



# Social Word-of-Mouth and Web Mining (社群口碑與網路探勘)

Time: 2012/10/31(Wed) 08:10-10:00

Place: 2F Rm.6203 ,杏春樓

College of Oral Medicine, Taipei Medical University

戴敏育

Min-Yuh Day

Assistant Professor

Dept. of Information Management, Tamkang University

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

2012-10-31



# Outline

1. Social Media (社群媒體)
2. Social Word-of-Mouth (社群口碑)
3. Web Mining (網路探勘)



# Social Media (社群媒體)



# #1 Activity on the Web?

# Social Media





Copyrighted Material

**Dave Evans** with Jake McKee

Foreword by Susan Bratton, CEO, Personal Life Media, Inc.

# Social Media Marketing

*The Next Generation of Business Engagement*



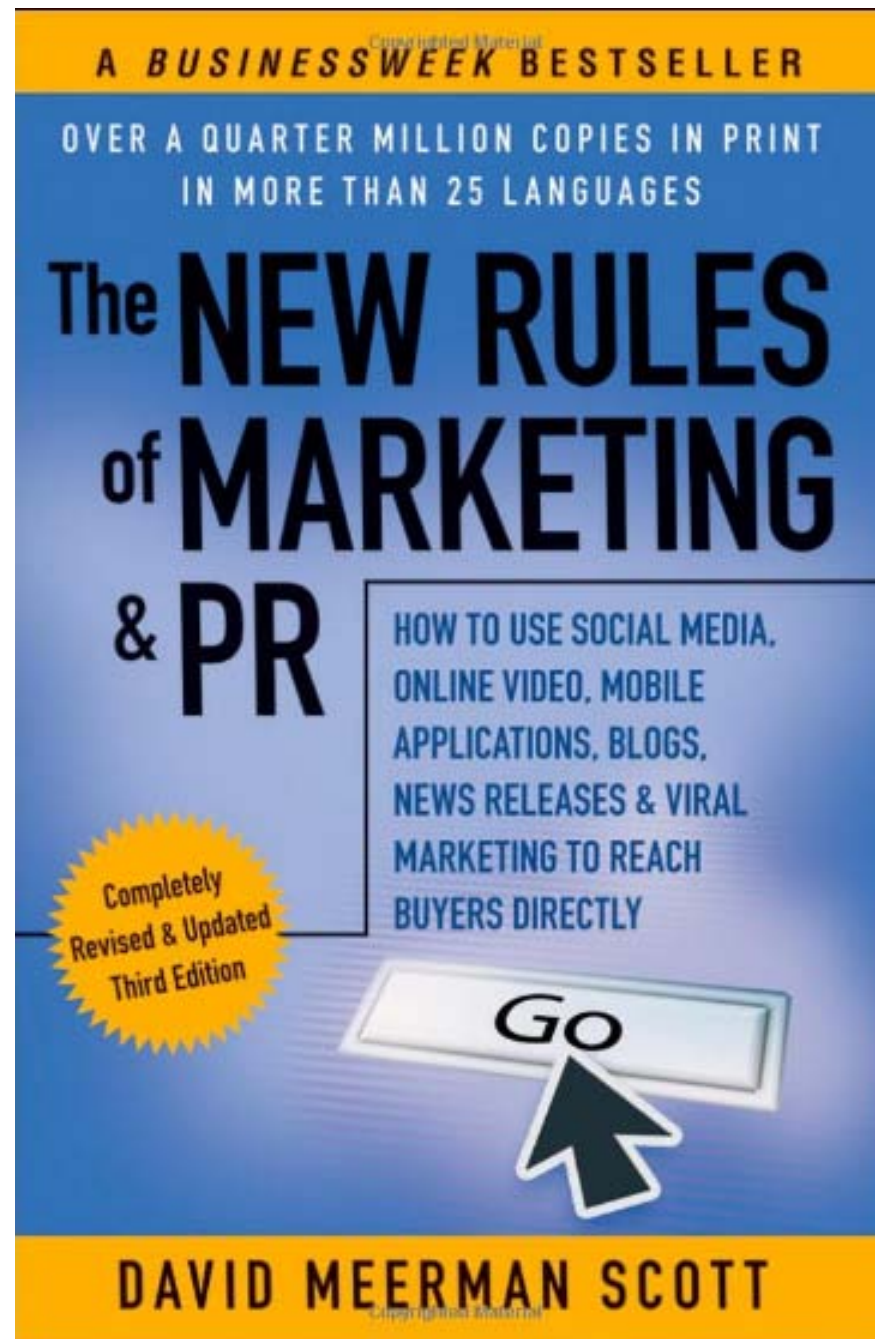


# Social Media

Strategies for Engaging in Facebook, Twitter & Other Social Media

# Marketing







Copyrighted Material

---

THE

---

# SOCIAL MEDIA MANAGEMENT HANDBOOK

---

EVERYTHING YOU NEED TO KNOW  
TO GET SOCIAL MEDIA WORKING  
IN YOUR BUSINESS



STRATEGY, CULTURE,  
METRICS, POLICIES,  
ROLES, AND  
RESPONSIBILITIES

NICK SMITH & ROBERT WOLLAN

WITH

CATHERINE ZHOU

Copyrighted Material





Copyrighted Material

**THE** | **SOCIAL**

**MEDIA**

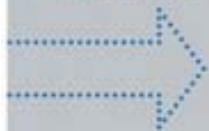


**BIBLE**

**LON  
SAFKO**

**SECOND  
EDITION**

**TACTICS, TOOLS & STRATEGIES  
FOR BUSINESS SUCCESS**



Copyrighted Material



Copyrighted Material

**#1**  
**BESTSELLER!**

*Completely Revised and Updated!*

# FACEBOOK MARKETING

Third Edition

Leveraging Facebook  
for Your Marketing Campaigns



JUSTIN LEVY

Copyrighted Material



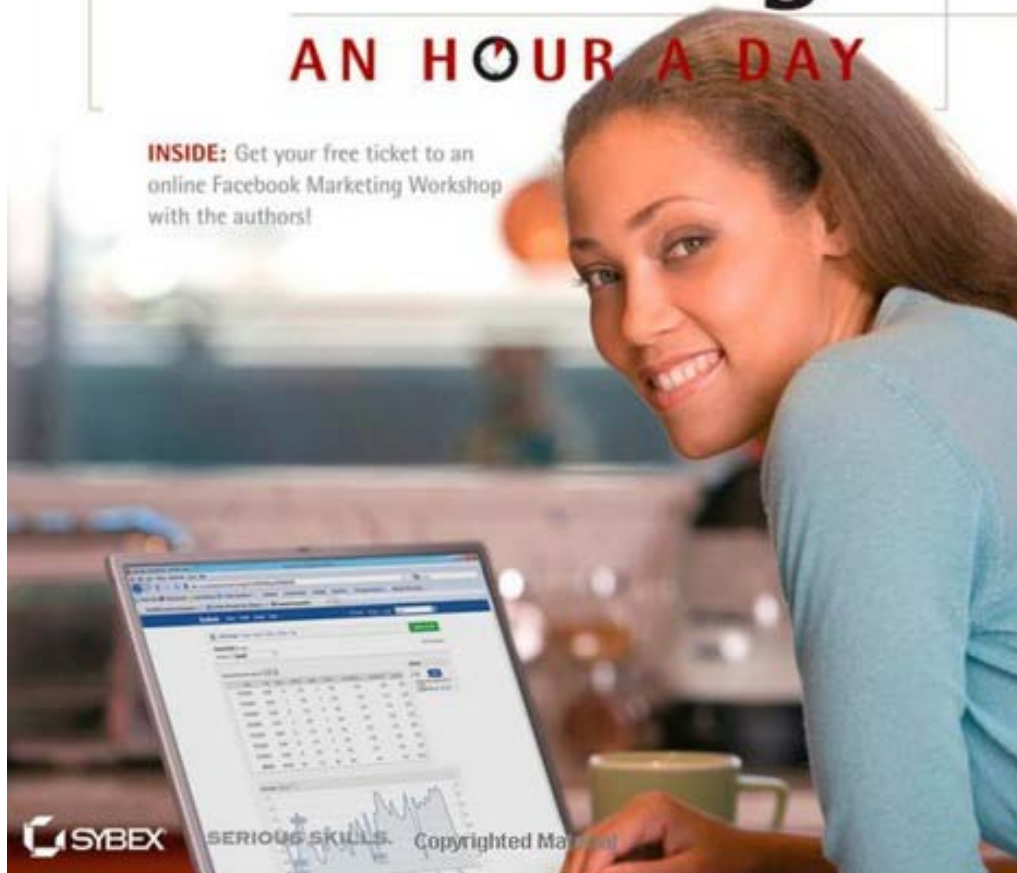
Copyrighted Material

Chris Treadaway and Mari Smith

# Facebook<sup>®</sup> Marketing

## AN HOUR A DAY

**INSIDE:** Get your free ticket to an online Facebook Marketing Workshop with the authors!





Michael Miller

**BESTSELLER!**

# YouTube®

Online Video Marketing for Any Business

# for Business

Second Edition

*This latest edition is a must-read book for any business owner wanting to implement a successful inbound video marketing campaign.*

—Rey Ybarra, Host/Producer of "The New Media Radio Hour"  
[www.newmediaradiohour.com](http://www.newmediaradiohour.com)



Copyrighted Material

# YouTube Marketing Handbook

Marc Bullard - YouTube Marketing Handbook

infomarketingexperts 1 book

Views

Marketing Efforts

0:00 / 5:19

Like Add to Share

443,731

1,324 likes, 0 dislikes

Updated by infomarketingexperts on Mar 25, 2011

Step by step instruction on how to use and market with just about every single feature of YouTube. Learn how to market, analyze, and research potential customers with FREE tools provided by YouTube.

Show more

by Marc Bullard

Video Marketing Specifically for YouTube

Copyrighted Material



**INSIDE:** Your Google AdWords™ gift card worth \$50—show your ads on YouTube!

Greg Jarboe

Foreword by Brian Cusack, Head of Display, Retail, and Canada, Google

# YouTube® and Video Marketing

**AN HOUR A DAY**

SECOND EDITION



# Social Media



# Management

Let Us Manage Your  
Social Media Campaigns



# Social Media Management Pyramid





# Social Media Marketing For Business





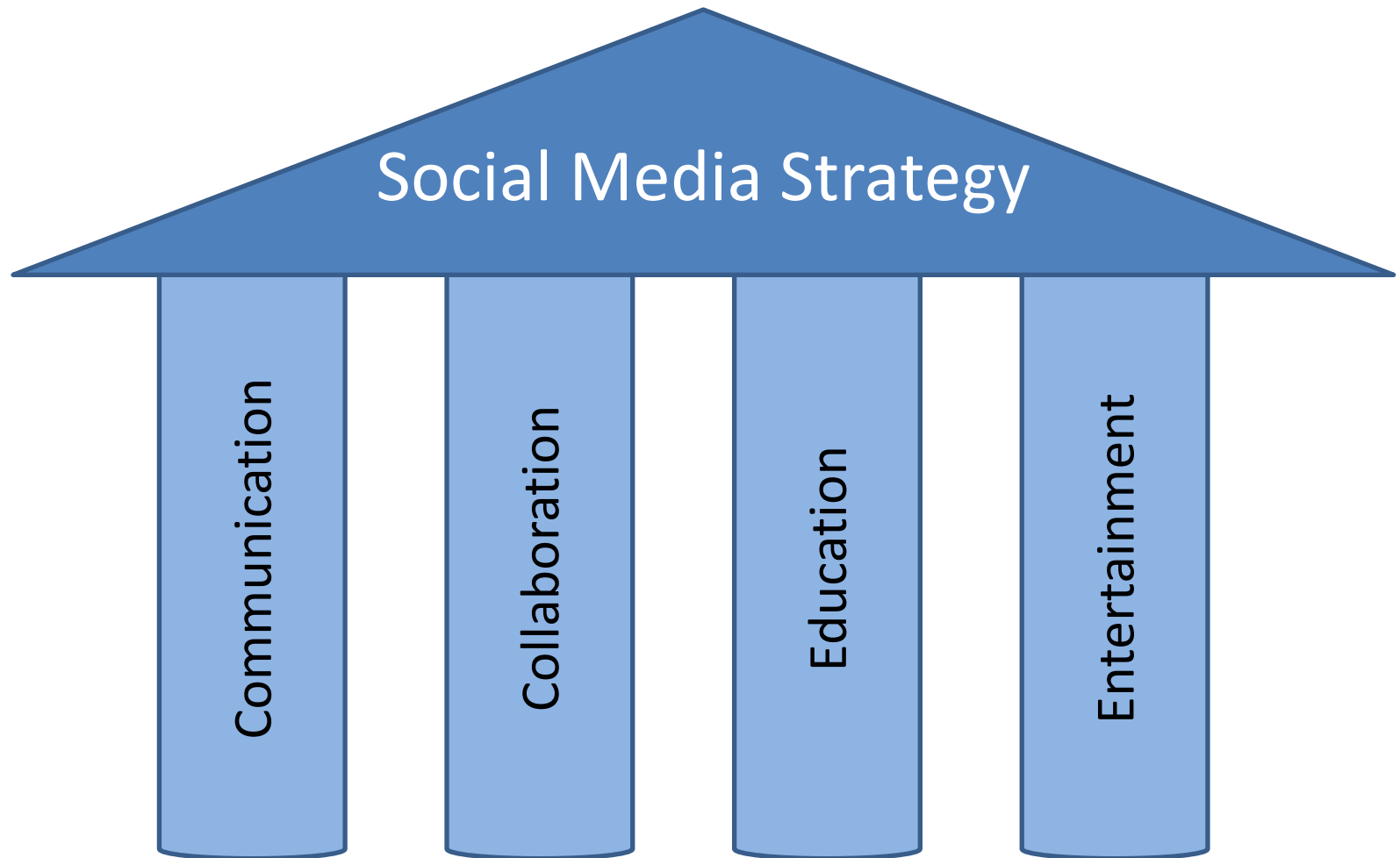
# Marketing 4P to 4C

- **Product** → **Customer solution**
- **Price** → **Customer Cost**
- **Place** → **Convenience**
- **Promotion** → **Communication**



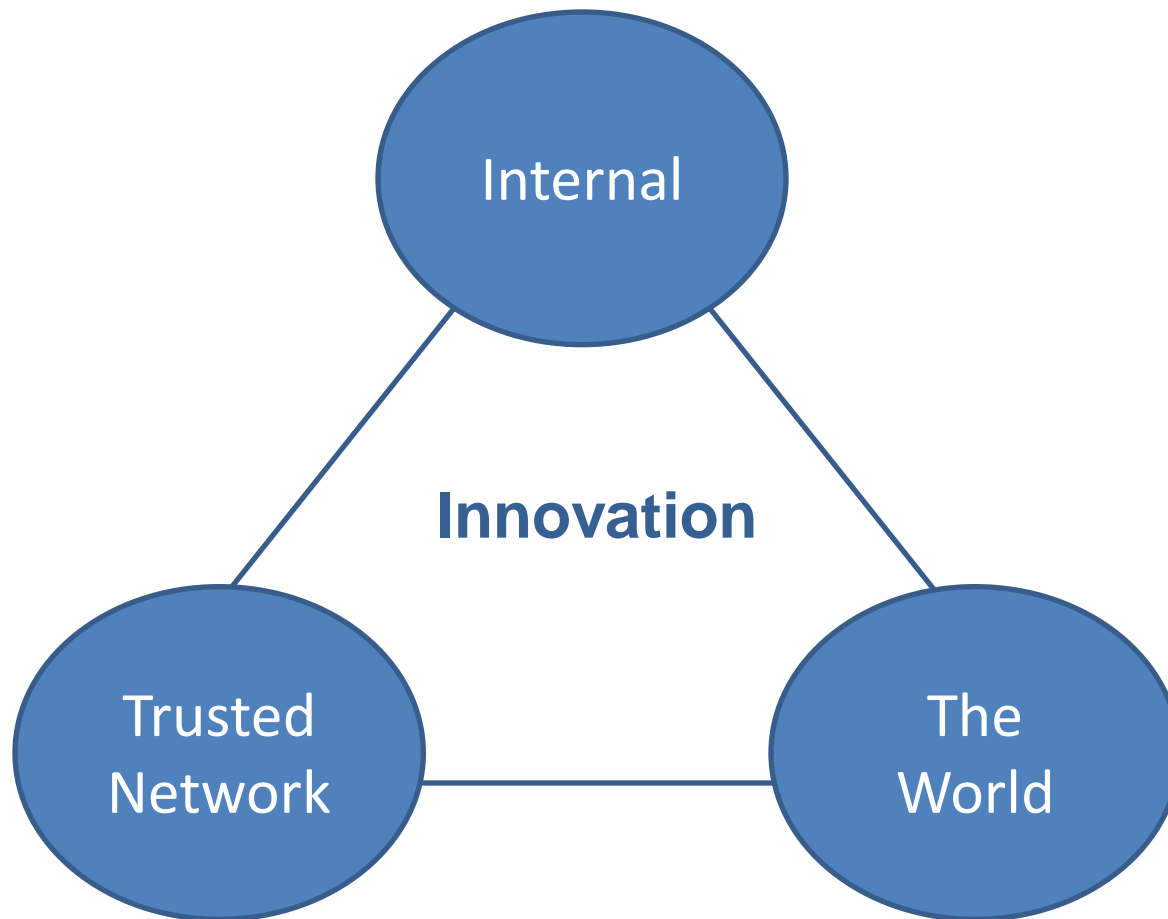
# Four Pillars of **Social Media Strategy**

**C<sup>2</sup>E<sup>2</sup>**





# 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





# 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





# THE WEB 2.0 REVOLUTION, **SOCIAL MEDIA**, AND INDUSTRY DISRUPTORS



# SOCIAL LUMAscape

MARKETER

CONSUMER

twitter  
facebook

**Social Marketing Management**  
 BUDDY#MEDIA shoutlet SYNCAPSE vitrue  
 Context Optional Spredfast involer sprinkl  
 hootsuite Social Publishing Platforms Flowtown  
 EXPION awareness tweet tapif  
 hear say MUTUALMIND SOCIALFLOW Socialware  
 socialvot SPROUT SOCIAL ArgyleSocial  
 offerpop Social Promotion Platforms Social Amp  
 5353factory Seismic Strutta votigo  
 SocialAppsHQ extole Fanzila  
 Zuberance Buzzgent CanSea

**URL Shorteners**  
 bitly  
 tiny arrows  
 TinyURL.com

**Stream Platforms**  
 UberMedia TweetDeck  
 twirl Aol Lifestream...

**Twitter Apps**  
 twitpic tweetmeme  
 StockTwits Cadmus  
 wefollow fixup!  
 Listorious twitvid

**Social Advertising Platforms**  
 TBG BLINK MEDIA spruceMEDIA Acaply Involved Media  
 Lexity epic social SOCIALTYZE  
 UNIFIED nanigans  
 Involved Media Ybrant  
 Marin SAM Taykey  
 Clickable KENSHOOD optimal EfficientFrontier

**Analytics**  
 AWB.SM bluefin  
 kontagent  
 Campalyut  
 mixpanel DMNITURE  
 webtrends  
 Simply Measured

**Content Curation**  
 hunch= memolane  
 BuzzFeed summarify  
 Ripbeat SkyGrid Storify

**Facebook Apps**  
 LIKESTER BranchOut  
 SNAP badoo smile  
 causes booshaka

**Social Ad Networks**  
 OneRiot rockyou across LIFESTREET  
 radium ONE XGRAPH  
 media6degrees

**Social Brand Engagement**  
 socialvibe mylikes  
 adly local response  
 appsavvy  
 DYNAMIC SIGNAL  
 sharethrough

**Social Commerce Platforms**  
 Payment Moontoast SHOP TAB  
 live gamer dotbox  
 Storenvy Mollie triapay  
 FLUID shop igniter  
 VendorShop miiyoni

**Facebook Gaming**  
 playfish zynga MetroGames  
 Playdom zuma CrossStar  
 socialpoint ZIPZAPPLAY  
 EA UBI SOFT EVERETT WILDNEEDLEGAMES

**Social Intelligence**  
 PostRank Trendrr ATTENUITY  
 synthesio crimson hexagon actionly  
 BuzzStream bottlen se colligent  
 trackur INFLUENCE PROFILES  
 systemx intlogy brandprotect  
 BACKOUTLABS M backtype  
 ALTERIAN conversion NETBASE

**Social Scoring**  
 KLOUT  
 empireavenue  
 PeerIndex  
 Kred

**Social Data**  
 Gnip DATA SIFT Topsy RapLeaf

**Social Search & Browsing**  
 Topsy Wink Aardvark  
 StumbleUpon greplin spokeo

**Social Business Software**  
 pasenger lithium JIVE telligent news gator  
 External (Customer) Facing  
 Ingage satisfaction mzinga  
 ISI Assistly TickApps Leverage Software Pluck  
 Internal (Employee) Facing  
 huddle cubetree acqua igloo  
 Watchitoo Yammer  
 moxie Socialtext sknoble SOCIALCAST

**Blogging Platforms**  
 tumblr  
 posterous  
 Blogger  
 WordPress  
 SQUARESPACE  
 Joomla!  
 JUX

**Social TV**  
 tunerfish Snappi PHILLO  
 GetGlue INTO-NOW miso FLINGO  
 UM&M clipsync

**Social Networks - Other**  
 Linked in TAGGED my  
 plawo Path CLUB PERIPHER  
 mynearby orkut Google+ #hashable renren  
 friend.ly

**Social/Mobile Apps & Games**  
 ingmoco  
 WAZZE IWIIX  
 Foodspotting glu h5  
 pelago playdom

**Social Shopping**  
 Zappi zappli  
 Swipely LOCKERZ  
 SUPPLY kaboodle

**Content Sharing (Reviews/Q&A/Docs)**  
 topix Pinterest fotopedia Scribd  
 yelp DocShare everplaces  
 Angie's list Quora

**Community Platforms**  
 yuku GROU.PS TickApps  
 BuddyPress NING!  
 mixxt Groupify

**Social Referral**  
 socialfoot  
 500friends TurnTo curebit

**Photo/Video Sharing**  
 SmugMug ZangZing photobucket olapic  
 Instagram fixable flickr Picasa iQ klip

**Plug-ins/Widgets**  
 tunc Bazaarvoice gigya Badgeville  
 ShareThis spinback janrain conduit  
 Add This meebp PowerReviews  
 Mass Relevance 5353factory

**Traditional Publishers**  
 DAILY CANDY CNN The New York Times  
 Glam Media The Wall Street Journal  
 ABC HEARST corporation  
 AOL. CONDÉ NAST

Denotes acquired company

© LUMA Partners LLC 2012



# 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



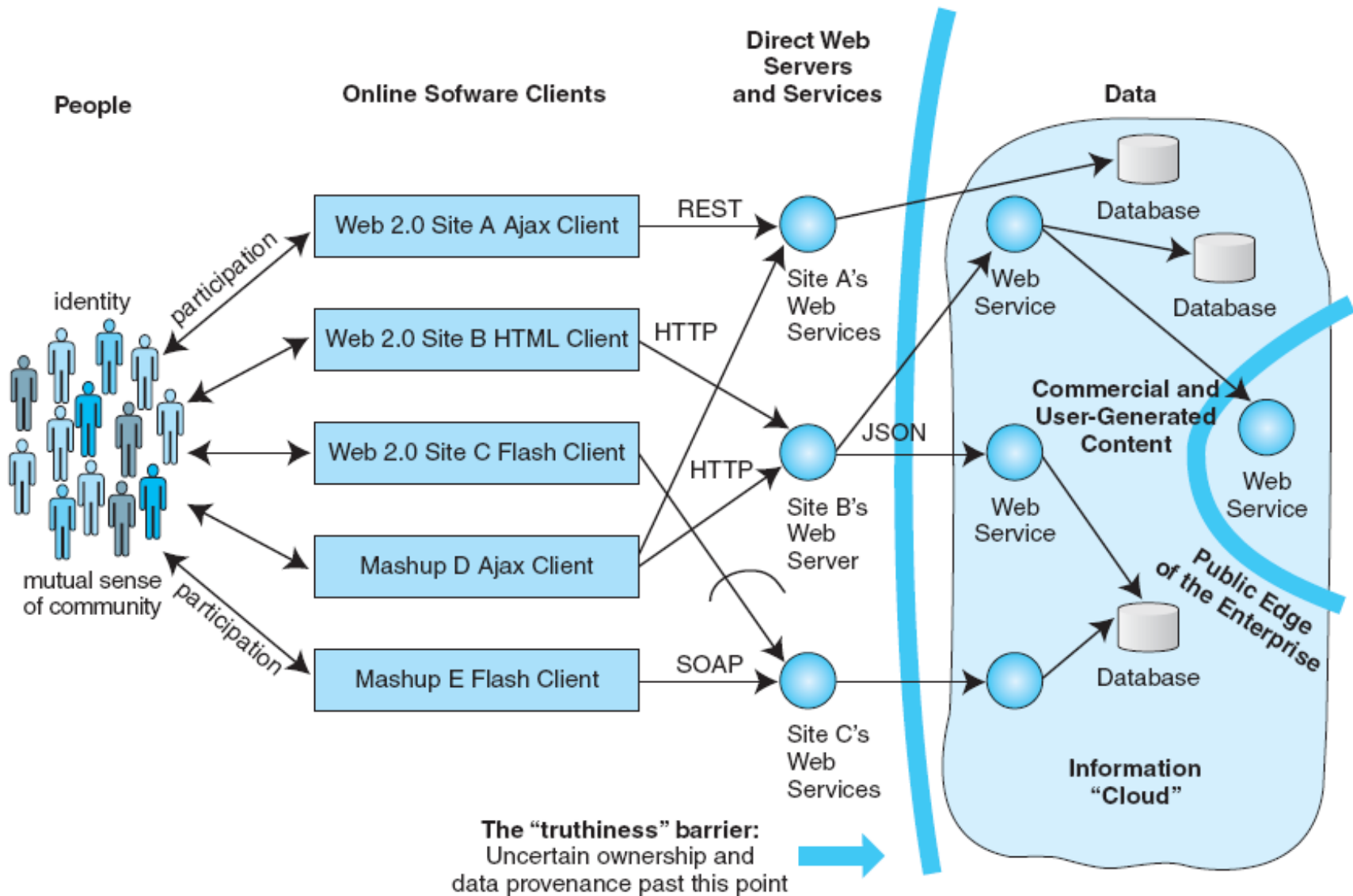
# THE WEB 2.0 REVOLUTION, SOCIAL MEDIA, AND INDUSTRY DISRUPTORS

- **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

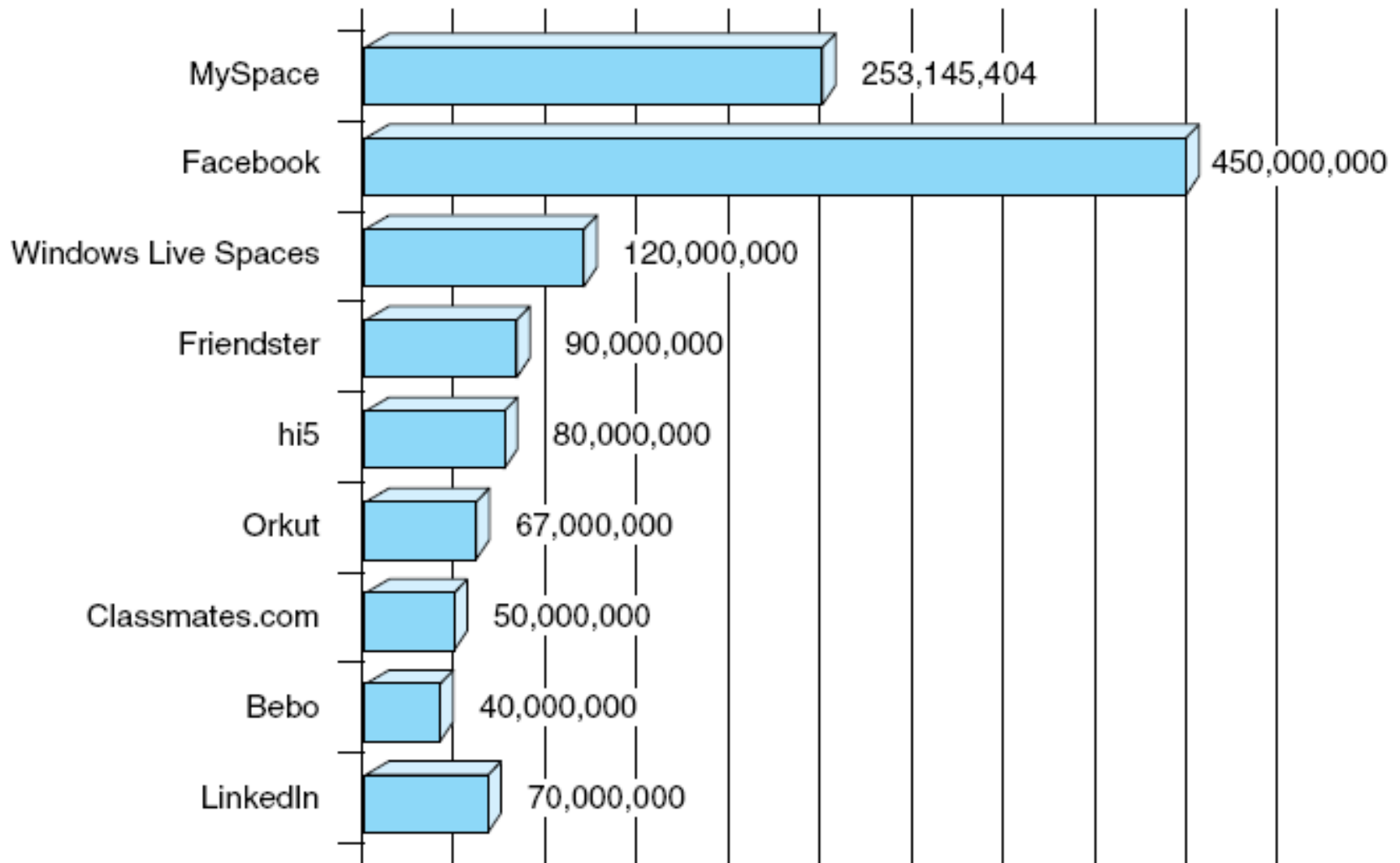
- **social networking**

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





# 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

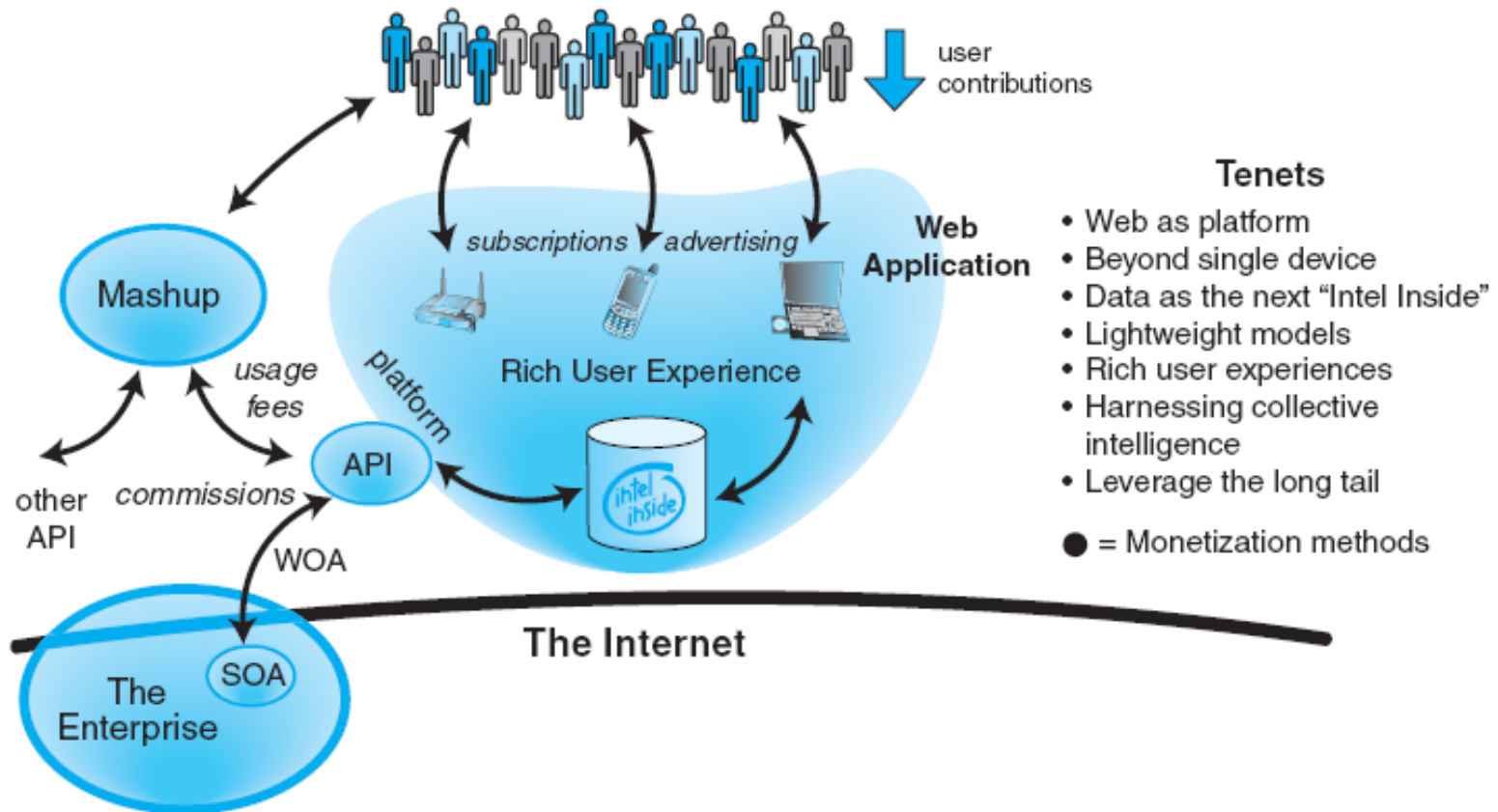
- **SHOPPING IN SOCIAL NETWORKS**
- **FEEDBACK FROM CUSTOMERS:  
CONVERSATIONAL MARKETING**
  - **Customer Feedback with Twitter**



# **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





# **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





# Social Word-of-Mouth (社群口碑)



# **Social Media Word-of-Mouth Marketing**





# How to Start Buzz

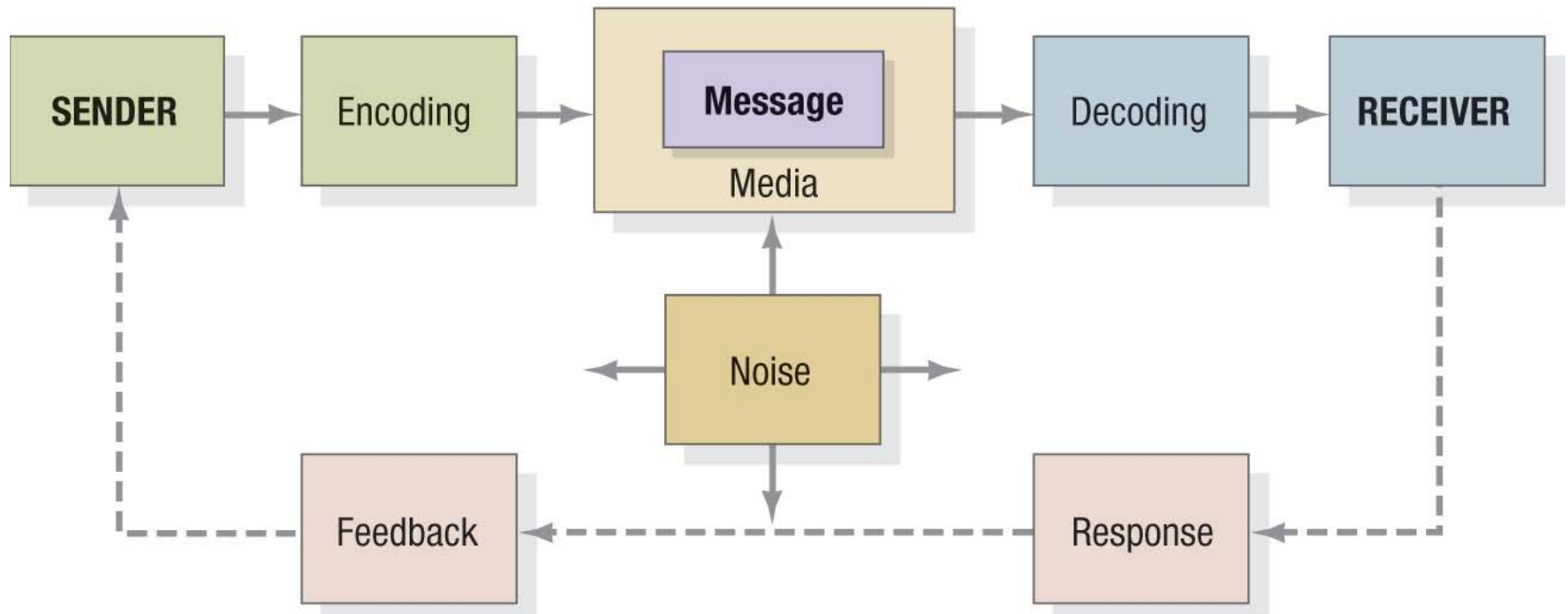
- Identify influential individuals and companies and devote extra effort to them
- Supply key people with product samples
- Work through community influentials
- Develop **word-of-mouth** referral channels to build business
- Provide compelling information that customers want to pass along



# Word-of-Mouth Marketing

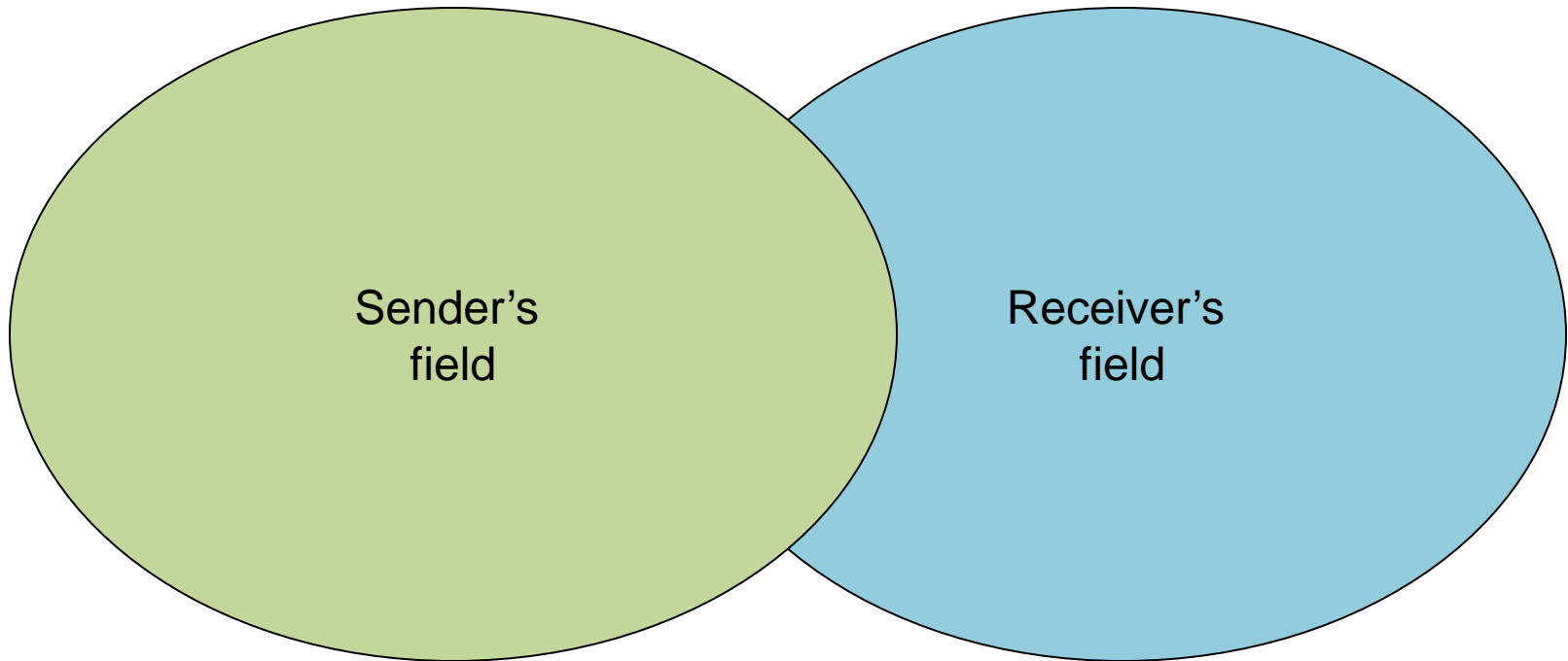
- Person-to-person
- Chat rooms
- Blogs
- Twitter, Plurk
- Facebook
- Youtube

# Elements in the Communications Process





# Field of Experience



# 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.



# 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



# Listen and Learn

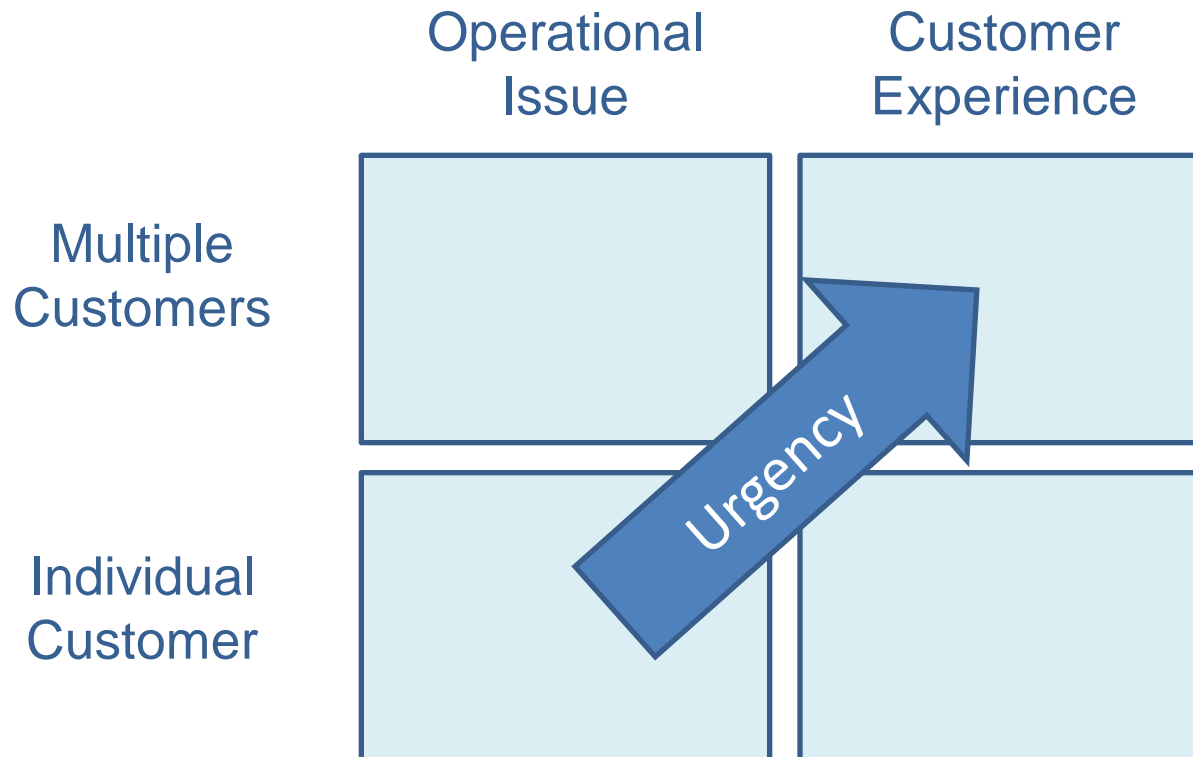
## Text Mining for VoC

- 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.



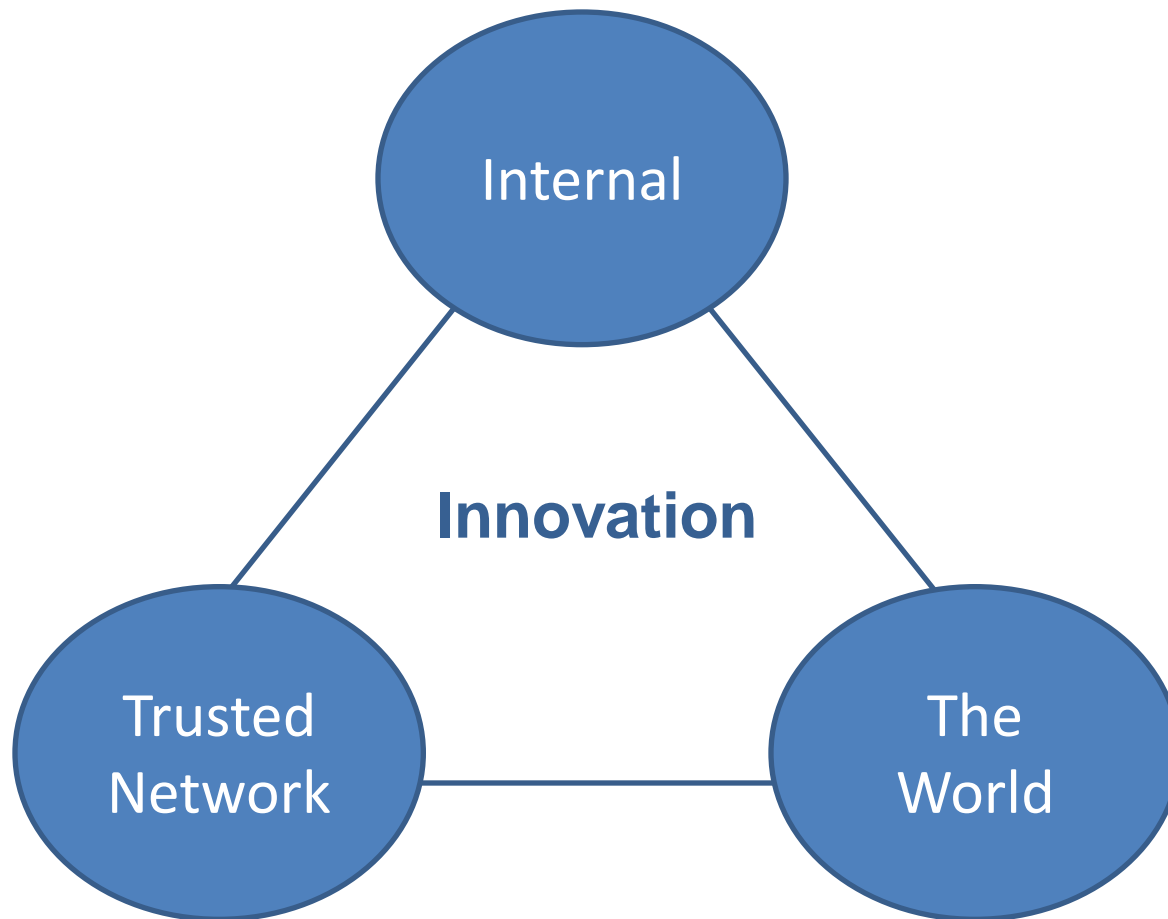
# Customers' Opinions About Operational versus Customer Experience Issues

## Reactive, Reputation Management





# 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

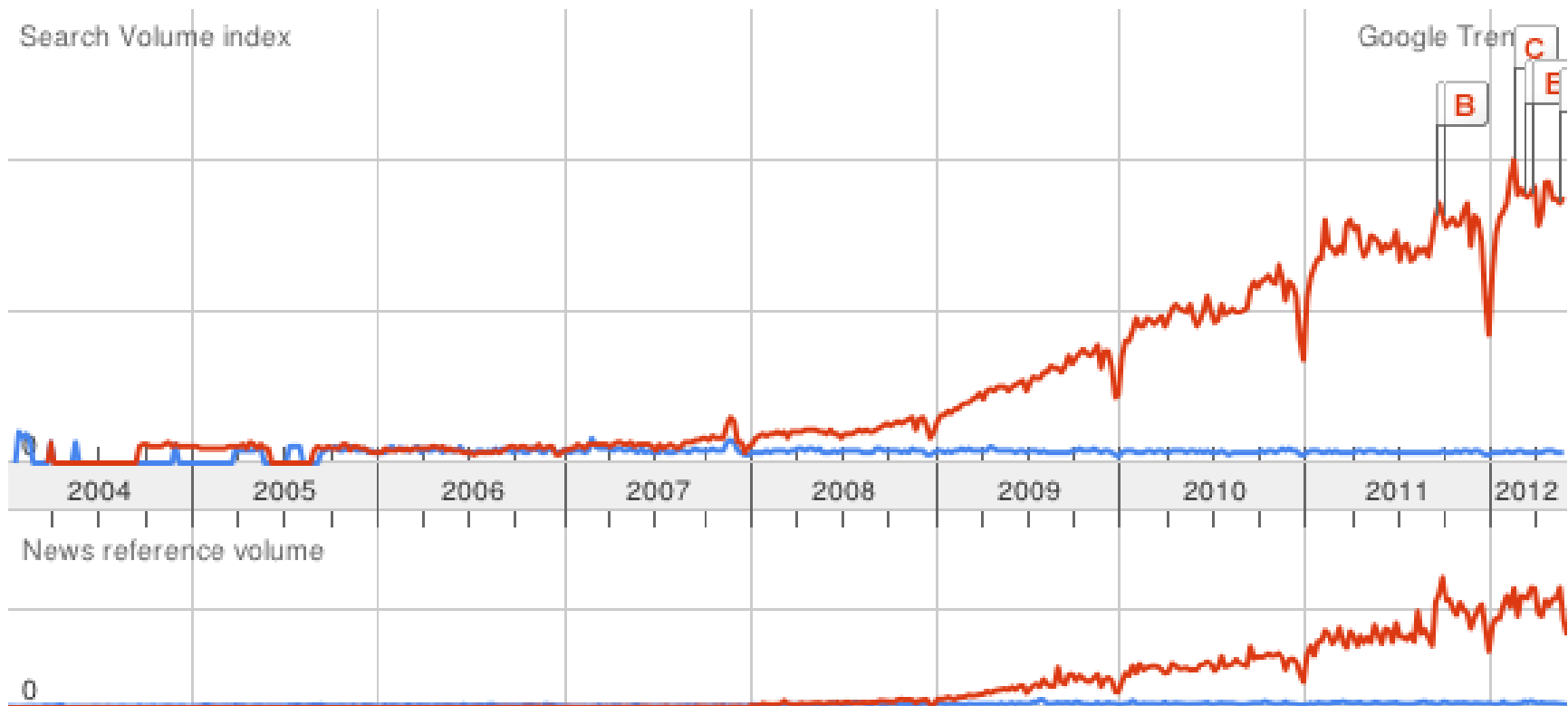




# Word-of-Mouth

## Social Media

word of mouth 1.00    social media 7.40



# Case Study: LenovoClub CareerLife 職場人生



拉NO佛國際集團 民國一百年隆重鉅獻

CH4 LTV 戲劇 | 頻道 | 活動 | 討論 | 購物

副總爭霸戰 觀眾選邊讚

特別介紹 姓沛的

人物關係圖

劇照 下載

## 職場人生

職場人生親似海深 誰擁姓沛的就穩升  
惡男奸女爭權謀，纏鬥三回定生死 非死不可選邊讚，若贏姓沛的跟你走！



# Case Study: LenovoClub CareerLife 職場人生



拉NO佛國際集團鉅獻... [www.lenovoclub.com.tw/careerlife/](http://www.lenovoclub.com.tw/careerlife/)

第一集 按怎潑攏免驚 熱烈上映中

[職場人生] 第一集 不管安怎潑攏免驚



0:00 / 2:02

拉NO佛國際集團空出副總職缺，  
平日是死對頭的行銷部經理麥可陳，和業務部經理尖妮ㄟㄟ，  
為求升官互相惡鬥，  
嗆聲過程中有人順勢把手中一杯水潑出……

浪浪 PLURK  
非死不可 facebook 分享

©2011 LTV CORPORATION AND LANGFO INTERNATIONAL CORPORATION CHA LTV

# Case Study: LenovoClub CareerLife 職場人生



YouTube player interface showing the video "【職場人生】第一集 不管安怎潑攏免驚". The video title is "【職場人生】第一集 不管安怎潑攏免驚". The channel is "Itvchn" with 13 videos. The video is at 0:03 / 2:02. The video content shows a night cityscape with the title "職場人生" in large red characters and "拉NO佛跨國企業集團鉅獻" in purple characters. The video has 124,781 views and 203 likes.

Video title: 【職場人生】第一集 不管安怎潑攏免驚

Channel: Itvchn (13 部影片)

Video description: 拉NO佛跨國企業集團鉅獻

Video duration: 0:03 / 2:02

Views: 124,781

Likes: 203 (Dislikes: 14)

Channel description: Itvchn 於 2011-05-06 上傳  
拉NO佛國際集團空出副總職缺，平日是死對頭的行銷部經理麥可陳和業務部經理尖妮口舌互嗆.....

Recommended videos:

- 擔心聖誕夜訂不到餐廳? (0:16)
- 【職場人生】第二集 姓沛的係按呢用ㄟ (2:31)
- 【職場人生】第三集 副總爭奪戰最終回 (3:23)
- [TVBS 報導]「筆電人生」鄉土劇 啥！廣告仿 (2:12)
- 【職場人生】第ㄟ篇 (0:18)
- 【職場人生】精彩預告(短) (0:11)

# Case Study: LenovoClub CareerLife 職場人生



[職場人生] 第一集 不管怎麼

www.youtube.com/watch?v=XRUVbFEnPig

124,781

總觀看次數：124,781

11-05-05 11-08-25 11-12-15

評分：217 評論：63 我的最愛：119

喜歡的：203 不喜歡的：14

### 重要探索活動

日期	活動	觀看次數
A	11-05-12 精選影片初次觀看次數	7,250
B	11-05-12 初次嵌入於 - static.ak.facebook.com	6,360
C	11-05-11 初次嵌入於 - share.youthwant.com.tw	2,946
D	11-05-10 首次透過廣告播放	7,727
E	11-05-10 初次嵌入於 - www.mobile01.com	2,820
F	11-05-08 初次嵌入於 - www.lenovoclub.com.tw	14,749
G	11-05-08 來自 YouTube 搜尋的首次推薦 - 職場人生	9,522
H	11-05-08 首次透過行動裝置播放	4,700
I	11-05-08 首次推薦媒介： - www.facebook.com	4,563
J	11-05-06 來自相關影片的首次推薦 - [職場人生] 第二集 姓沛的係按呢用	8,749

### 觀眾

此影片最受下列人士歡迎：

性別 年齡

此影片在以下地點最受歡迎：

- [職場人生] 精彩預告(短) by Itvchn 觀看次數：6,944
- 世界最小台筆電 by cbate 觀看次數：27,623
- 頭夜市D第一集片段-江一官向警察挑釁(附字幕) by lan08438 觀看次數：31,780
- 超級偶像-崩崩哥 by wenhot 觀看次數：539,566
- 2010-03-31 夜市人生72集有志跆拳道第一名 大風如 by a1245648 觀看次數：19,463
- 職場人生-孤島教師下集.mp4 by PhoneMovieTv 觀看次數：3,862
- 夜市人生 爆橋拳 (格鬥天王版) by st10043 觀看次數：489,652
- 職場人生-辦桌總舖師上

# Case Study: LenovoClub CareerLife 職場人生



[職場人生] 第一集 不管安怎 x

www.youtube.com/watch?v=XRUVbFEnPig

**觀眾**  
此影片最受下列人士歡迎：

性別	年齡
男性	35-44
男性	25-34
男性	18-24

此影片在以下地點最受歡迎：

更多  
較少

Itvchn 於 2011-05-06 上傳

拉NO佛國際集團空出副總裁，平日是死對頭的行銷部經理麥可陳和業務部經理尖妮ㄟㄟ互哈.....  
[http://www.lenovoclub.com.tw/careerlife/?utm\\_source=Youtube&utm\\_medium=vi...](http://www.lenovoclub.com.tw/careerlife/?utm_source=Youtube&utm_medium=vi...)

顯示更多

最佳評論

不簡單的宣傳片段!! 裡面完全沒帶半點廣告氣色, 亦完全沒有感到是在播廣告! 而且他完全不會想在電視來放!! 因為在電視大家就會以為是在看廣告!! 但在電腦世界大家就會以為是真的什麼新的連續劇!! 成功的廣告, 不會在廣告裡說那是什麼產品!! 兩個字已經完全地帶出廣告意思!!! 不簡單的廣告!! 不簡單的姓沛和拉NO佛  
nomuchmore 7個月以前 14

"我就跟你姓尖" 笑死XDDD  
p90349 6個月以前 10

203人喜歡, 14人不喜歡

影片出處:  
OA拆組達人-OA辦公家具  
買賣拆組專家

- by st10043  
觀看次數: 489,652
- 職場人生-辦桌總舖師 上集.mp4  
by PhoneMovieTV  
觀看次數: 1,001
- Nothing New No Response  
by vutuanduy  
觀看次數: 159,593
- 職場人生-木船師 上集.mp4  
by PhoneMovieTV  
觀看次數: 734
- 職場人生 木船師 下集  
by PhoneMovieTV  
觀看次數: 345
- 職場人生-辦桌總部師 下集.mp4  
by PhoneMovieTV  
觀看次數: 553
- 老婆的夜市人生~第一集~射飛鏢  
by jasir120  
觀看次數: 135
- 信義房屋分手快樂 真相篇  
by jamestuo  
觀看次數: 542,319



# Web Mining

## (網路探勘)

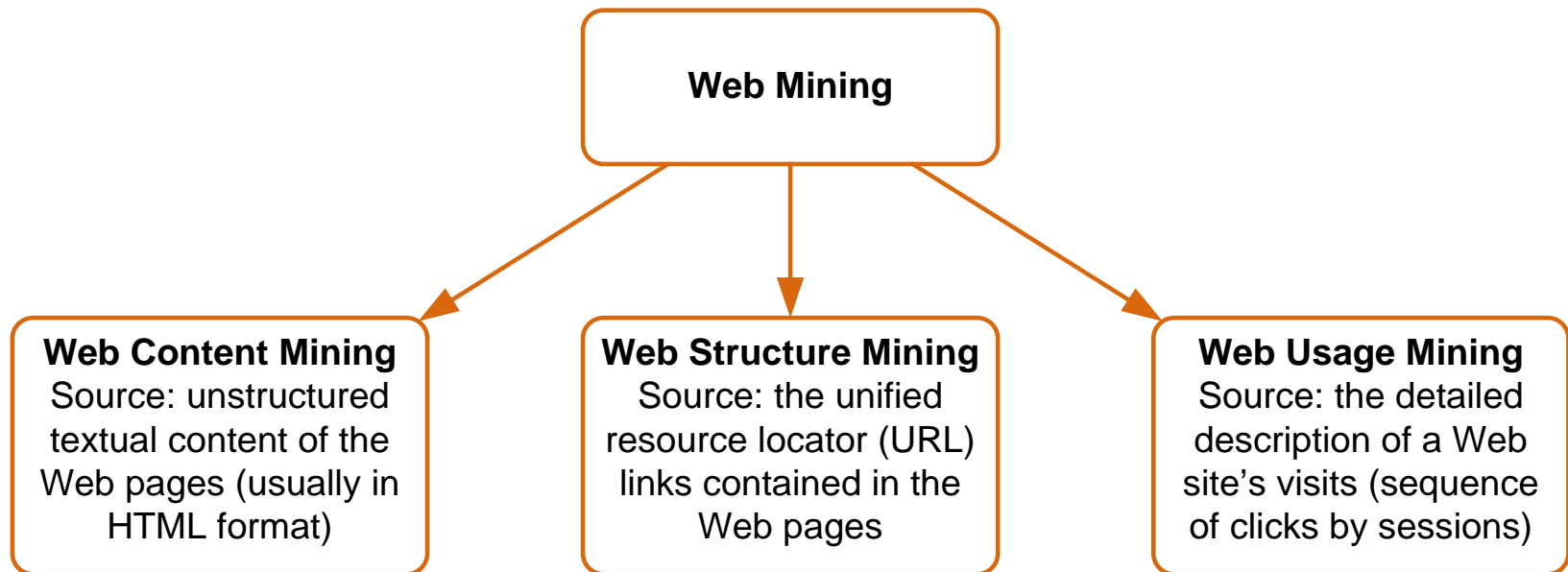


# ACM Categories and Subject Descriptors

- I.2.7 [Artificial Intelligence]
  - Natural Language Processing
    - Text analysis
- H.2.8 [Database Management]
  - Database Applications
    - Data mining

# Web Mining

- Web mining (or Web data mining) is the process of discovering intrinsic relationships from Web data (textual, linkage, or usage)





# Web Content/Structure Mining

- Mining of the textual content on the Web
- Data collection via Web crawlers
- Web pages include hyperlinks
  - Authoritative pages
  - Hubs
  - hyperlink-induced topic search (HITS) alg





# Web Usage Mining

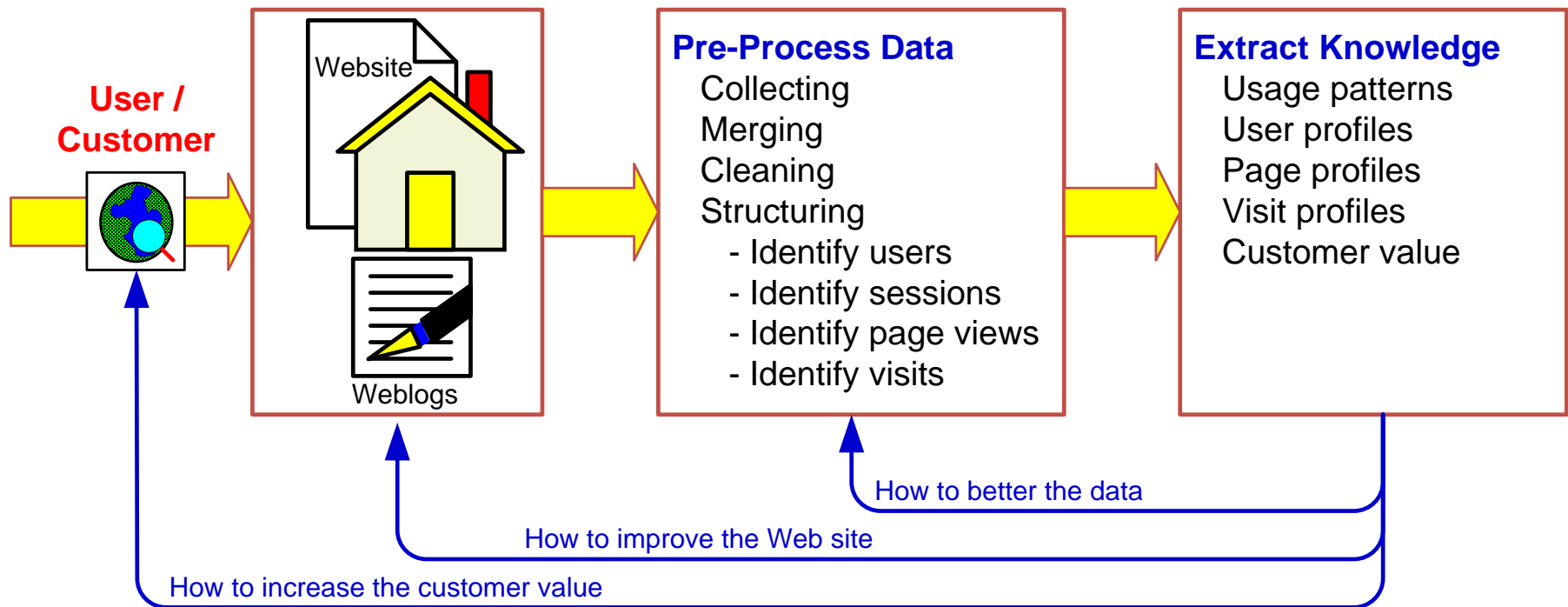
- Extraction of information from data generated through Web page visits and transactions...
  - data stored in server access logs, referrer logs, agent logs, and client-side cookies
  - user characteristics and usage profiles
  - metadata, such as page attributes, content attributes, and usage data
- Clickstream data
- Clickstream analysis



# Web Usage Mining

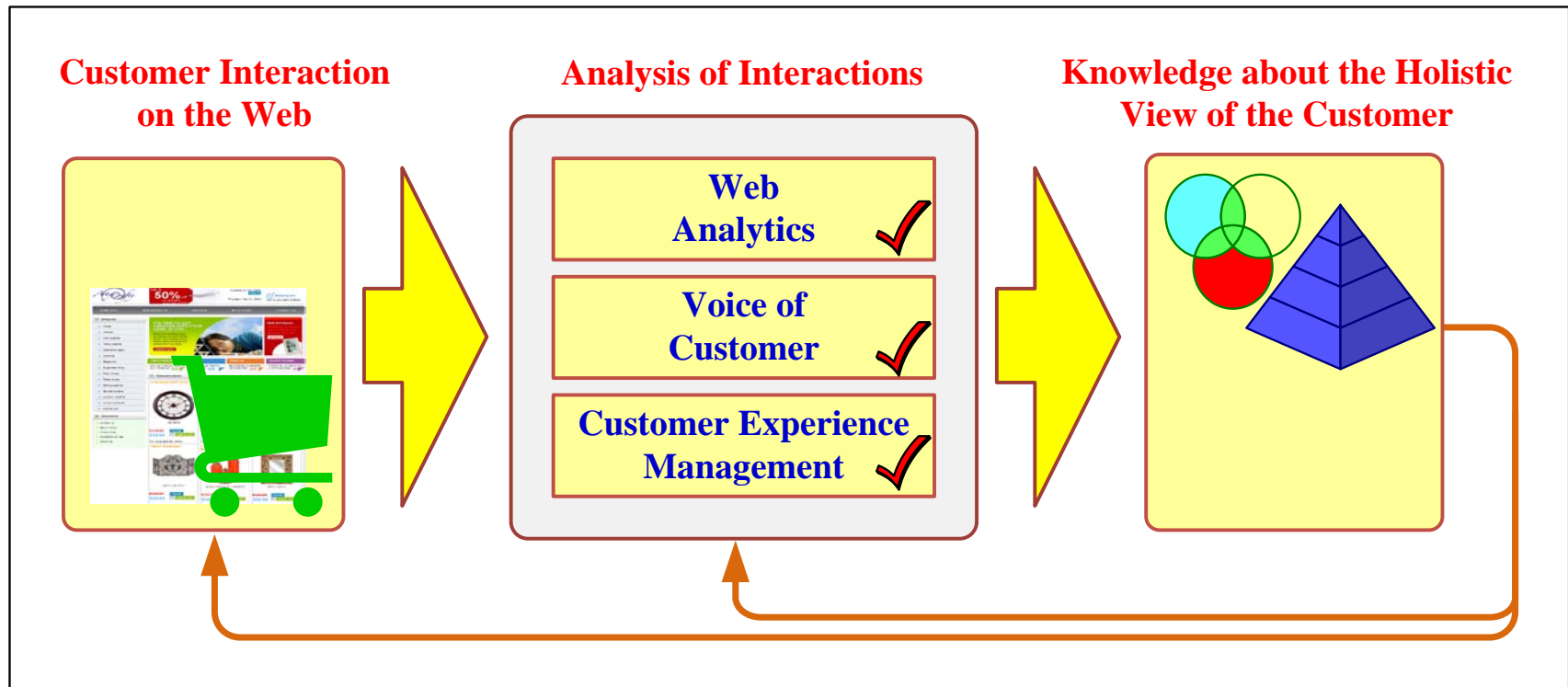
- Web usage mining applications
  - Determine the lifetime value of clients
  - Design cross-marketing strategies across products.
  - Evaluate promotional campaigns
  - Target electronic ads and coupons at user groups based on user access patterns
  - Predict user behavior based on previously learned rules and users' profiles
  - Present dynamic information to users based on their interests and profiles...

# Web Usage Mining (clickstream analysis)



# Web Mining Success Stories

- Amazon.com, Ask.com, Scholastic.com, ...
- Website Optimization Ecosystem

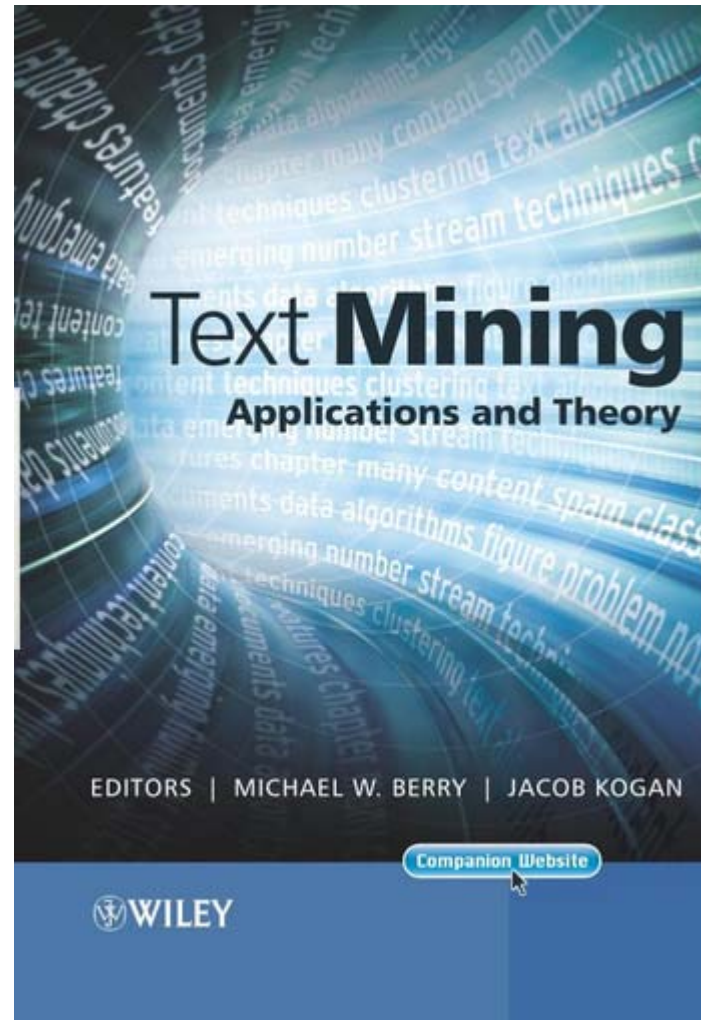




# Text and Web Mining

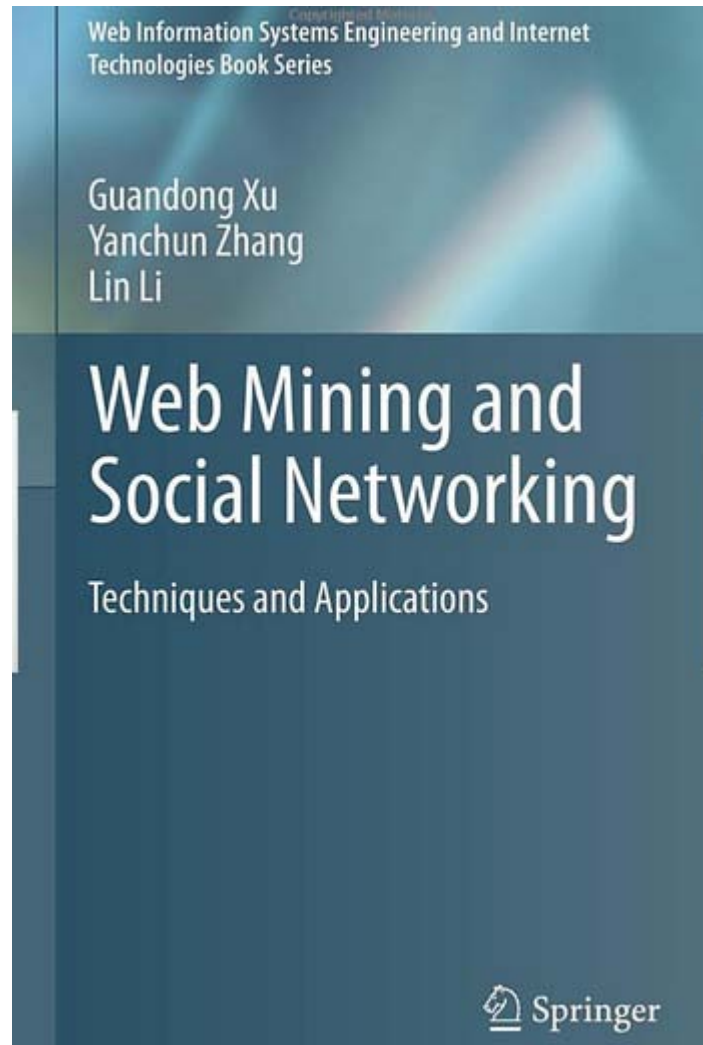
- Text Mining: Applications and Theory
- Web Mining and Social Networking
- Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites
- Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data
- Search Engines – Information Retrieval in Practice

# Text Mining





# Web Mining and Social Networking





# 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*



Mining the  
Social Web

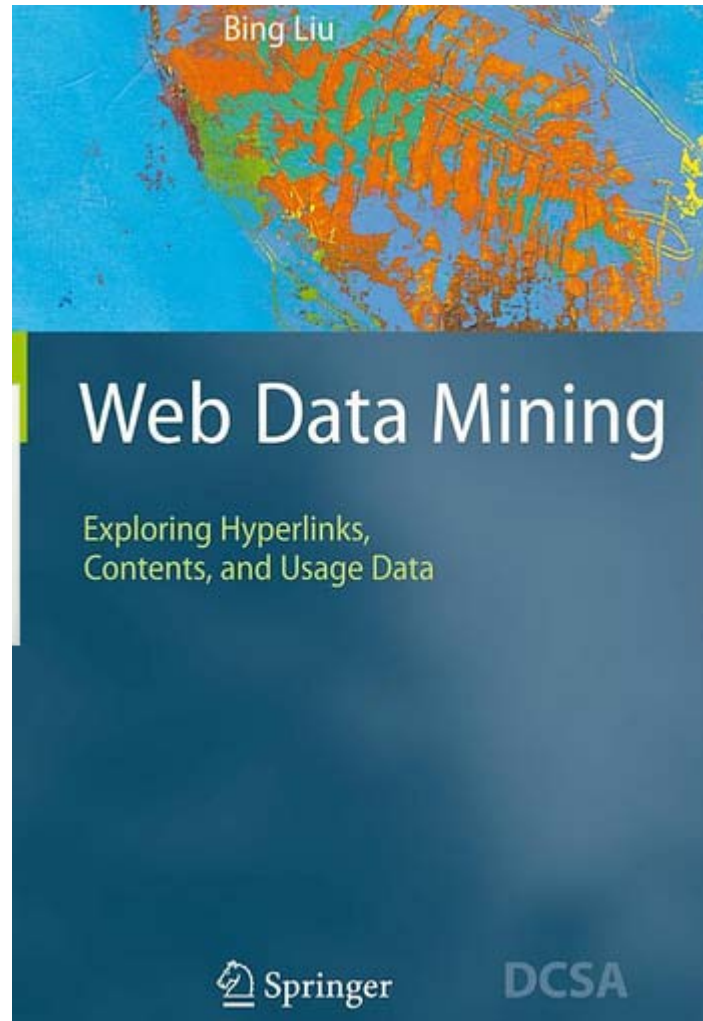
O'REILLY®

*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



# 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, 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





# 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/>

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. Analyze.

Watch Video

Command Center Video

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

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



# Clarabridge: Sentiment and Text Analytics Software

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

Sentiment and Text Analy x

www.clarabridge.com

Home About Us News & Events Blog Login Contact Us

clarabridge®

WHY TEXT ANALYTICS PRODUCT SERVICES CUSTOMERS PARTNERS RESOURCES ABOUT US

## The First Sentiment and Text Analytics Solution Built Specifically for Business.

The Clarabridge sentiment and text analytics software provides enterprises with a universal view of their customers.

Learn more about how Clarabridge works >

Customers

NISSAN BEST BUY Marriott sage H&R BLOCK CHOICE HOTELS INTERNATIONAL Wendy's GWLORD HOTELS BE GROUPSPACE DOLL More >

Clarabridge Text Analytics

Choose Your Edition

Clarabridge for Enterprises  
Ideal for companies seeking an enterprise class text analytics solution.

Clarabridge Webinar  
Hypatia Research Group presents on Social

<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

f

t

YouTube

g+

radian6 Community

Enter email address GO

**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.

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

[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/

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.

#### 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

“ The great thing about SAS is that it's so powerful and has such a broad offering. ”

—Jonathan Prantner  
Manager of Statistics  
Organic

Read full story

#### Product Demo



What do tweeples think ab x

www.tweetfeel.com/index.php#iPhone4s

FAQ | Contact Us

# tweetfeel

iPhone4s

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

Search:    
 User  Topic

Feedback

**Sentiments**

Index[0-100]: 50.00

Positive: 8

Negative: 1

Neutral: 13

Total Analyzed: 0

**Profile**


Education:

Flamboyance:

Slang:

Gender:

Age:



1 minute ago by Flow\_\_Show

1 minute ago by designer\_sayaka

1 minute ago by abhay01007

1 minute ago by SuckMy\_TwitNuts

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](#)

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

-議題分析-

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

連結類型	百分比
照片連結	73%
文章連結	10%
影片連結	17%
純發文	0%

#### 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
<b>馬英九總統</b>  粉絲數 1419591	<b>蔡英文 Tsai</b>  粉絲數 669928	<b>蘇貞昌</b>  粉絲數 317035

#### 排行榜

##### Traffic 交通

	<b>第1名</b> FORD (FIESTA)
	第2名 PORSCHE (911)
	第3名 SUBARU (legacy)

[更多](#)

##### Traffic 交通

	<b>第1名</b>
--	------------

#### 進口車

- 吵了這麼久, 就是不要「超速」超車, 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」超車, 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





The screenshot shows a web browser window with the URL [www.eland.com.tw/solutions](http://www.eland.com.tw/solutions). The page features the eLAND Cloud Services logo and a navigation menu with items: 產品, 雲端服務, 臺銀標專區, 我們的客戶, 新聞與活動, 支援, and 關於意藍. The main content area is titled "OpView Service 您的輿論觀測站" and includes a sub-header "連上OpView · 品牌形象 · 輿論觀點監測好輕鬆". Below this are five circular icons representing social media and analytics. A large image shows a dashboard with a bar chart and a list of data points. A search bar is located at the bottom right of the main content area. The footer contains a "相關文件下載" section with links to "OpView Insight DM", "OpView品牌口碑雷達 DM", and "OpView品牌口碑雷達訂購單".



# 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**

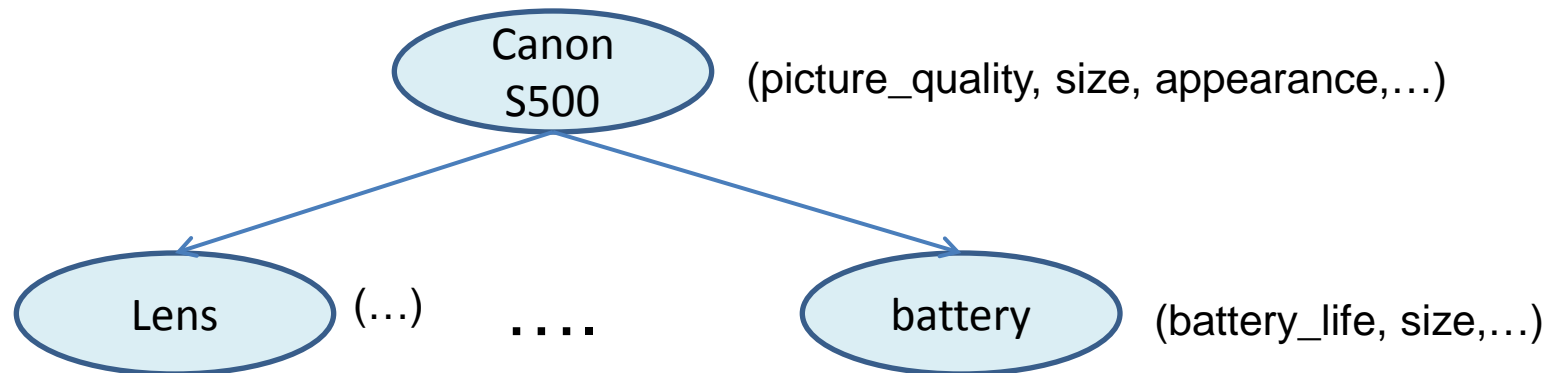


# Entity and aspect

- 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**

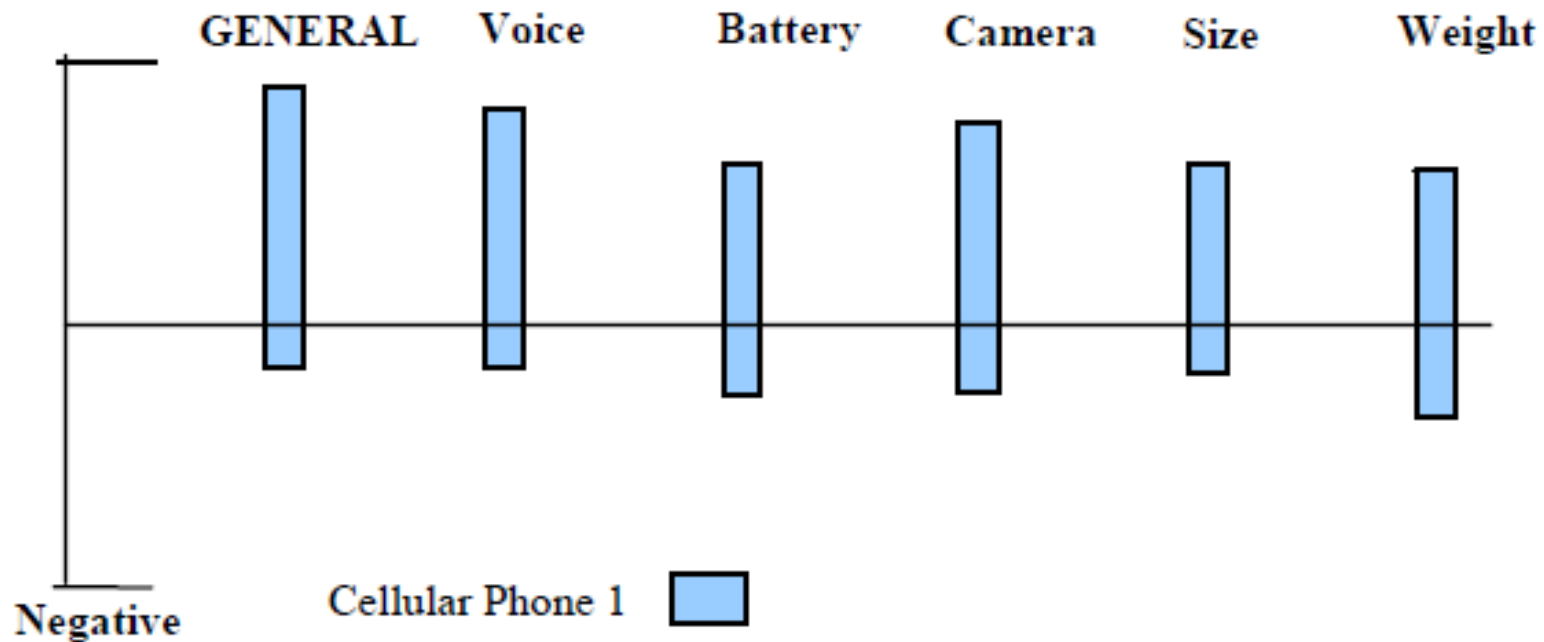
Positive: 80 <individual review sentences>

Negative: 12 <individual review sentences>

...

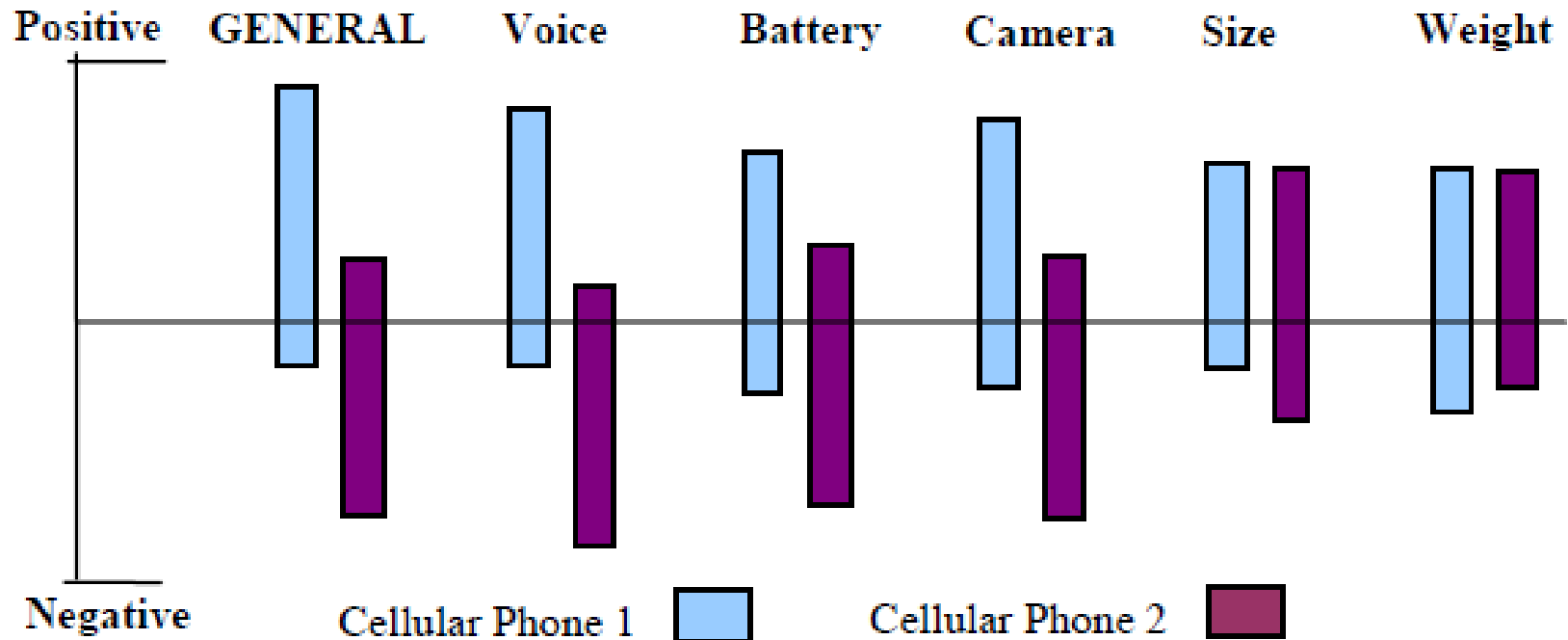


# Visualization of aspect-based summaries of opinions





# 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
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)



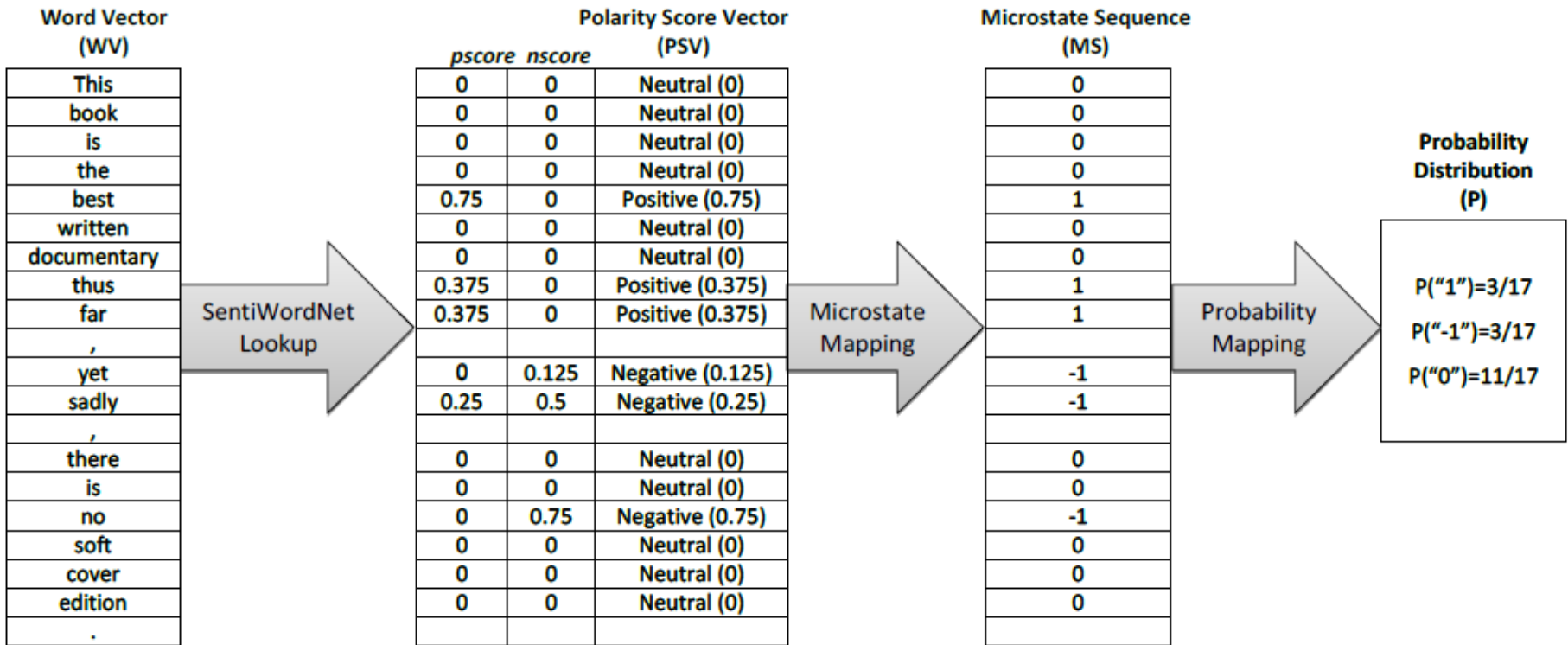
# Word-of-Mouth (WOM)

- “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.”



Word	POS
This	DT
book	NN
is	VBZ
the	DT
best	JJS
written	VCN
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

1. Social Media (社群媒體)
2. Social Word-of-Mouth (社群口碑)
3. Web Mining (網路探勘)



# 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](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 Mining (社群口碑與網路探勘)

## Q & A

戴敏育

Min-Yuh Day

Assistant Professor

Dept. of Information Management, Tamkang University

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

2012-10-31