Social Word-of-Mouth and Web Content Analysis
(社群口碑與網路文本分析)

戴敏育
Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail. tku.edu.tw/myday/
2012-06-04

Tamkang University

淡江大學

1

Outline

1. 社群媒體的特性
2. 社群媒體發展趨勢
3. 社群媒體商業應用
4. 社群口碑趨勢分析
5. 網路文本分析

社群媒體的特性

#1 Activity on the Web?

Social Media

4

Social Media Management Pyramid

Social Media Marketing For Business

Marketing 4P to **4C**

- **P**roduct → **C**ustomer solution
- **P**rice → Customer **C**ost
- **P**lace → **C**onvenience
- **P**romotion → **C**ommunication

Source: Kotler and Keller (2008)

Four Pillars of Social Media Strategy C²E²

Social Media Strategy

- Communication
- Collaboration
- Education
- Entertainment

Social Media Can Help Orchestrate Three Spheres to Influence to Boost a Company's Innovation Efforts

Internal

Innovation

Trusted Network

The World

Source: Robert Wollan, Nick Smith, Catherine Zhou, The Social Media Management Handbook, John Wiley, 2011.

Examples of Social Media Selling Strategies in the Market Today

Strategy #3 – "Appealing to Influencers": Target Influencers Who Can Move the Masses

Strategy #1 – "Accessing social Consumers": Use Social Media as a New Channel to Individuals

Engaging the Advocates

"Pro-sumer" collaboration

Influencer-Led Development

Social Media Community

User Reviews

Social Media Wildfire

Creating Urgency/ Spontaneous Selling

Customers as "Community Organizers"

"Pass it along" promptions

Recruiting others/ Group Seles

Strategy #2 – "Engaging the Hive": Get Customers to Mobilize Their Personal Networks

Source: Robert Wollan, Nick Smith, Catherine Zhou, The Social Media Management Handbook, John Wiley, 2011.

Social Media Landscape

社會媒體 (social media)的定義
(Kaplan & Haenlein, 2010)

建立在Web 2.0概念與技術的基礎上，
以網路為平台的應用系統
(Internet-based applications)，
讓網路使用者可以
方便產生與交流使用者建立的內容
(user generated content; UGC)。

社會媒體服務
(Social Media Services)

提供使用者在網路環境中使用
社會媒體應用系統的線上服務
(online services)
Google+, Youtube, Facebook, Plurk

社群媒體發展趨勢

THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

SOCIAL LUMAscape

© LUMA Partners LLC 2012

Denotes acquired company

Web 2.0

- The popular term for advanced Internet technology and applications, including blogs, wikis, RSS, and social bookmarking.

- One of the most significant differences between Web 2.0 and the traditional World Wide Web is greater collaboration among Internet users and other users, content providers, and enterprises.

THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

- **REPRESENTATIVE CHARACTERISTICS OF WEB 2.0**
  - The ability to tap into the collective intelligence of users
  - Data is made available in new or never-intended ways
  - Web 2.0 relies on user-generated and user-controlled content and data
  - The virtual elimination of software-upgrade cycles makes everything a *work in progress* and allows rapid prototyping

THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

- Users can access applications entirely through a browser
- An architecture of participation encourages users to add value to the application
- A major emphasis on social networks and computing
- Strong support of information sharing and collaboration
- Rapid and continuous creation of new business models

THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

- **WEB 2.0 COMPANIES AND NEW BUSINESS MODELS**

- **social media**

  The online platforms and tools that people use to share opinions, experiences, insights, perceptions, and various media, including photos, videos, and music, with each other.

EXHIBIT 7.1   The Emergence and Rise of Mass Social Media
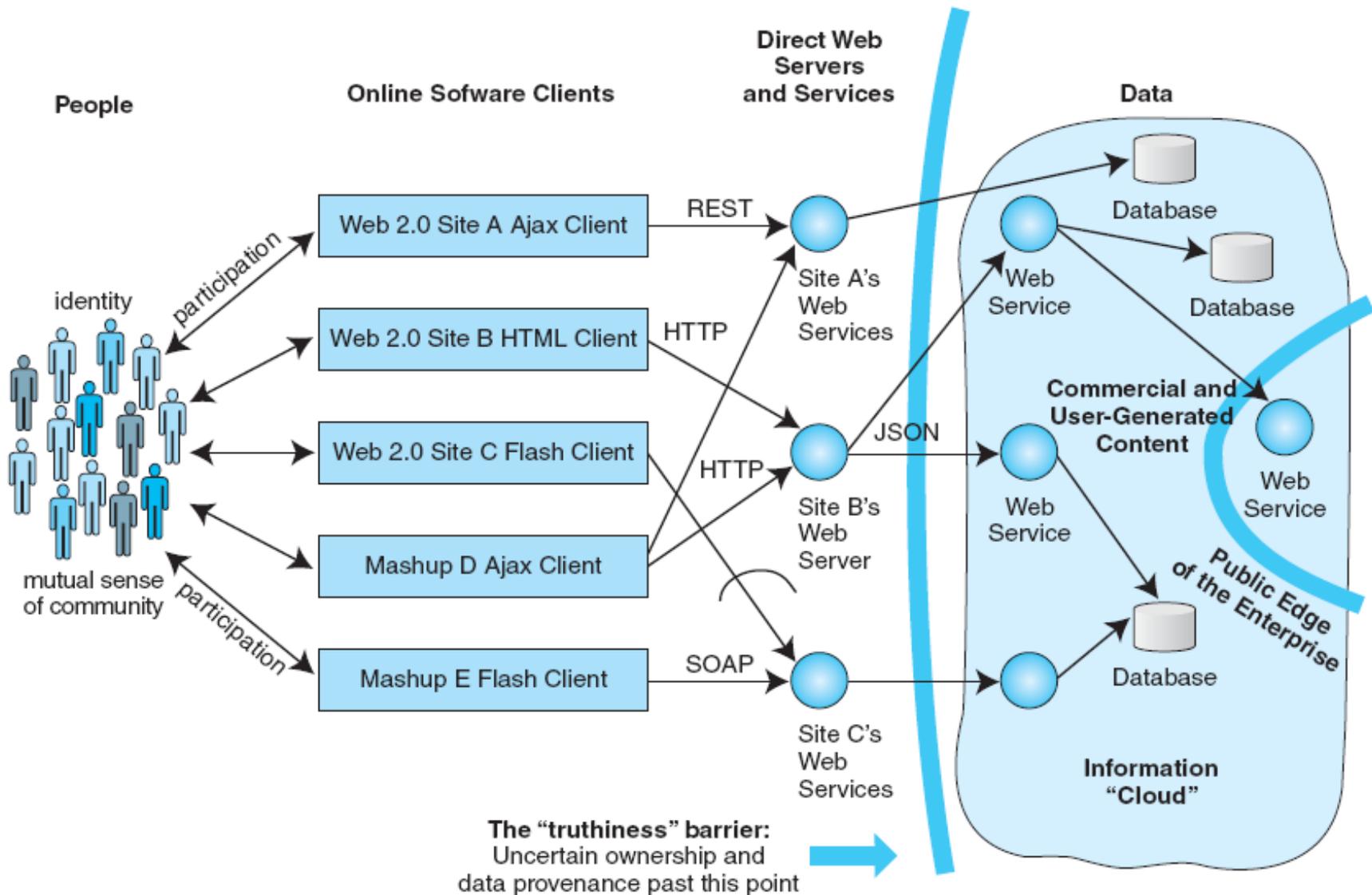
THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

- **INDUSTRY AND MARKET DISRUPTORS**
  - **disruptors**

    Companies that introduce a significant change in their industries, thus causing a disruption in normal business operations.

ONLINE SOCIAL NETWORKING: BASICS AND EXAMPLES

- **social networking**

  Social networks and activities conducted in social networks. It also includes activities conducted using Web 2.0 (e.g., wikis, microblogs) not within social networks.
  - **The Size of Social Network Sites**
  - **New Business Models**

EXHIBIT 7.4 **The Top Nine Social Networking Sites**

Site	Members
MySpace	253,145,404
Facebook	450,000,000
Windows Live Spaces	120,000,000
Friendster	90,000,000
hi5	80,000,000
Orkut	67,000,000
Classmates.com	50,000,000
Bebo	40,000,000
LinkedIn	70,000,000
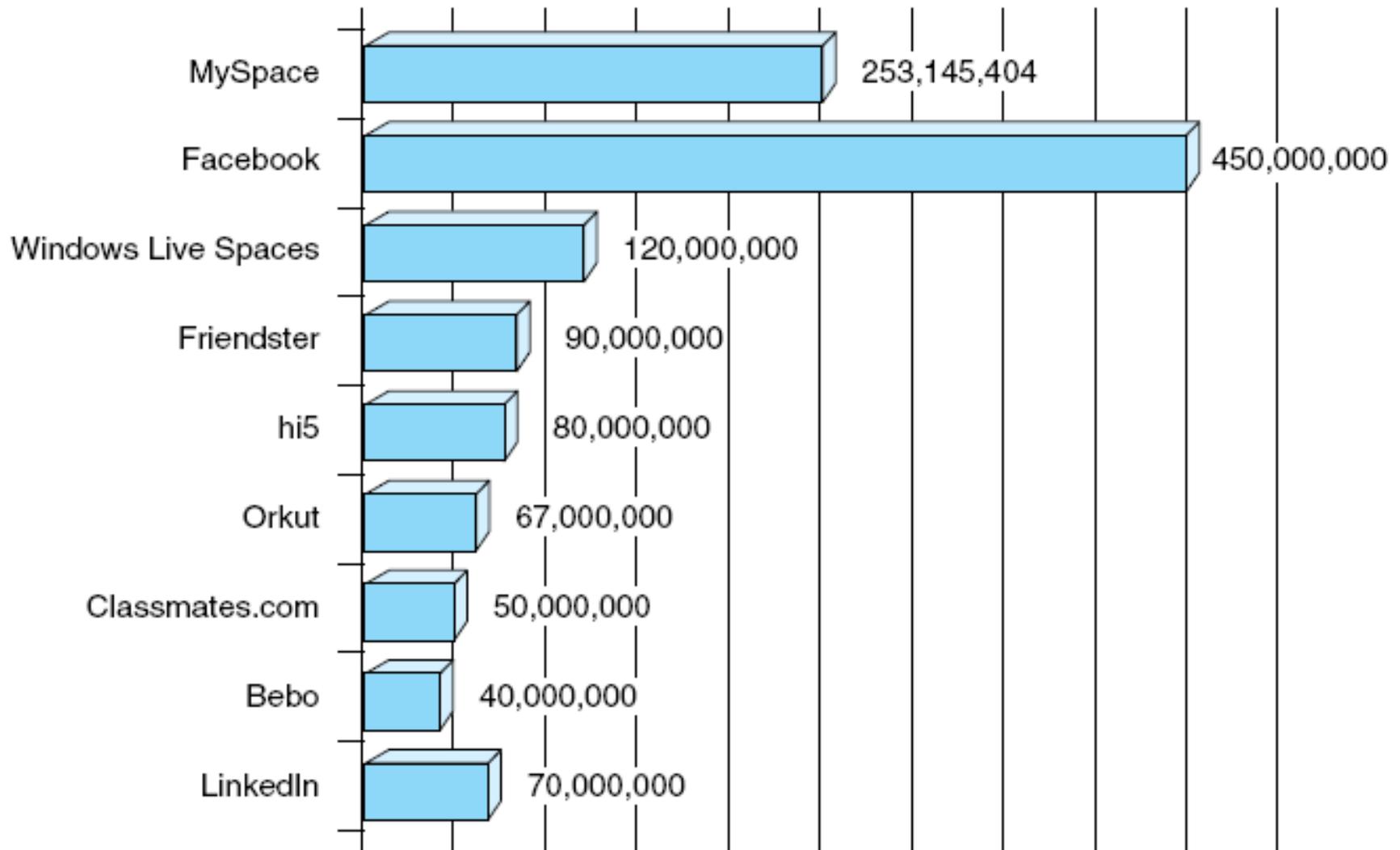
ONLINE SOCIAL NETWORKING: BASICS AND EXAMPLES

– **social network analysis (SNA)**

The mapping and measuring of relationships and information flows among people, groups, organizations, computers, and other information- or knowledge-processing entities. The nodes in the network are the people and groups, whereas the links show relationships or flows between the nodes. SNAs provide both visual and a quantitative analysis of relationships.

BUSINESS AND ENTERPRISE SOCIAL NETWORKS

- The major reasons to use or deploy a business social network are to:
    - Build better customer relationships
    - Improve knowledge management
    - Facilitate recruiting and retention
    - Increase business opportunities
    - Build a community
    - Gain expert advice
    - Improve trade show experiences
    - Improve communication and collaboration

THE FUTURE: WEB 3.0 AND WEB 4.0

- **Web 3.0**

  A term used to describe the future of the World Wide Web. It consists of the creation of high-quality content and services produced by gifted individuals using Web 2.0 technology as an enabling platform.

THE FUTURE: WEB 3.0 AND WEB 4.0

– **Semantic Web**

An evolving extension of the Web in which Web content can be expressed not only in natural language, but also in a form that can be understood, interpreted, and used by intelligent computer software agents, permitting them to find, share, and integrate information more easily.

THE FUTURE: WEB 3.0 AND WEB 4.0

– **Web 4.0**

The Web generation after Web 3.0. It is still mostly an unknown entity. However, it is envisioned as being based on islands of intelligence and as being ubiquitous.

– **Future Threats**

- Security concerns
- Lack of Net neutrality
- Copyright complaints
- Choppy connectivity

COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

- **WHY IS THERE AN INTEREST?**
  - Web 2.0 applications are spreading rapidly, and many of them cater to a specific *segment of* the population (e.g.,music lovers, travelers, game lovers, and car fans), enabling segmented advertising
  - Many users of Web 2.0 tools are young, and they will grow older and have more money to spend

COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

- **ADVERTISING USING SOCIAL NETWORKS, BLOGS, AND WIKIS**
  - **Viral (Word-of-Mouth) Marketing**
    - **viral blogging**

      Viral (word-of-mouth) marketing done by bloggers.
  - **Classified Ads, Job Listings, and Recruitment**
  - **Special Advertising Campaigns**
  - **Mobile Advertising**

COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

- **SHOPPING IN SOCIAL NETWORKS**

- **FEEDBACK FROM CUSTOMERS: CONVERSATIONAL MARKETING**
  - **Customer Feedback with Twitter**

COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

- **COMMERCIAL ACTIVITIES IN BUSINESS AND ENTERPRISE SOCIAL NETWORKS**
  - **Finding and Recruiting Workers**
  - **Management Activities and Support**
  - **Training**
  - **Knowledge Management and Expert Location**
  - **Enhancing Collaboration**
  - **Using Blogs and Wikis Inside the Enterprise**

EXHIBIT 7.5   Generating Revenue from Web 2.0 Applications

user contributions

subscriptions   advertising

Web Application

Mashup

platform

Rich User Experience

usage fees

API

other API

commissions

WOA

intel inside

The Enterprise

SOA

The Internet

Tenets
- Web as platform
- Beyond single device
- Data as the next "Intel Inside"
- Lightweight models
- Rich user experiences
- Harnessing collective intelligence
- Leverage the long tail

● = Monetization methods
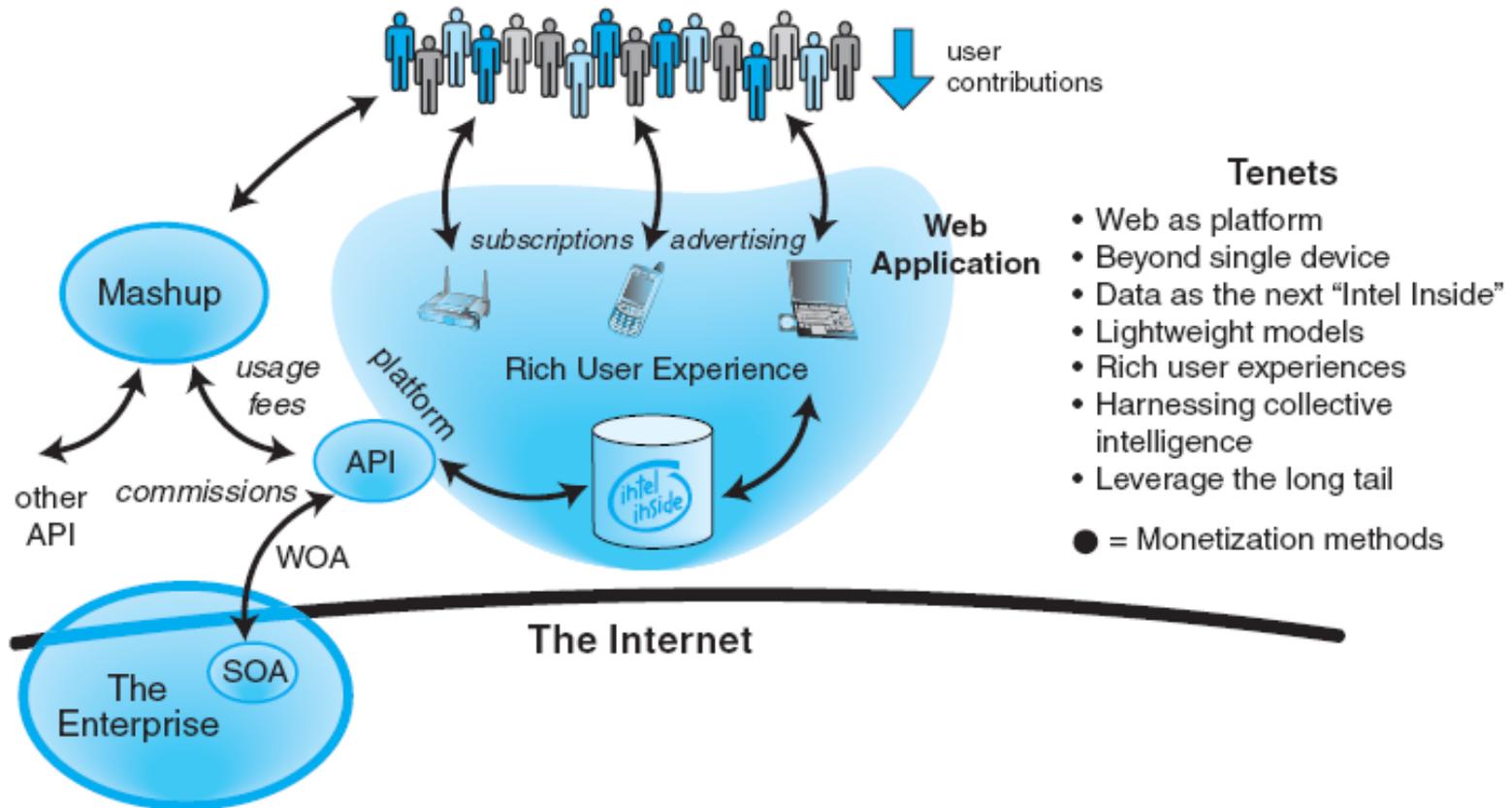
COMMERCIAL ASPECTS OF WEB 2.0 AND SOCIAL NETWORKING APPLICATIONS

- **REVENUE-GENERATION STRATEGIES IN SOCIAL NETWORKS**

  - **Increased Revenue and Its Benefit**

- **RISKS AND LIMITATIONS WHEN INTERFACING WITH SOCIAL NETWORKS**

- **JUSTIFYING SOCIAL MEDIA AND NETWORKING**

ENTERTAINMENT WEB 2.0 STYLE: FROM SOCIAL NETWORKS TO MARKETPLACES

- **MOBILE WEB 2.0 DEVICES FOR ENTERTAINMENT AND WORK**
  - **iPhone and Its Clones**

iPhone 3G

社群媒體商業應用

Social Media Word-of-Mouth Marketing

How to Start Buzz

- Identify influential individuals and companies and devote extra effort to them
- Supply key people with product samples
- Work through community influentials
- Develop word-of-mouth referral channels to build business
- Provide compelling information that customers want to pass along

Source: Kotler and Keller (2008)

Word-of-Mouth Marketing

- Person-to-person
- Chat rooms
- Blogs
- Twitter, Plurk
- Facebook
- Youtube

Elements in the Communications Process

Field of Experience

Sender's field

Receiver's field

Source: Kotler and Keller (2008)
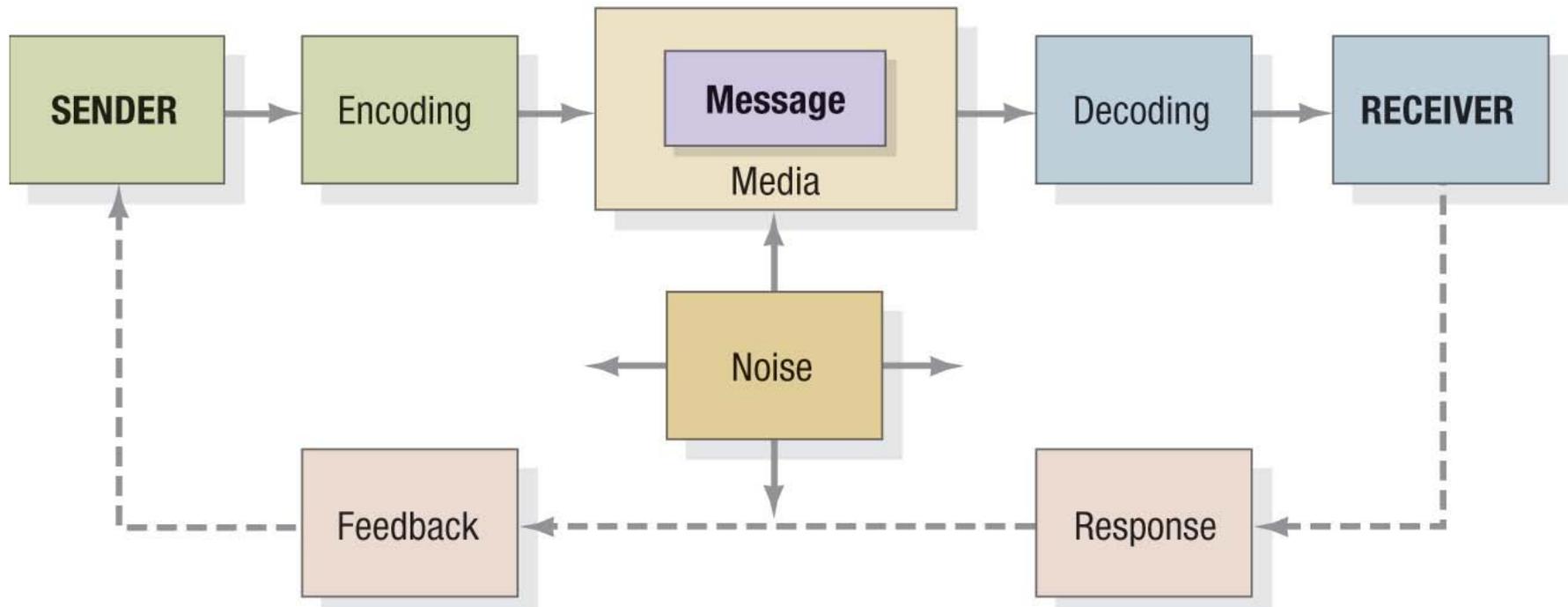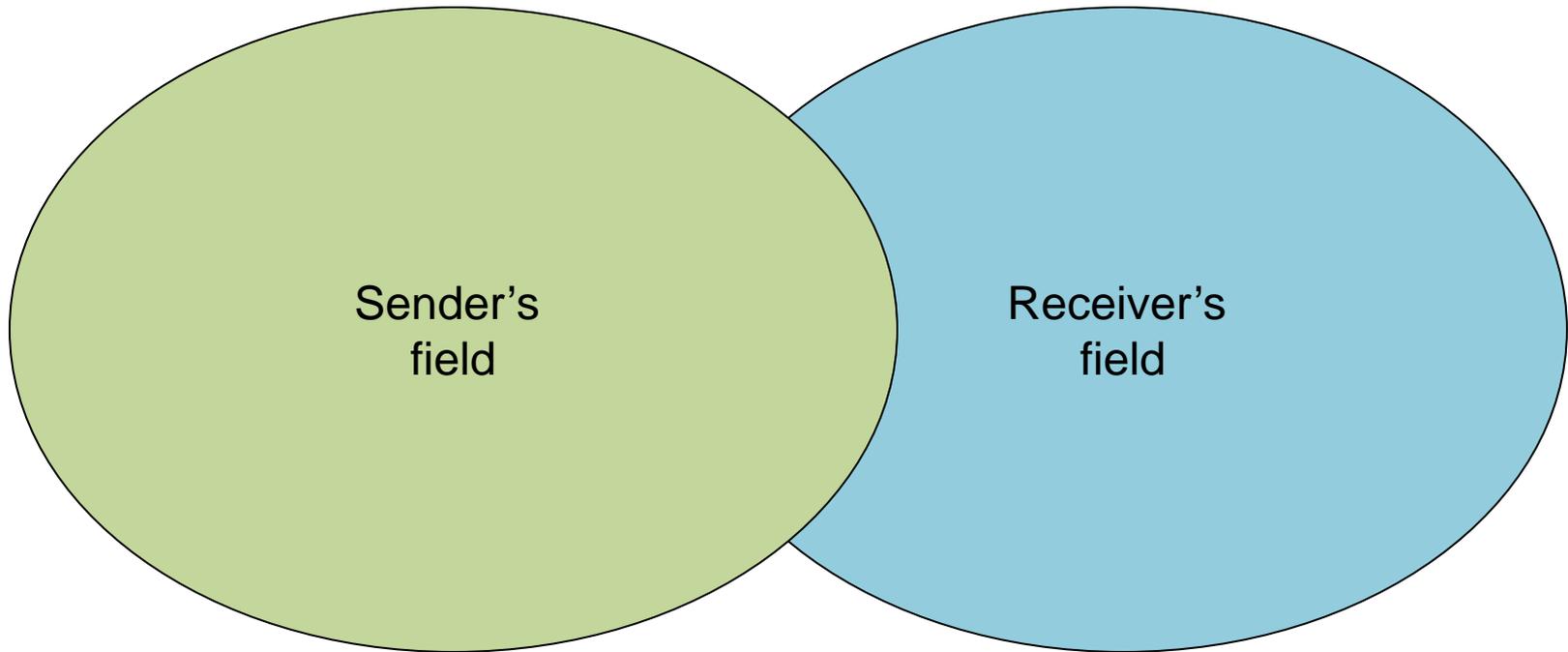
The Communications Process

Selective attention

Selective distortion

Selective retention

Social Media Marketing For Business

Social Media Marketing

- Scorecard for Social Media
  - 4 - Extremely Valuable
  - 3 - Very Valuable
  - 2 - Somewhat Valuable
  - 1 - Not Very Valuable
  - 0 - No Value

Scorecard for Social Media

Social Media Tool	Internal Value	External Value
Facebook	4 3 2 1 0	4 3 2 1 0
LinkedIn	4 3 2 1 0	4 3 2 1 0
Blogger	4 3 2 1 0	4 3 2 1 0
SlideShare	4 3 2 1 0	4 3 2 1 0
Wikipedia	4 3 2 1 0	4 3 2 1 0
Flickr	4 3 2 1 0	4 3 2 1 0
Picasa	4 3 2 1 0	4 3 2 1 0
iTunes	4 3 2 1 0	4 3 2 1 0
Podcast	4 3 2 1 0	4 3 2 1 0
Youtube	4 3 2 1 0	4 3 2 1 0
Twitter	4 3 2 1 0	4 3 2 1 0
Plurk	4 3 2 1 0	4 3 2 1 0

Scorecard for Social Media
4 - Extremely Valuable, 3 - Very Valuable, 2 – Somewhat Valuable, 1 - Not Very Valuable, 0 - No Value

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.

Source: Robert Wollan, Nick Smith, Catherine Zhou, The Social Media Management Handbook, John Wiley, 2011.

Accenture's SLOPE Model for Listening to the Social Voice of the Customer

Social Voice of the Customer

Synchronize → Listen & Learn → Optimize & Operationalize → Personalize & Propagate → Execution & Expectations

Source: Robert Wollan, Nick Smith, Catherine Zhou, The Social Media Management Handbook, John Wiley, 2011.

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.

Customers' Opinions About Operational versus Customer Experience Issues

Reactive, Reputation Management
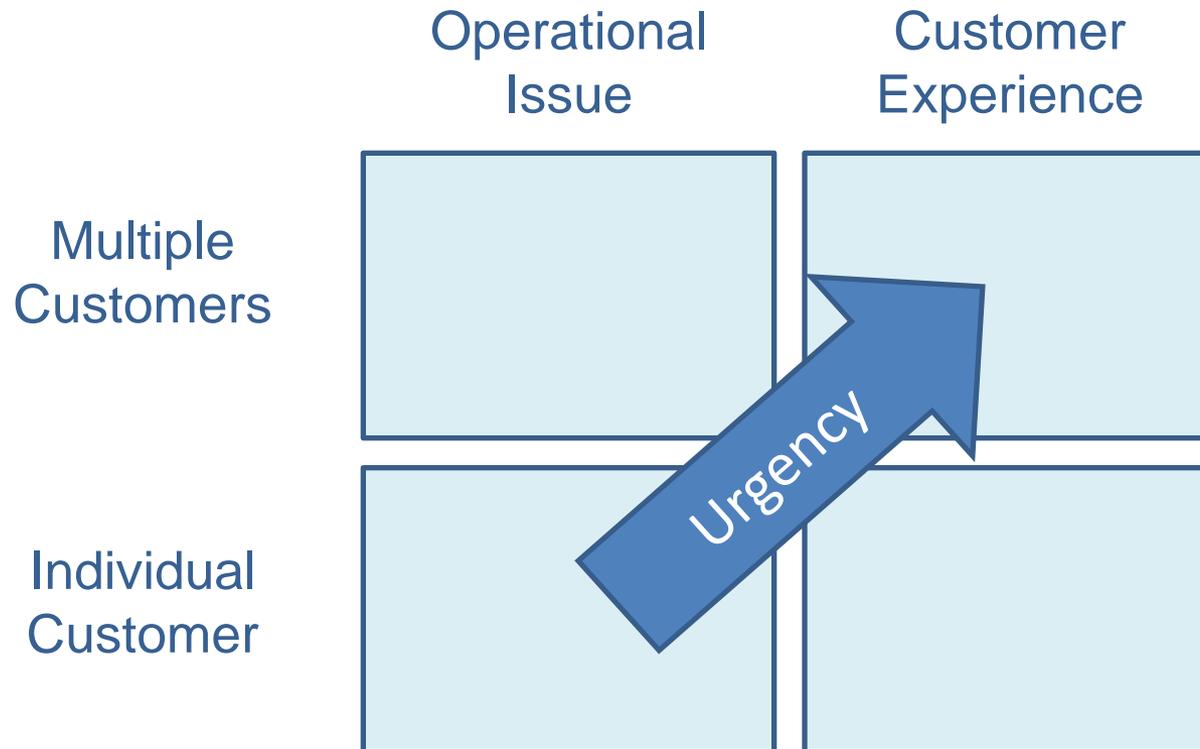
Source: Robert Wollan, Nick Smith, Catherine Zhou, The Social Media Management Handbook, John Wiley, 2011.

Social Media Can Help Orchestrate Three Spheres to Influence to Boost a Company's Innovation Efforts

Examples of Social Media Selling Strategies in the Market Today

Strategy #3 – "Appealing to Influencers": Target Influencers Who Can Move the Masses

Strategy #1 – "Accessing social Consumers": Use Social Media as a New Channel to Individuals

Engaging the Advocates

"Pro-sumer" collaboration

Influencer-Led Development

User Reviews

Social Media Wildfire

Creating Urgency/ Spontaneous Selling

Social Media Community

Customers as "Community Organizers"

"Pass it along" promptions

Recruiting others/ Group Seles

Strategy #2 – "Engaging the Hive": Get Customers to Mobilize Their Personal Networks

Source: Robert Wollan, Nick Smith, Catherine Zhou, The Social Media Management Handbook, John Wiley, 2011.

社群口碑趨勢分析

Word-of-Mouth
Social Media

word of mouth 1.00   social media   7.40

Case Study: LenovoClub CareerLife 職場人生

Case Study: LenovoClub CareerLife 職場人生

Case Study: LenovoClub CareerLife 職場人生

Case Study: LenovoClub CareerLife 職場人生

Case Study: LenovoClub CareerLife 職場人生

網路文本分析

ACM Categories and Subject Descriptors

- I.2.7 [Artificial Intelligence]
  - Natural Language Processing
    - Text analysis

- H.2.8 [Database Management]
  - Database Applications
    - Data mining

Text and Web Mining

- Text Mining: Applications and Theory

- Web Mining and Social Networking

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

- Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data

- Search Engines – Information Retrieval in Practice

Text Mining

Web Mining and Social Networking
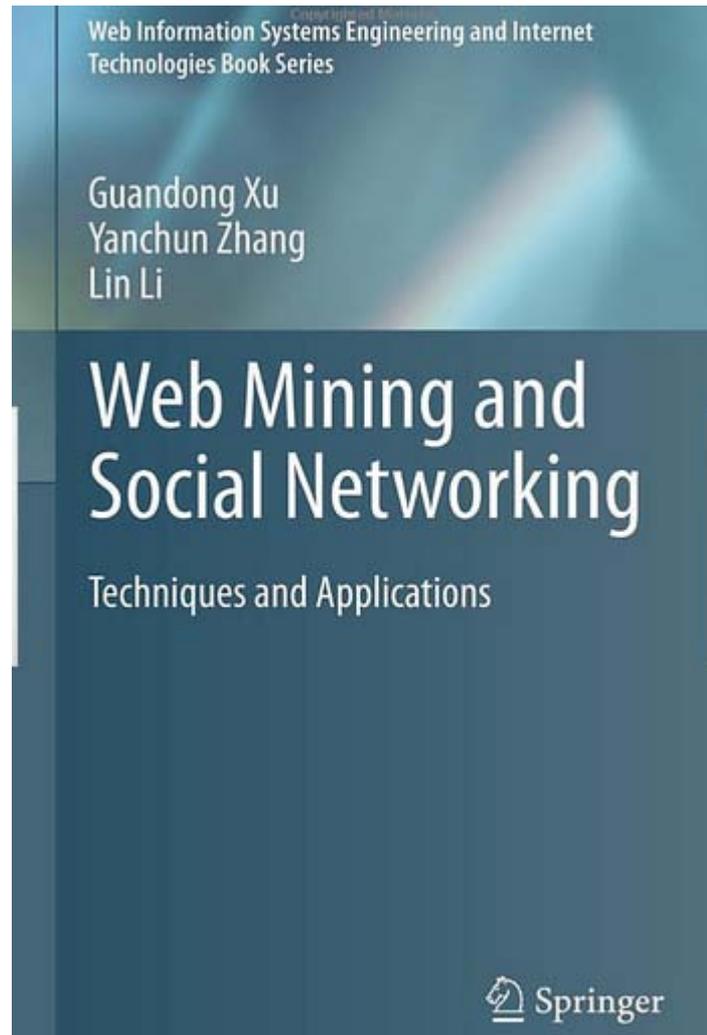
Web Information Systems Engineering and Internet Technologies Book Series

Guandong Xu
Yanchun Zhang
Lin Li

Web Mining and Social Networking

Techniques and Applications

Springer

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

Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites

O'REILLY®   Matthew A. Russell

Web Data Mining:
Exploring Hyperlinks, Contents, and Usage Data
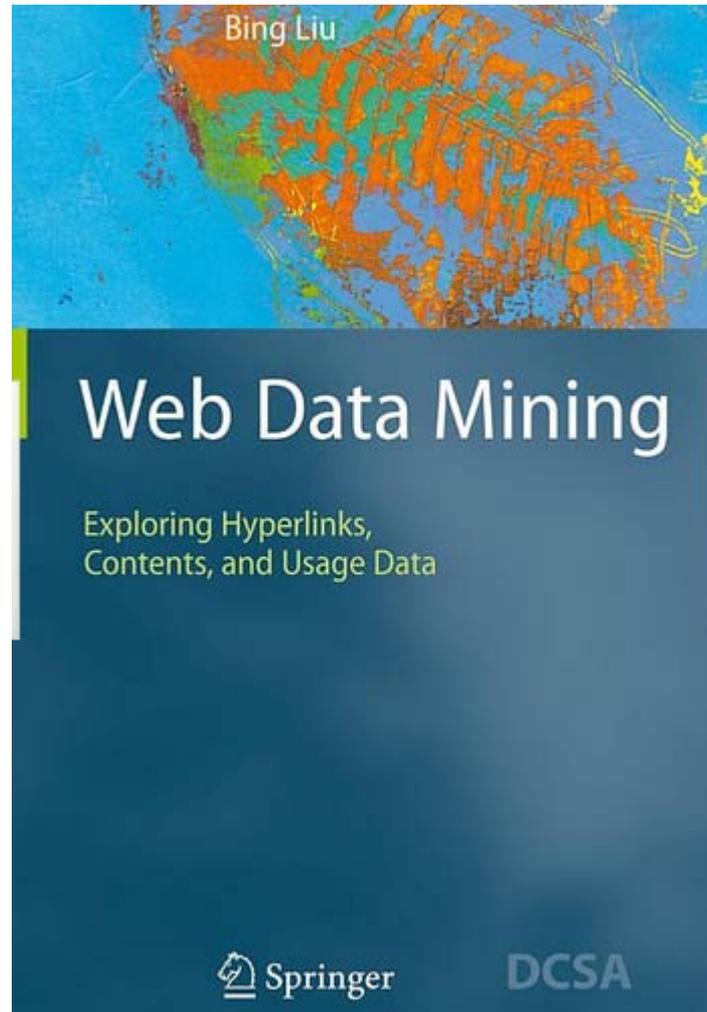
Web Data Mining

Exploring Hyperlinks, Contents, and Usage Data

1. Introduction
2. Association Rules and Sequential Patterns
3. Supervised Learning
4. Unsupervised Learning
5. Partially Supervised Learning
6. Information Retrieval and Web Search
7. Social Network Analysis
8. Web Crawling
9. Structured Data Extraction: Wrapper Generation
10. Information Integration
11. Opinion Mining and Sentiment Analysis
12. Web Usage Mining

Source: http://www.cs.uic.edu/~liub/WebMiningBook.html

Text Mining

- Text mining (text data mining)
  - the process of deriving high-quality information from text
- 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.

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

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

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,
  ets., 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 <u>iPhone</u> a few days ago.

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

(3) The <u>touch screen</u> was really **cool**.

+Positive Opinion

(4) The <u>voice quality</u> 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 **<u>expensive</u>**, 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,

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

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

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

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

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

http://www.radian6.com/

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

http://www.tweetfeel.com

http://tweetsentiments.com/

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

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

http://opview-eland.blogspot.tw/2012/05/blog-post.html

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.

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	
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

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
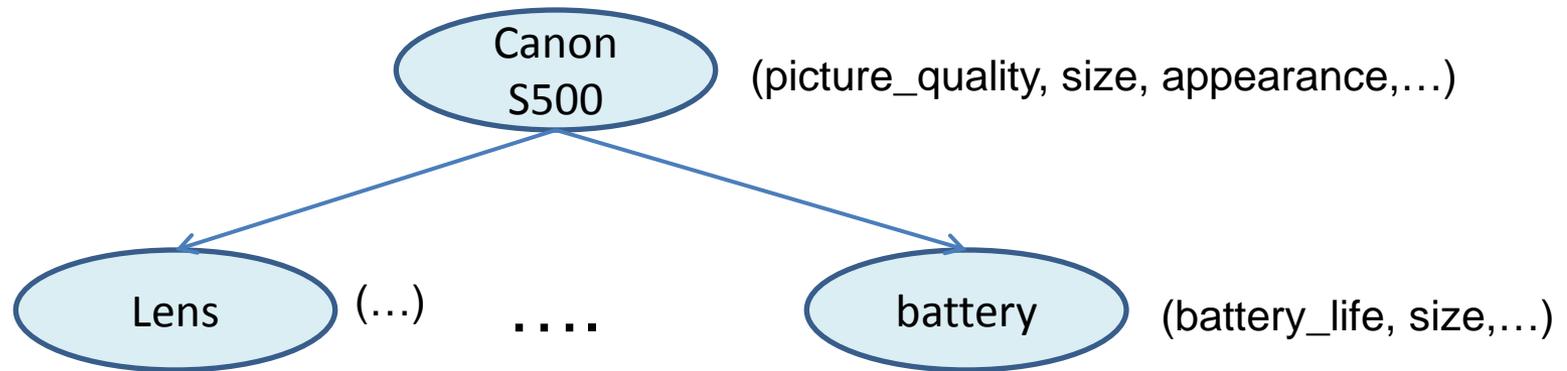
Canon S500 (picture_quality, size, appearance,…)

Lens (…) …. battery (battery_life, size,…)

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 <> 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
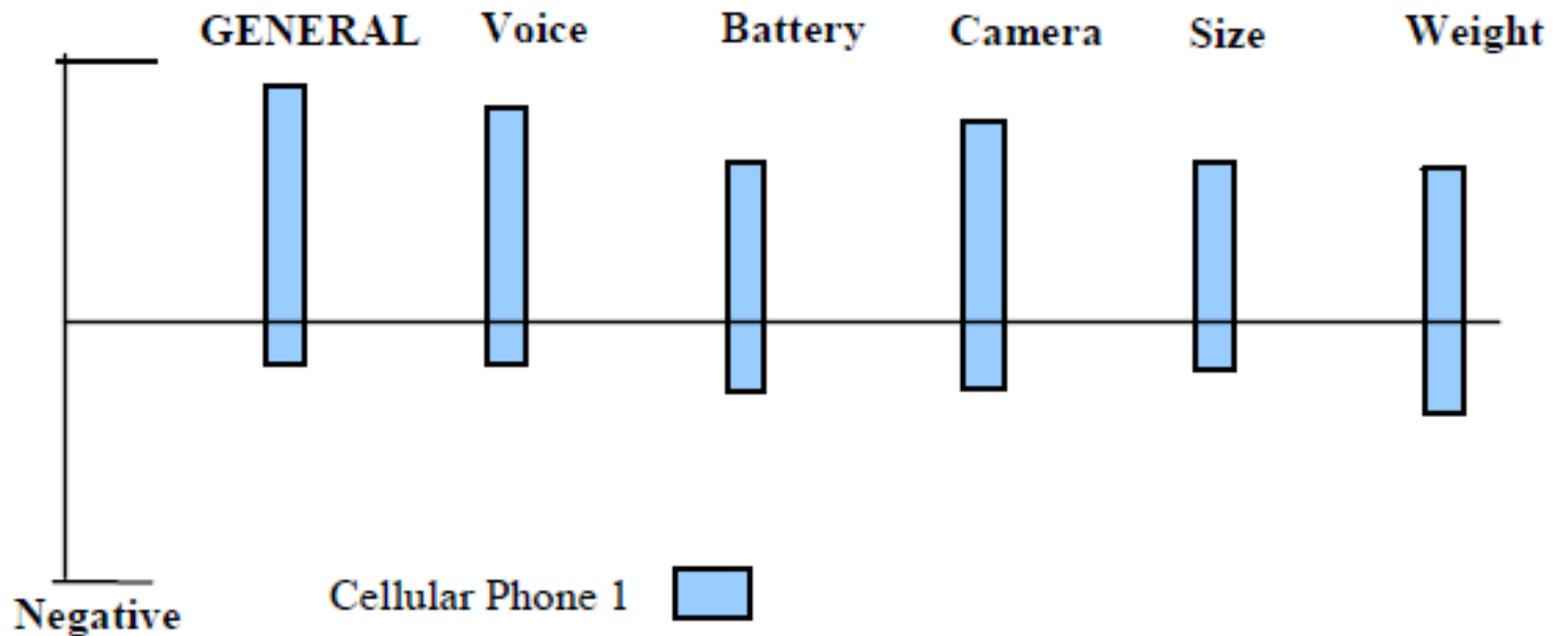
Visualization of aspect-based summaries of opinions
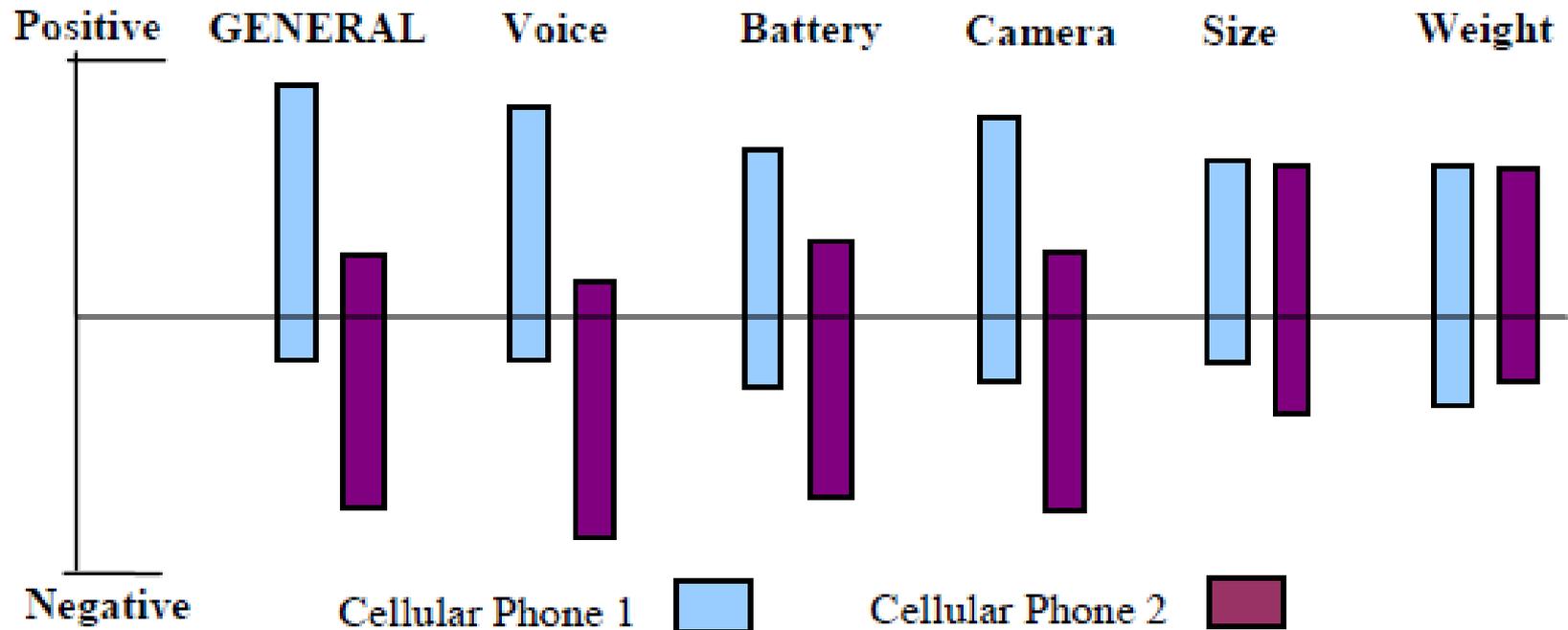
Classification Based on Supervised Learning

- 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>

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

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)

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,

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

Conversion of text representation

Word Vector (WV)				
This				
book				
is				
the				
best				
written				
documentary				
thus				
far				
,				
yet				
sadly				
,				
there				
is				
no				
soft				
cover				
edition				
.				

SentiWordNet Lookup →

Polarity Score Vector (PSV)		
pscore	nscore	
0	0	Neutral (0)
0	0	Neutral (0)
0	0	Neutral (0)
0	0	Neutral (0)
0.75	0	Positive (0.75)
0	0	Neutral (0)
0	0	Neutral (0)
0.375	0	Positive (0.375)
0.375	0	Positive (0.375)
0	0.125	Negative (0.125)
0.25	0.5	Negative (0.25)
0	0	Neutral (0)
0	0	Neutral (0)
0	0.75	Negative (0.75)
0	0	Neutral (0)
0	0	Neutral (0)
0	0	Neutral (0)

Microstate Mapping →

Microstate Sequence (MS)
0
0
0
0
1
0
0
1
1
-1
-1
0
0
-1
0
0
0

Probability Mapping →

Probability Distribution (P)

P("1")=3/17

P("-1")=3/17

P("0")=11/17

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)

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

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

1. 社群媒體的特性
2. 社群媒體發展趨勢
3. 社群媒體商業應用
4. 社群口碑趨勢分析
5. 網路文本分析

References

- Bing Liu (2011) , "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data," Springer, 2nd Edition, 2011, http://www.cs.uic.edu/~liub/WebMiningBook.html

- Efraim Turban, Ramesh Sharda, Dursun Delen (2011), "Decision Support and Business Intelligence Systems," Pearson , Ninth Edition, 2011.

- 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.

Social Word-of-Mouth and Web Content Analysis
(社群口碑與網路文本分析)

Q & A

戴敏育
Min-Yuh Day
Assistant Professor
Dept. of Information Management, Tamkang University

http://mail. tku.edu.tw/myday/

2012-06-04