinjello: This 12 year old has an iPhone4s wtf

 So my 9 year old little sister has a iPhone4s . Wtf bruh?!

 This 12 year old has an iPhone4s wtf

 So my sister has a android and i dont even have a phone and she gets a brand new iPhone4s - \_\_\_\_ - #Wtf

 iPhone4s is funny ass a bitch

 -Ohwell .. a new iPhone4s won't hurt , aha.

[Read our FAQ](#) [Legal Stuff](#) [100% Guarantee](#)  [Share](#)

Follow us  Email us  Brought to you by **conversion**  Powered by **twitter**

Browser window: Tweet Sentiments - Know ' x

Address bar: tweetsentiments.com/analyze?utf8=✓&q=iphone4s&topic=true&commi

Language: Eng

Logo: **tweet** Sentiments v0.8.6

Navigation: Dashboard Analyze Recent Top 1000 Users Maps Charts About Login

### Tweet Sentiment Analysis

Input:  User ☐ Topic ☒


Feedback

Sentiments Index[0-100]: 50.00

Positive: 8  
Negative: 1  
Neutral: 13  
Total Analyzed: 0

Profile  
Education:  
Flamboyance:  
Slang:  
Gender:  
Age:

50.00



1 Niggie, if i dont jailbreak my Iphone4s, u mean i cant play Street fighter4? wtf @louistekneeq 1 minute ago by Flow\_Show

2 リフォロー100%です！ #Arashi #韓fan #wstcg #板野友美 #followme #アメブロ #kimiboku #ntb #autofollow #apple #mt2 #相互フォローの輪 #iphone4s #Arashi #twitter #前田敦子 1 minute ago by designer\_sayaka

3 Rumor has it that #Apple #iPhone5 will have a 4" screen (compared 2 3.5" in #iPhone4S ), gving direct competition 2... <http://t.co/zsIGPxbX> 1 minute ago by abhay01007

4 RT @BallMe\_Dollar: The iPhone4S look better then the iPhone5. 1 minute ago by SuckMy\_TwitNuts

5 #iphoneography #iphonography #bahrain #iphone4s #photography #sunrise #scenery #beautiful #sun <http://t.co/Ee4VNmsL> 2 minutes ago by thementaldawg

i-Buzz網路口碑研究中心. 網 x

www.i-buzz.com.tw/index.asp

iBuzz中國 | 關於我們

歡迎光臨。請 [加入會員](#) 或 [登錄](#)



艾博司網路口碑研究中心  
Internet Buzz Research Center

Home

議題分析

排行榜

研究專題

數據報告

產品服務

解決方案

7-ELEVEN

純發文 0%

文章連結 10%

影片連結 17%

照片連結 73%

議題分析

血汗醫院全民皆憤 網路熱門

研究專題

便利商店衝人氣 臉書加持少

數據報告

永慶房屋穩據房仲龍頭 資訊

f 臉書粉絲團排行榜

更多

政治人物

名人

演藝人員

馬英九總統

蔡英文 Tsai

蘇貞昌

粉絲數 1419591

粉絲數 669928

粉絲數 317035

排行榜

Traffic 交通



第1名

FORD (FIESTA)

第2名

PORSCHE (911)

第3名

SUBARU (legacy)

更多

Traffic 交通



第1名

進口車

吵了這麼久，就是不要「超速」起車，還不懂 (224)

讓一下，很難嗎。。。 (184)

請問大家認為這樣要賠多少?? (148)

當Audi 一直被鍵盤魔人攻擊時，表示已經威 (134)

ELANTRA系列 vs ALTIS 系列之比較 (109)

更多

高級進口車

吵了這麼久，就是不要「超速」起車，還不懂 (224)

研究專題

便利商店衝人氣 臉書加持少不得

研究專題

LUMIX與Samsung相機口碑比拼 粉絲只顧「按讚」不「留言」



社交媒介監看

了解社群監測

26

OpView 服務總覽 | 意藍科技 x

← → ↻ ↑ [www.eland.com.tw/solutions](http://www.eland.com.tw/solutions) ☆

**eLAND** Cloud Services 意藍科技

產品 雲端服務 臺銀標專區 我們的客戶 新聞與活動 支援 關於意藍

## OpView Service

您的輿論觀測站

連上OpView, 品牌形象、輿論觀點監測好輕鬆

OpView 服務總覽

搜尋

### OpView 服務總覽

什麼是OpView服務?

OpView是協助您蒐集、處理、分析網路資訊的雲端服務。

OpView服務, 猶如提供您無盡網路情報能量的發電廠。以雲端架構為基礎, OpView服務蒐集、處理、分析各類型網路資訊與情報, 並以雲端服務平台供客戶使用。OpView服務的資料涵蓋範圍包括台灣最具代表性的新聞網站、

相關文件下載

- OpView Insight DM
- OpView品牌口碑雷達 DM
- OpView品牌口碑雷達訂購單

更多

# 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 politicians' 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.

# Abstraction (1) :

## 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

# 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

# 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.”

# 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

# Sentiment Analysis

## VS.

# Subjectivity Analysis

Sentiment Analysis	Subjectivity Analysis
Positive	Subjective
Negative	
Neutral	Objective

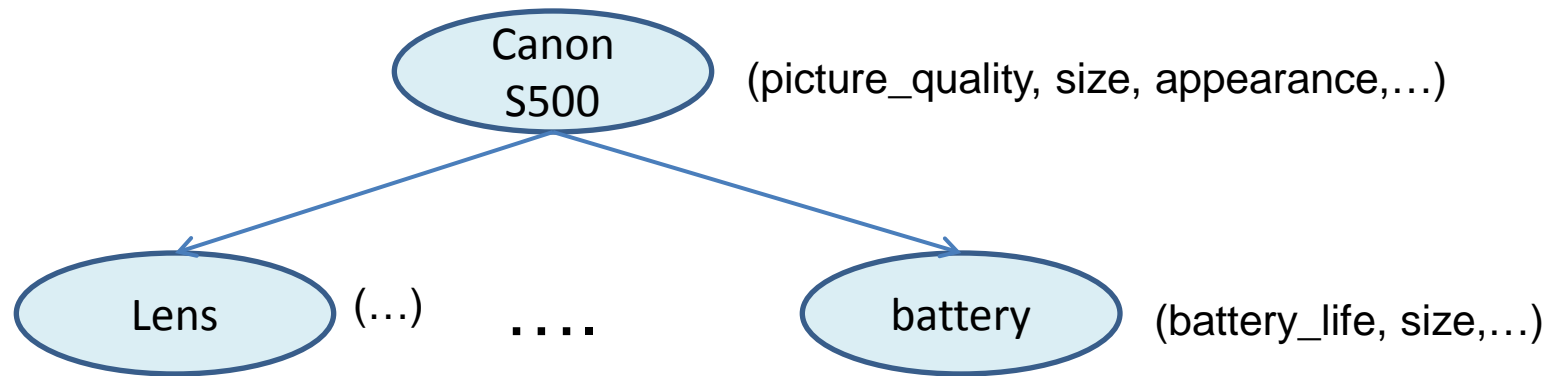
# A (regular) opinion

- **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**

# 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

# Entity and aspect



# 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.

# 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

# Terminologies

- **Entity**: object
- **Aspect**: feature, attribute, facet
- **Opinion holder**: opinion source
- **Topic**: entity, aspect
- Product features, political issues

# 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.

# Emotion

- Six main emotions
  - Love
  - Joy
  - Surprise
  - Anger
  - Sadness
  - Fear

# Abstraction (2): opinion summary

- With a lot of opinions, a summary is necessary.
  - A multi-document summarization task
- For factual texts, summarization is to select the most important facts and present them in a sensible order while avoiding repetition
  - 1 fact = any number of the same fact
- But for opinion documents, it is different because opinions have a quantitative side & have targets
  - 1 opinion  $\leftrightarrow$  a number of opinions
  - Aspect-based summary is more suitable
  - Quintuples form the basis for opinion summarization

# An aspect-based opinion summary

*Cellular phone 1:*

Aspect: **GENERAL**

Positive: 125 <individual review sentences>

Negative: 7 <individual review sentences>

Aspect: **Voice quality**

Positive: 120 <individual review sentences>

Negative: 8 <individual review sentences>

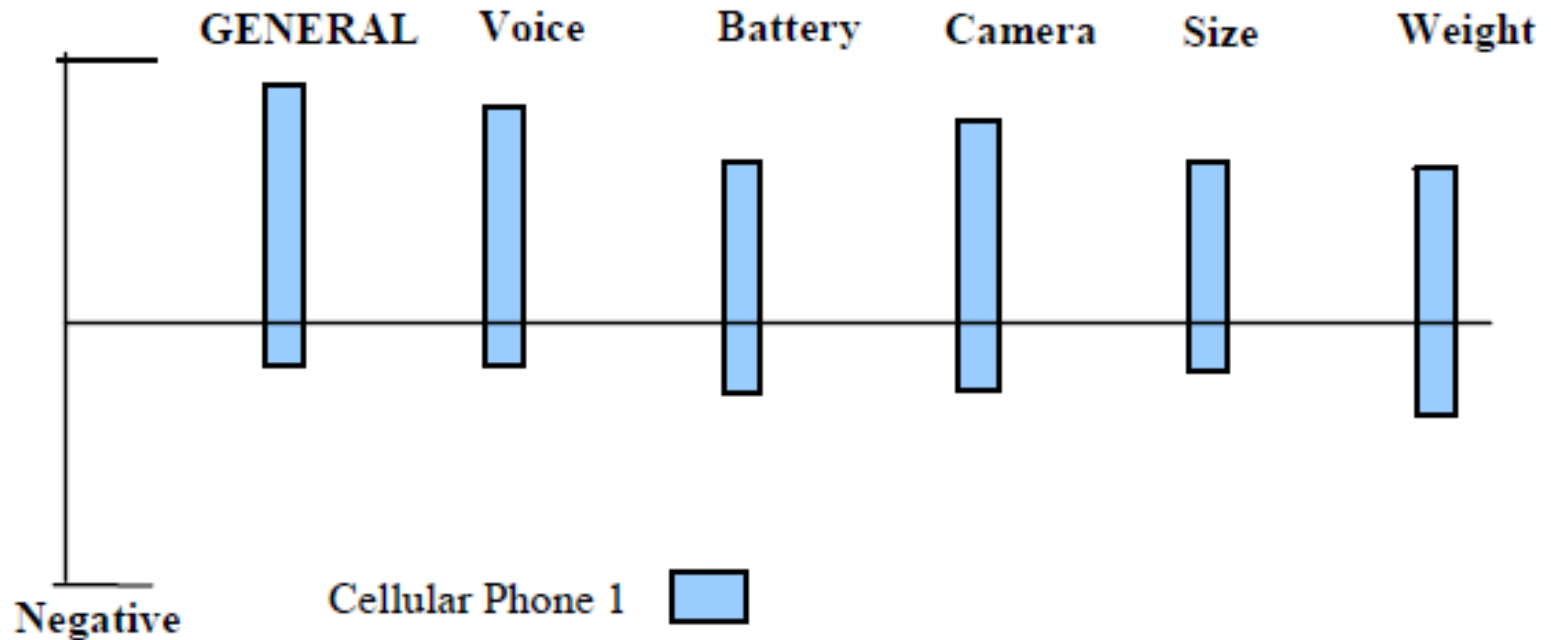
Aspect: **Battery**

Positive: 80 <individual review sentences>

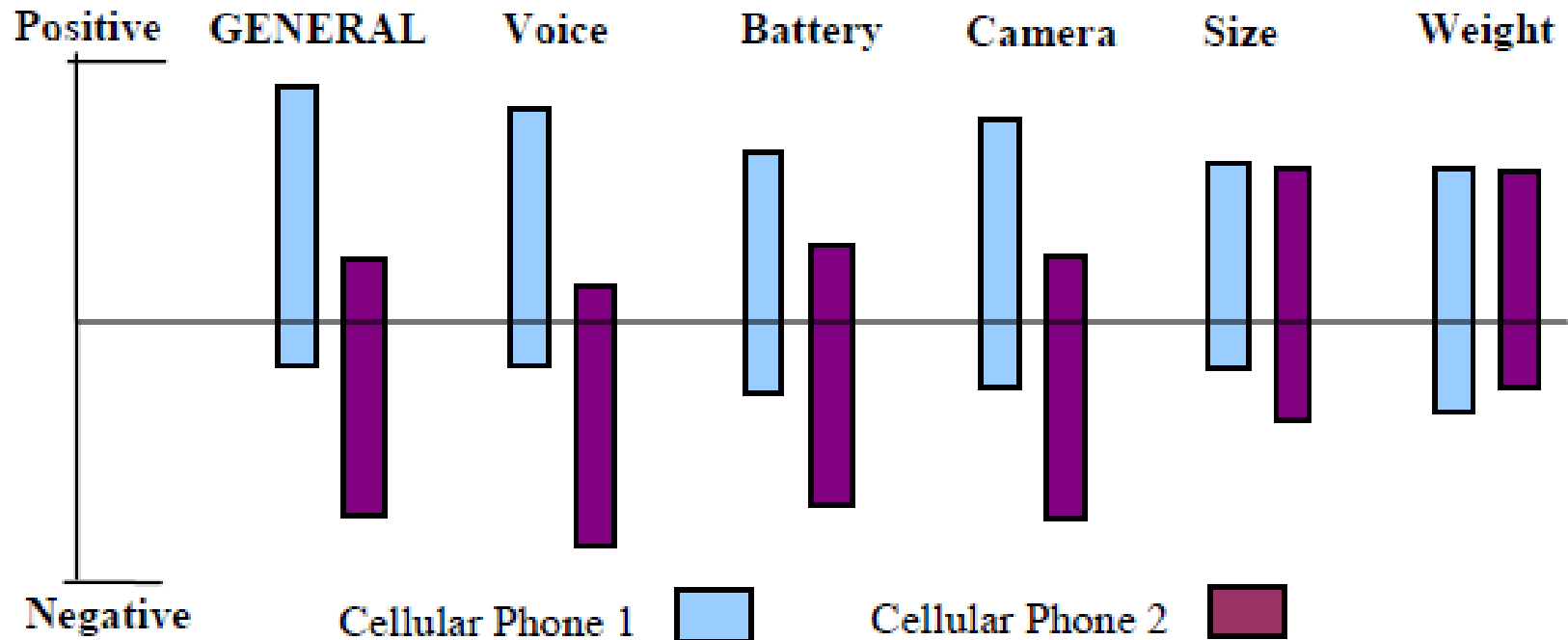
Negative: 12 <individual review sentences>

...

# Visualization of aspect-based summaries of opinions



# Visualization of aspect-based summaries of opinions



# Classification Based on Supervised Learning

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

# 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*

# 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*

# ***Rules of opinions***

## **Syntactic template**

<subj> passive-verb

<subj> active-verb

active-verb <dobj>

noun aux <dobj>

passive-verb prep <np>

## **Example pattern**

<subj> was satisfied

<subj> complained

endorsed <dobj>

fact is <dobj>

was worried about <np>

# 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
Airolidi, 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)

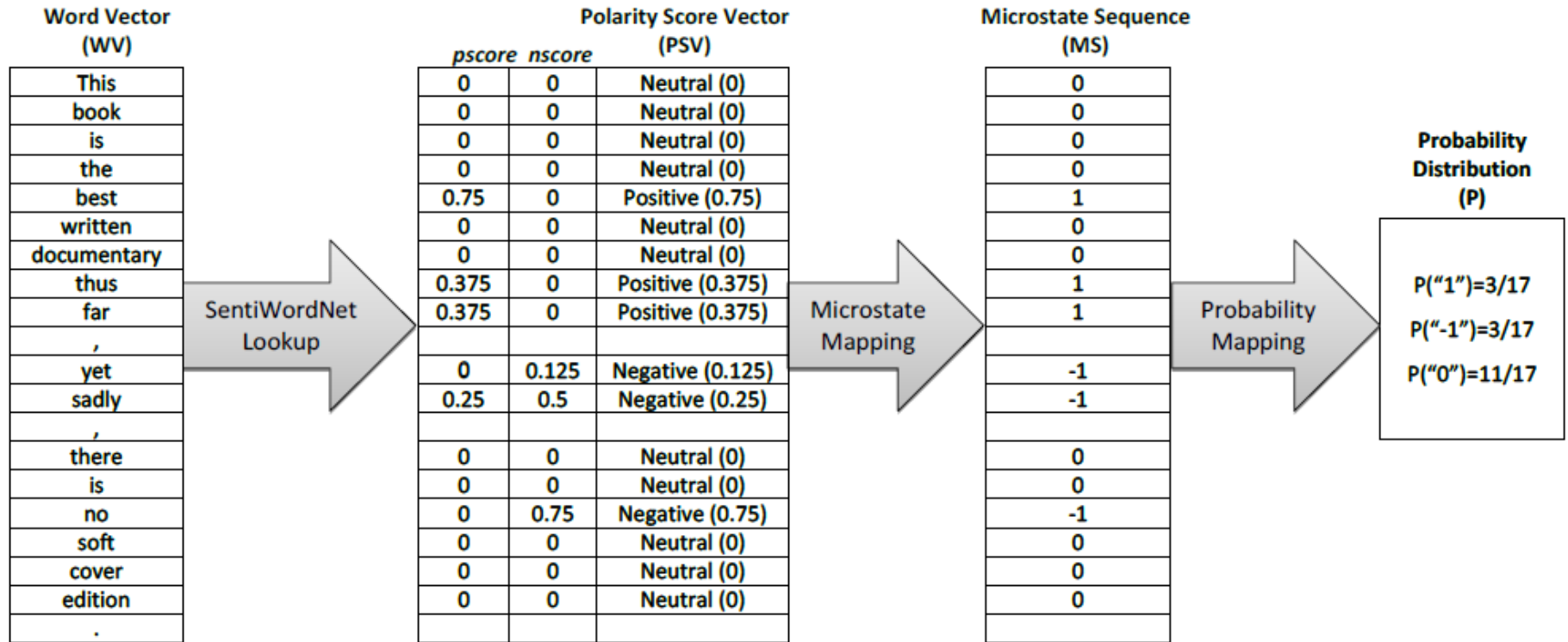
# Word-of-Mouth (WOM)

- “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.”

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
.	.

# Conversion of text representation



# Datasets of Opinion Mining

- Blog06
  - 25GB TREC test collection
  - [http://ir.dcs.gla.ac.uk/test collections/access to data.html](http://ir.dcs.gla.ac.uk/test%20collections/access%20to%20data.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)

# 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](http://www.keenage.com/html/c_bulletin_2007.htm)

# 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

# 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

# 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 ...

# 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 ...

# 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

# Summary

- Introduction to Opinion Mining and Sentiment Analysis
- Social Media Monitoring/Analysis
- Resources of Opinion Mining

# References

- Bing Liu (2011) , “Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data,” 2<sup>nd</sup> Edition, Springer.  
<http://www.cs.uic.edu/~liub/WebMiningBook.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](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,” Pearson , Ninth Edition, 2011.