

## 社群網路行銷分析

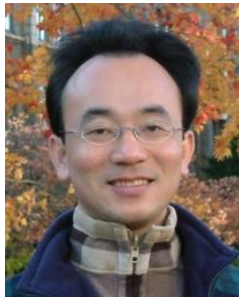
### 社群網路情感分析

## (Sentiment Analysis on Social Media)

1032SMMA09

TLMXJ1A (MIS EMBA)

Fri 12,13,14 (19:20-22:10) D326



Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

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

2015-06-05



# 課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
1	2015/02/27	和平紀念日補假(放假一天)
2	2015/03/06	社群網路行銷分析課程介紹 (Course Orientation for Social Media Marketing Analytics)
3	2015/03/13	社群網路行銷分析 (Social Media Marketing Analytics)
4	2015/03/20	社群網路行銷研究 (Social Media Marketing Research)
5	2015/03/27	測量構念 (Measuring the Construct)
6	2015/04/03	兒童節補假(放假一天)
7	2015/04/10	社群網路行銷個案分析 I (Case Study on Social Media Marketing I)
8	2015/04/17	測量與量表 (Measurement and Scaling)
9	2015/04/24	探索性因素分析 (Exploratory Factor Analysis)

# 課程大綱 (Syllabus)

週次 (Week)	日期 (Date)	內容 (Subject/Topics)
10	2015/05/01	社群運算與大數據分析 (Social Computing and Big Data Analytics) [Invited Speaker: Irene Chen, Consultant, Teradata]
11	2015/05/08	期中報告 (Midterm Presentation)
12	2015/05/15	確認性因素分析 (Confirmatory Factor Analysis)
13	2015/05/22	社會網路分析 (Social Network Analysis)
14	2015/05/29	社群網路行銷個案分析 II (Case Study on Social Media Marketing II)
15	2015/06/05	社群網路情感分析 (Sentiment Analysis on Social Media)
16	2015/06/12	期末報告 I (Term Project Presentation I)
17	2015/06/19	端午節補假 (放假一天)
18	2015/06/26	期末報告 II (Term Project Presentation II)

# Outline

- Affective Computing and Social Computing
- Opinion Mining and Sentiment Analysis
- Social Media Monitoring/Analysis
- Resources of Opinion Mining
- Opinion Spam Detection

# **Affective Computing and Social Computing**

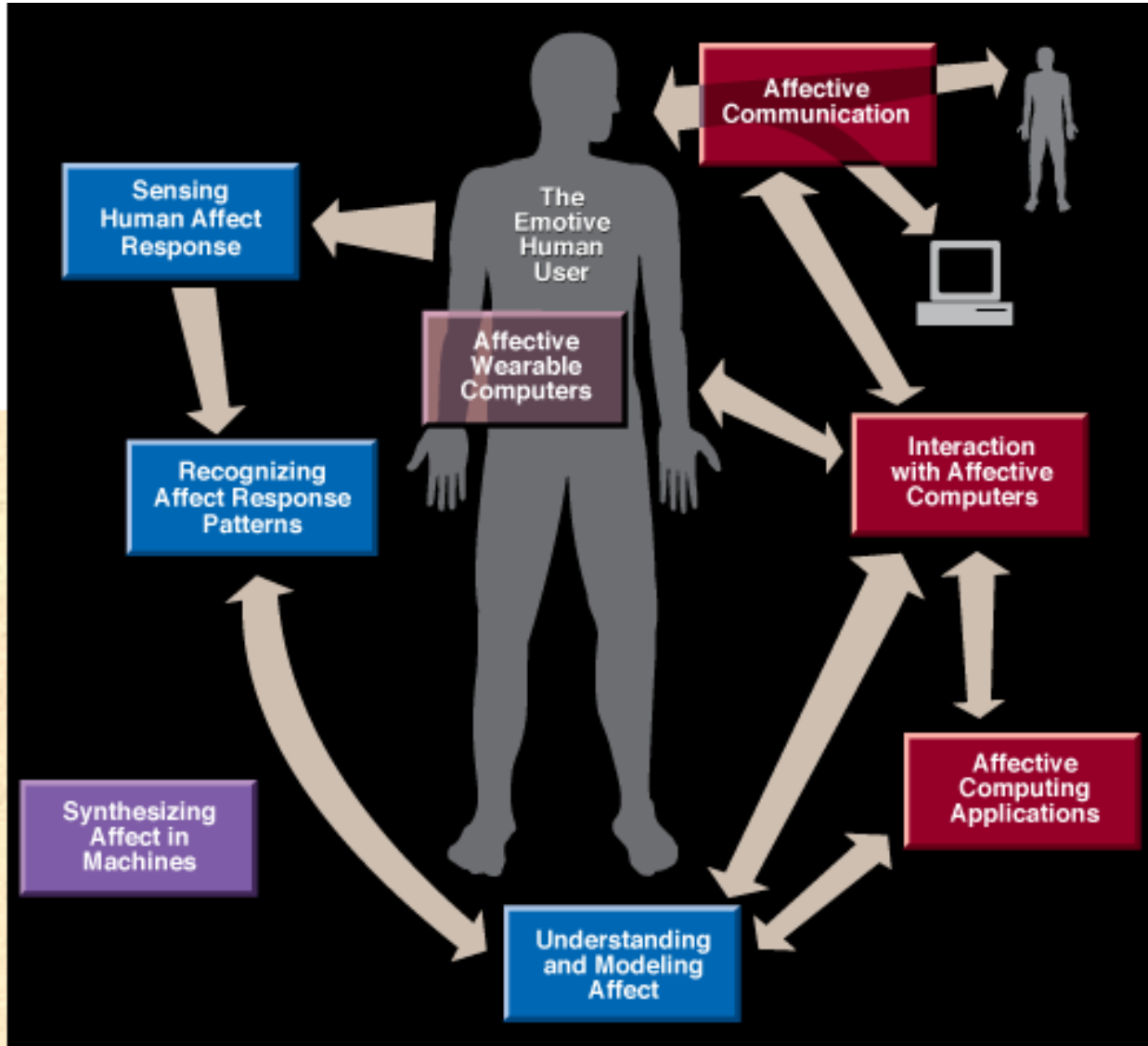
# Affective Computing

# Rosalind W. Picard, Affective Computing, The MIT Press, 2000

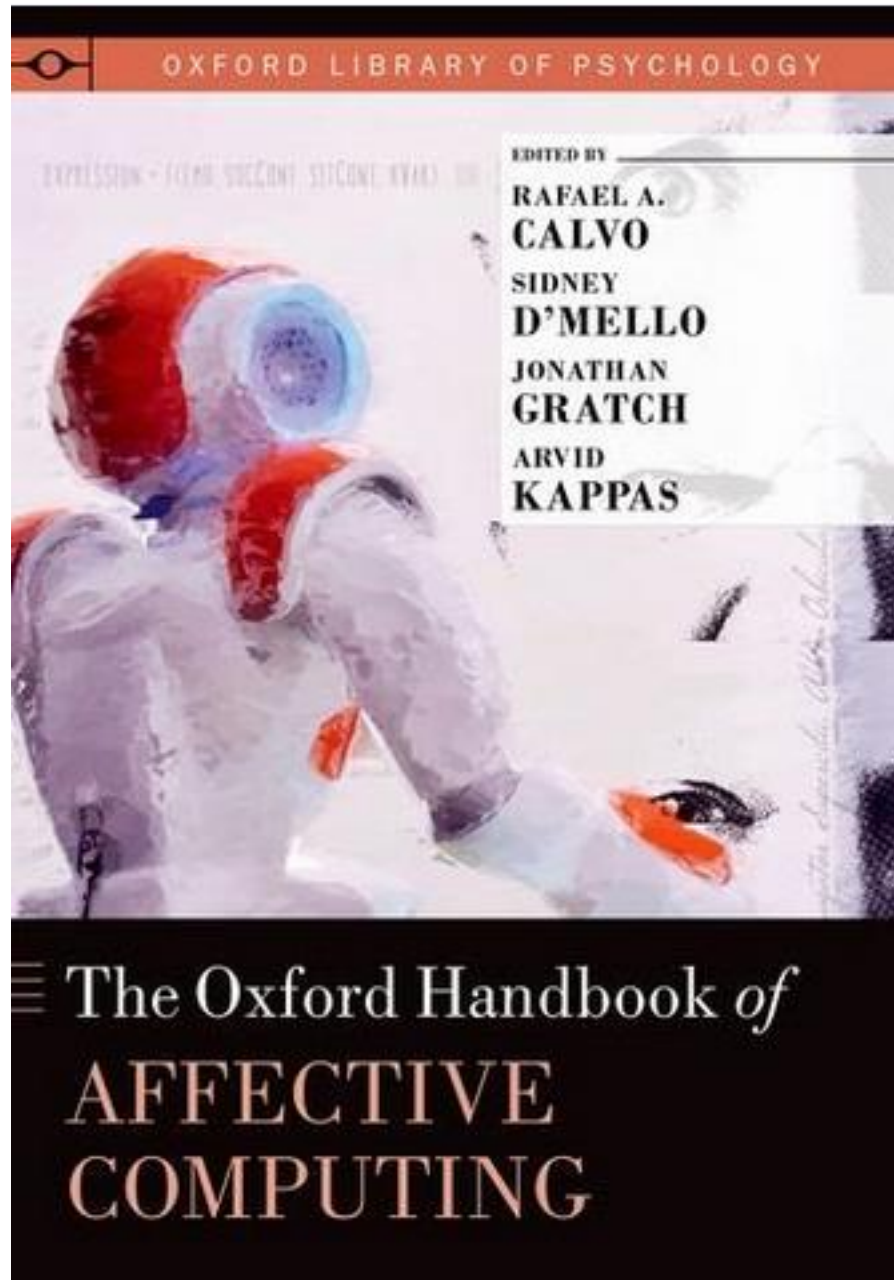


# Affective Computing

## Research Areas







# Affective computing

is the study and development of

systems and devices

that can

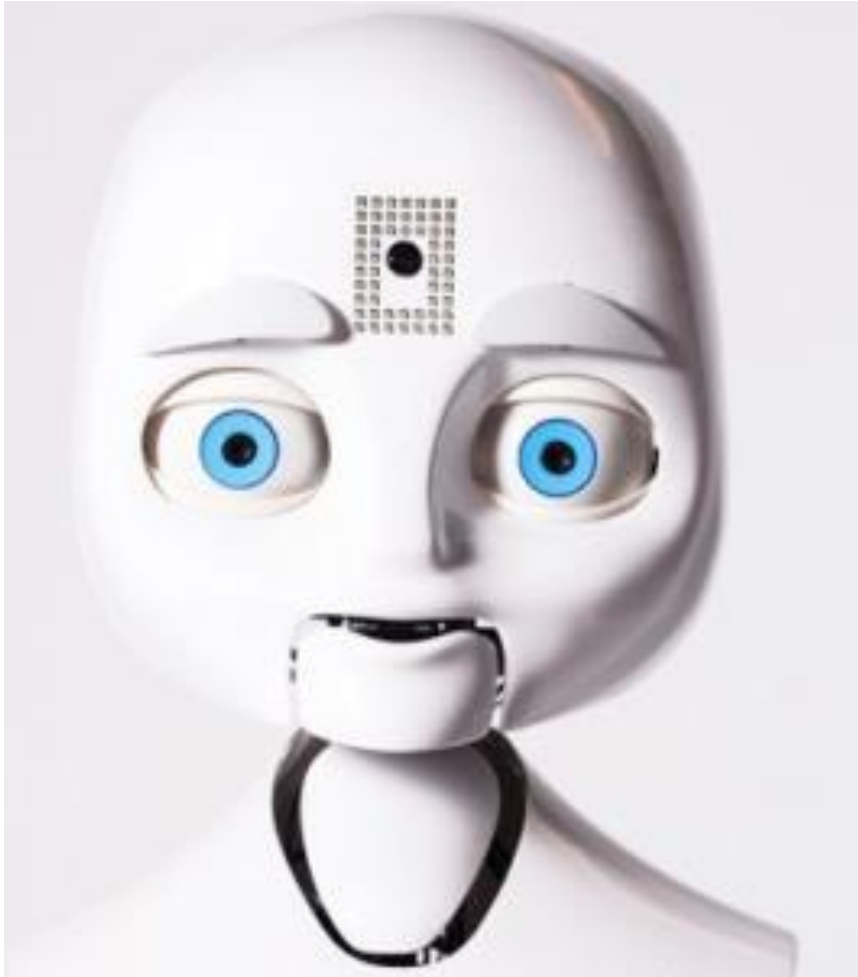
recognize, interpret,  
process, and simulate

human affects.

# Affective Computing

- **Affective Computing** research combines **engineering and computer science** with psychology, cognitive science, neuroscience, sociology, education, psychophysiology, value-centered design, ethics, and more.

# Affective Computing





# Wearable Tech

Discover the Gear™ that works best for your life.



Galaxy Gear™

Enjoy a personal assistant right on your wrist.



Gear™ 2

Take your best music with you on a stand-alone music player.



Gear™ 2 Neo

Answer calls, emails and texts, directly from your wrist.



Gear™ Fit

The world's first curved 1.84" Super AMOLED® display.



Gear™ Live

Now with the power of Google® Android Wear.



## Advancing Human Mobility

ASIMO drives more than just robotics research. Leading edge technologies developed for ASIMO provide a springboard for other Honda product development projects such as the ones shown below.



[Click to learn more](#)

## STRIDE MANAGEMENT ASSIST



Honda's Stride Management Assist device is designed to help those with weakened leg muscles but who are still able to walk. A motor helps lift each leg at the thigh as it moves forward and backward. This lengthens the user's stride, making it easier to cover longer distances at a greater speed.

# Emotions

Love

Anger

Joy

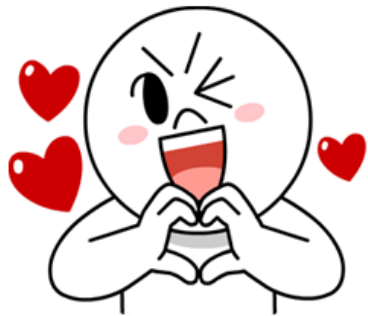
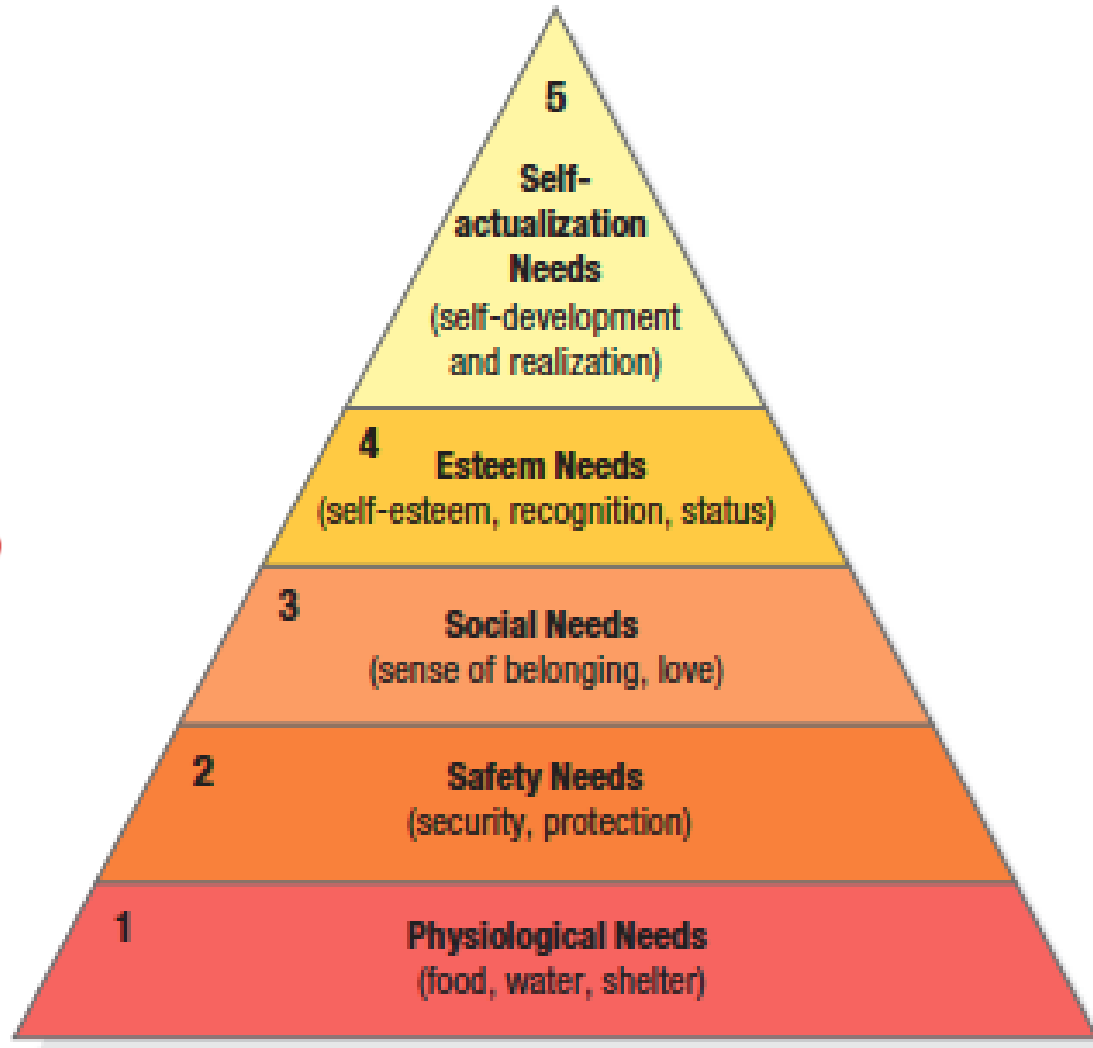
Sadness

Surprise

Fear

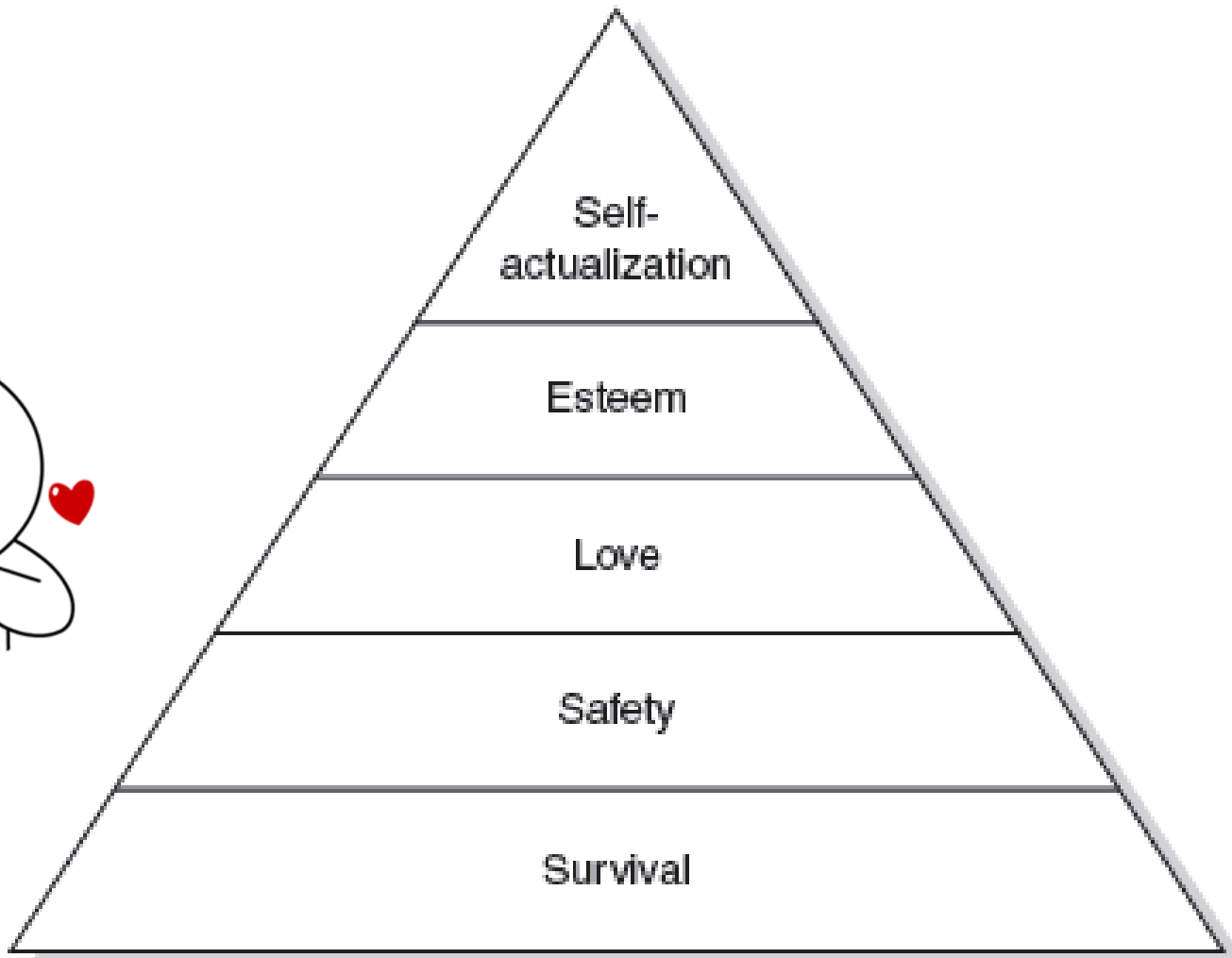


# Maslow's Hierarchy of Needs

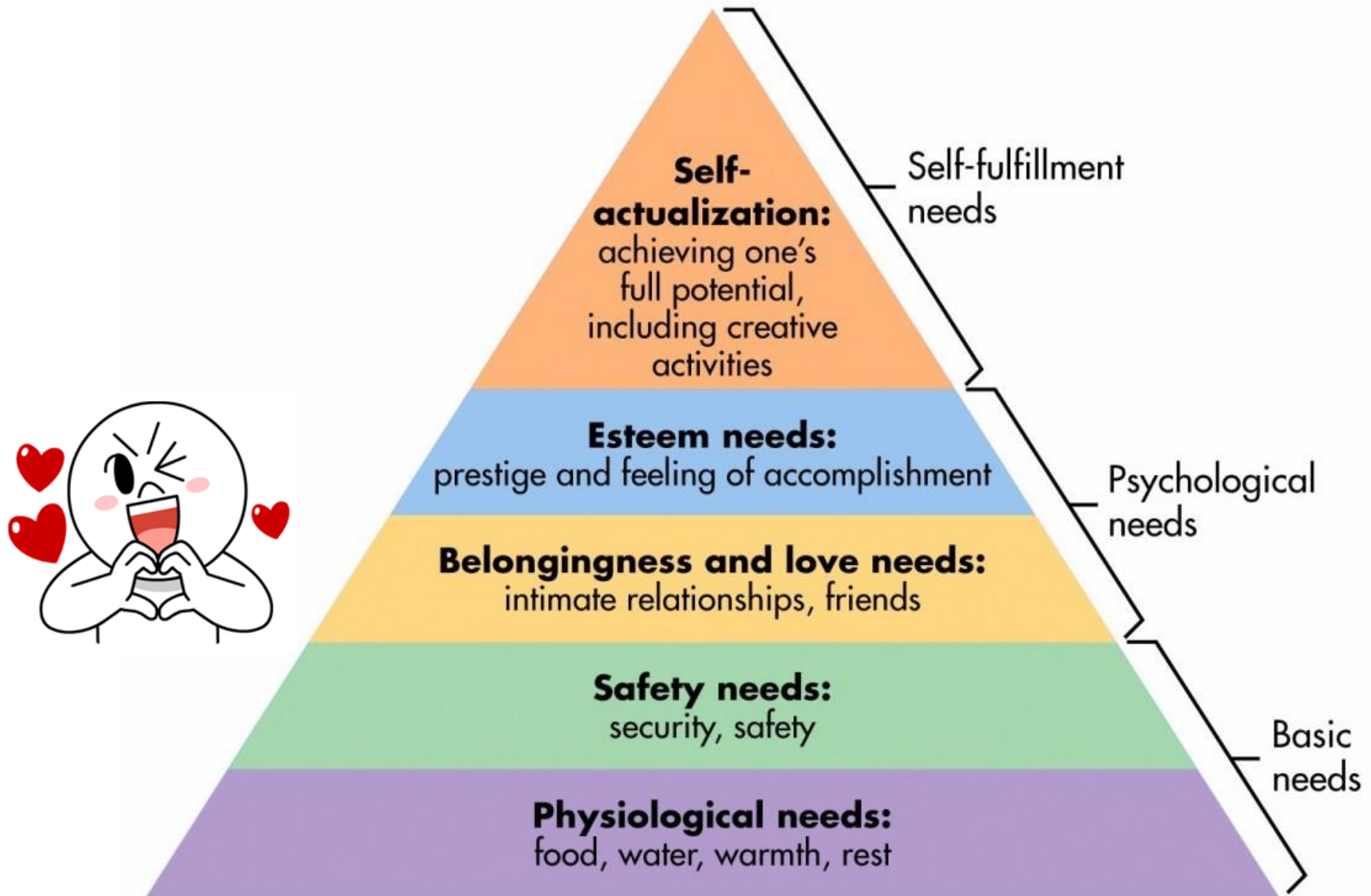


# Maslow's hierarchy of human needs

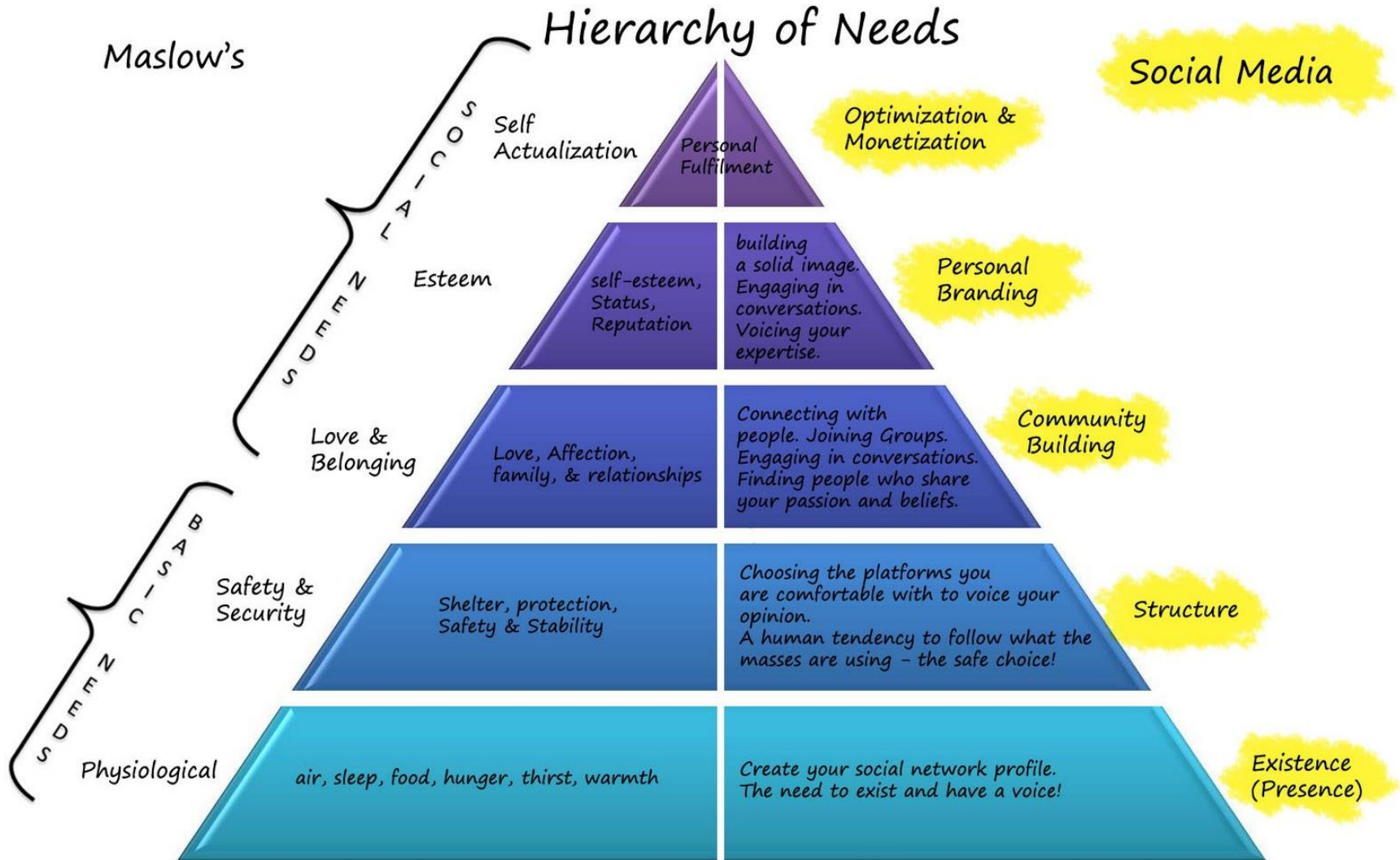
(Maslow, 1943)



# Maslow's Hierarchy of Needs

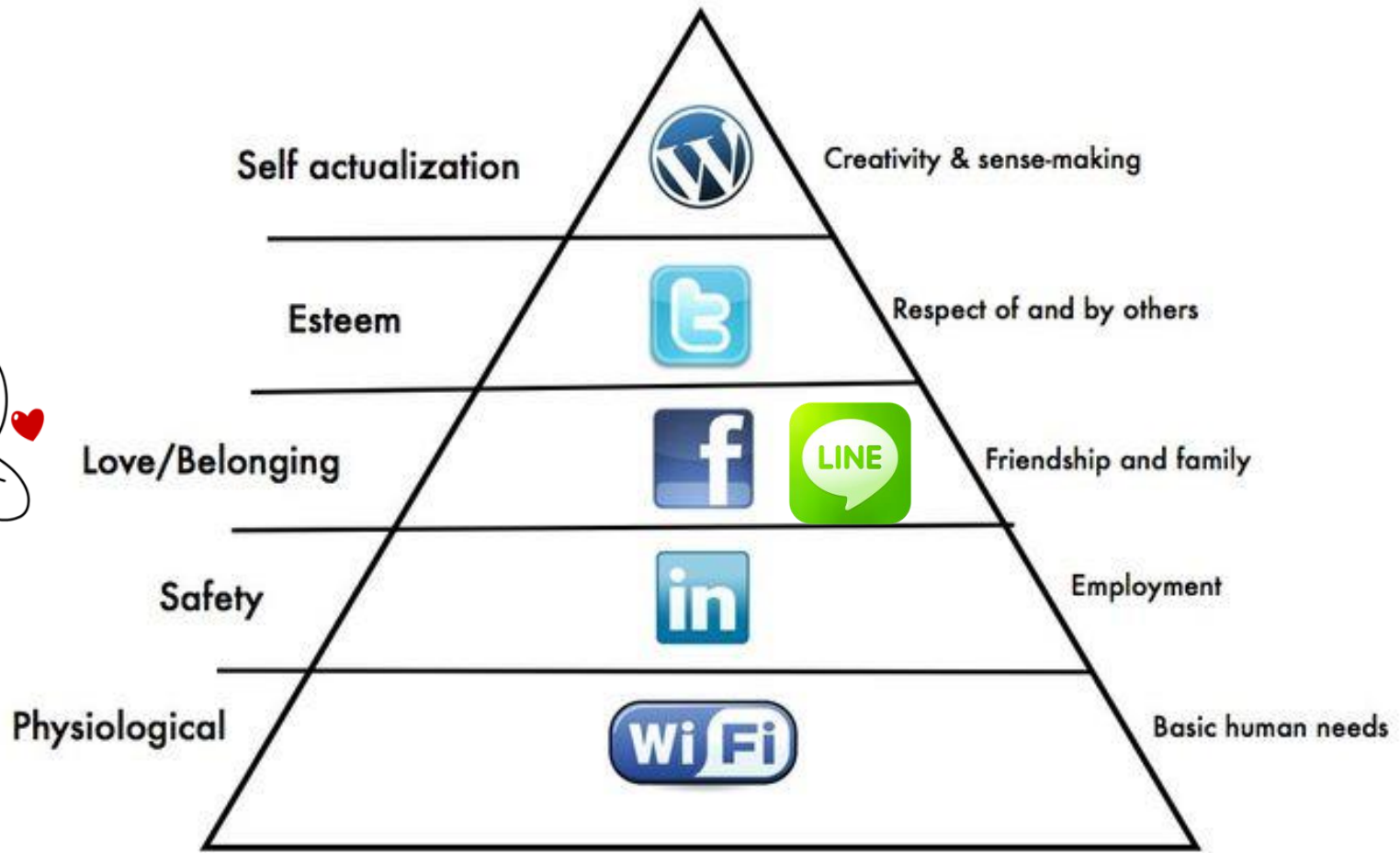


# Social Media Hierarchy of Needs



Social Media Hierarchy of Needs - by John Antonios

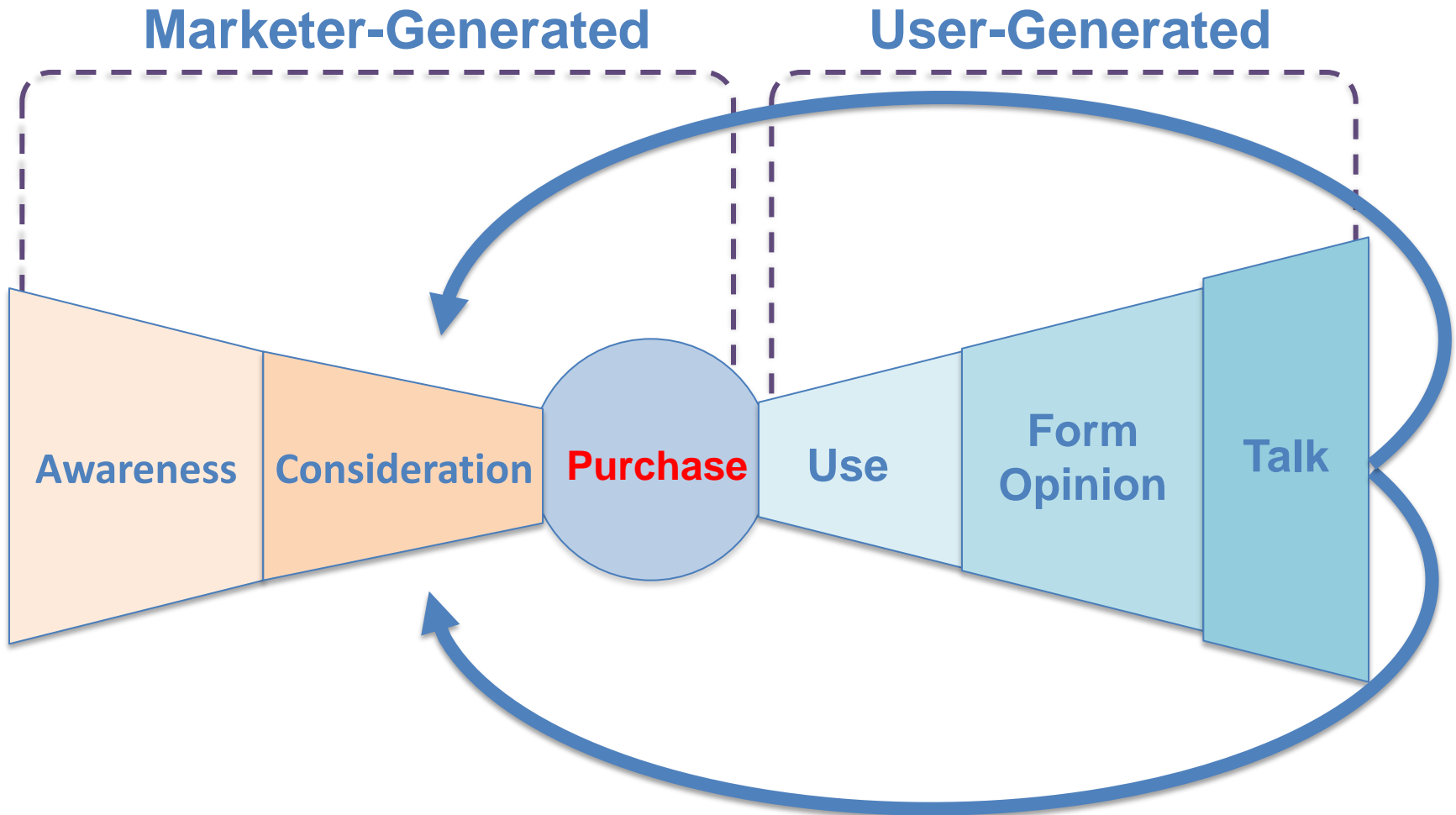
# Social Media Hierarchy of Needs



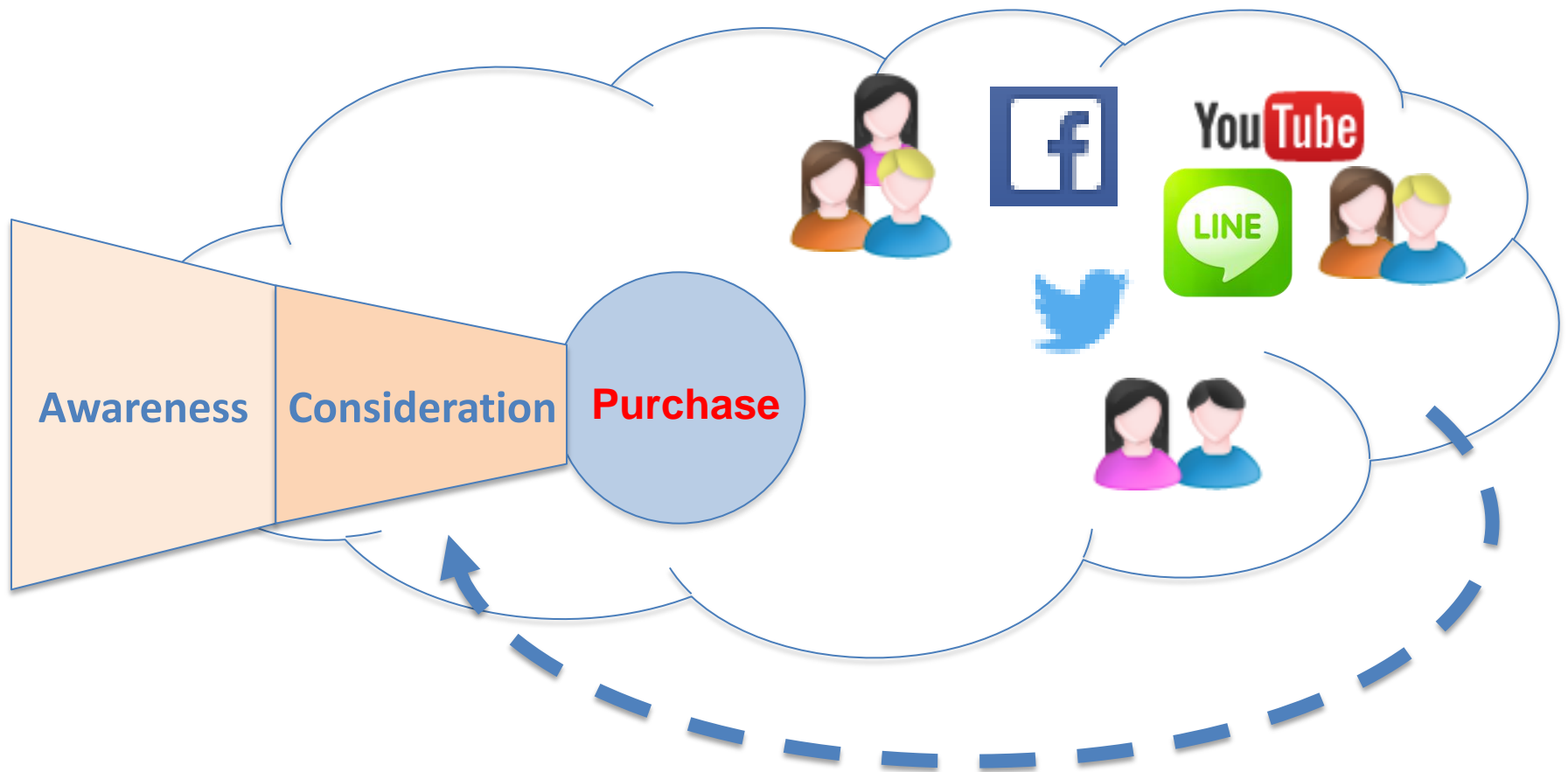
@daveduarte

# The Social Feedback Cycle

## Consumer Behavior on Social Media



# The New Customer Influence Path



# Social Computing



# Social Computing

- Social Computing
  - Business Computing
- Business Application
  - Content
  - Context
- Social Media Monitoring/Analysis
- Social Network Analysis

# 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

# Social Media and the Voice of the Customer

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

# Listen and Learn

## Text Mining for VoC

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

# **Opinion Mining and Sentiment Analysis**

# 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

# 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

# **Social Media Monitoring/Analysis**

# Existing Tools

## ("Social Media Monitoring/Analysis")

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



# Existing Tools

## ("Social Media Monitoring/Analysis")

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

# Word-of-mouth

## Voice of the Customer

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

# Attensity: Track social sentiment across brands and competitors

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

The screenshot displays the Attensity website homepage. At the top, there is a navigation bar with the Attensity logo on the left, a language selector set to 'English', and links for 'Contact', 'Resources', 'Support', and 'Blog'. A search bar is located on the right side of the navigation bar. Below the navigation bar, there are five main menu items: 'Products', 'Solutions', 'Services', 'Customers', and 'Partners'. The central hero section features a large dark grey box with the text 'Your real-time window into the social web.' and a quote from Yahoo! stating, 'Teaming with a leading analytics provider like Attensity offers Yahoo! a great opportunity to deliver the key news and analysis that matter.' Below the quote is a green 'Learn More' button. To the right of the hero section, there are several overlapping images of the Attensity dashboard, showing various charts and graphs. On the left side of the hero section, there is a vertical list of five menu items: 'Social Analytics', 'Social Response', 'Customer Analytics', 'Industry Solutions', and 'Why Attensity'. At the bottom of the page, there are several smaller sections: 'Attensity for Marketing', 'Attensity for Customer Service', 'Attensity for IT', 'Success Story' (featuring JetBlue Airways), 'About Attensity', and 'Watch Video' (featuring a 'Command Center Video' player).

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

# Clarabridge: Sentiment and Text Analytics Software

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

The image shows a screenshot of a web browser displaying the Clarabridge website. The browser's address bar shows the URL [www.clarabridge.com](http://www.clarabridge.com/). The website features a navigation menu with links for Home, About Us, News & Events, Blog, Login, and Contact Us. Below the navigation is a dark grey menu with categories: WHY TEXT ANALYTICS, PRODUCT, SERVICES, CUSTOMERS, PARTNERS, RESOURCES, and ABOUT US. The main content area has a blue background with the headline "The First Sentiment and Text Analytics Solution Built Specifically for Business." and a sub-headline "The Clarabridge sentiment and text analytics software provides enterprises with a universal view of their customers." A "Learn more about how Clarabridge works >" link is positioned at the bottom right of this section. Below the main content is a "Customers" section displaying logos for Nissan, Best Buy, Marriott, Sage, H&R Block, Choice Hotels International, Wingy's, GWLORD HOTELS, BE ERDFORCE, and DOLL, followed by a "More >" link. The footer contains three promotional boxes: "Clarabridge Text Analytics", "Choose Your Edition" (with a sub-section for "Clarabridge for Enterprises" described as ideal for enterprise-class text analytics solutions), and "Clarabridge Webinar" (presented by Hypatia Research Group 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.

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

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

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

**Newsletter** Sign up and get the regular Radian6 goods.  
Enter email address GO

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

JUST Get the Skinny  
WEBINAR / June 7th at 2pm est  
CASE STUDY

radian6 Community

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

Social Media Monitoring x

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

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.

**“ 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**

**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 screenshot shows a web browser window with the URL [www.tweetfeel.com/index.php#iPhone4s](http://www.tweetfeel.com/index.php#iPhone4s). The page features the 'tweetfeel' logo with a blue bird icon. A search bar contains the text 'iPhone4s' and a yellow 'Search' button. Below the search bar, it displays 'Try some Twitter trends: [Tomorrow is June](#) [H&M](#) [Defense of Marriage Act](#) [Diddy's](#) [Bloomberg](#) [UCLA](#) [ESPN](#)'. A sentiment analysis graphic shows a green smiley face with '40' below it, a red frowny face with '41' below it, an equals sign, and '51%' in red. Below this, a message reads: '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.](#)'

Below the message are six tweet snippets, each with a small profile picture on the left:

- 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.

The footer contains links for 'Read our FAQ', 'Legal Stuff', '100% Guarantee', and 'Share'. It also includes social media icons for 'Follow us' and 'Email us', a logo for 'conversion' with the text 'Brought to you by', and the 'Powered by twitter' logo.

Browser window: Tweet Sentiments - Know ' x  
Address bar: tweetsentiments.com/analyze?utf8=✓&q=iphone4s&topic=true&commi ☆

Lang: Eng  
Eng **tweet** v0.8.6  
Dashboard Analyze Recent Top 1000 Users Maps Charts About Login


### Tweet Sentiment Analysis

iphone4s  
User  Topic   
Analyze Tweets

Feedback



**Sentiments**  
Index[0-100]: 50.00  
Positive: 8  
Negative: 1  
Neutral: 13  
Total Analyzed: 0

**Profile**  
Education:  
Flamboyance:  
Slang:  
Gender:  
Age:





50.00



---

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



---

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



---

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

---

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

---

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



i-Buzz網路口碑研究中心. 網 x

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

iBuzz中國 | 關於我們 歡迎光臨。請 [加入會員](#) 或 [登錄](#)

**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

血汗醫院全民皆慌 網路熱門

#### 研究專題

便利商店衝人氣 臉書加持少

#### 數據報告

永慶房屋得據房仲龍頭 資訊

#### f 臉書粉絲團排行榜

政治人物 名人 演藝人員

馬英九總統	蔡英文 Tsai	蘇貞昌
粉絲數 1419591	粉絲數 669928	粉絲數 317035

#### 排行榜

##### Traffic 交通

	第1名
FORD (FIESTA)	
	第2名
PORSCHE (911)	
	第3名
SUBARU (legacy)	

[更多](#)

##### Traffic 交通

	第1名
--	-----

#### 進口車

- 吵了這麼久，就是不要「超速」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 a navigation bar with the eLAND logo and menu items: 產品, 雲端服務, 壹銀標專區, 我們的客戶, 新聞與活動, 支援, and 關於意藍. The main content area has a large banner for 'OpView Service 您的輿論觀測站' with the tagline '連上OpView · 品牌形象、輿論觀點監測好輕鬆'. Below the banner are five circular icons representing different services. To the right is a 3D-style image of the OpView software interface. Below the banner is a sidebar with the heading 'OpView 服務總覽' and a search box. The sidebar contains the text 'OpView 服務總覽', '什麼是OpView服務?', and 'OpView是協助您蒐集、處理、分析網路資訊的雲端服務。'. Below this is a '相關文件下載' section with a list of documents: 'OpView Insight DM', 'OpView品牌口碑雷達 DM', and 'OpView品牌口碑雷達訂購單'. A '更多' link is at the bottom of the download section.

# 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

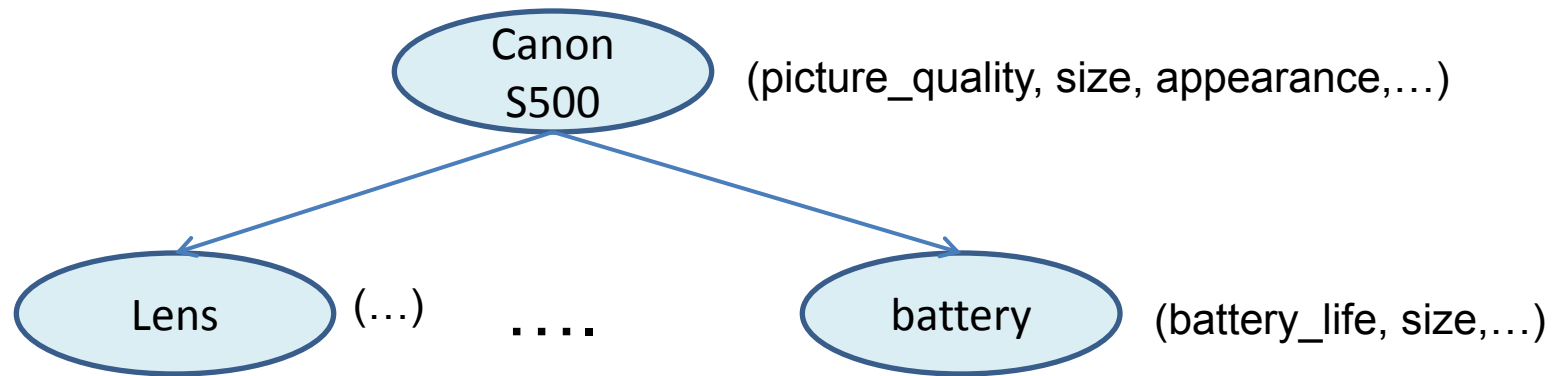
# A (regular) opinion

- **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>

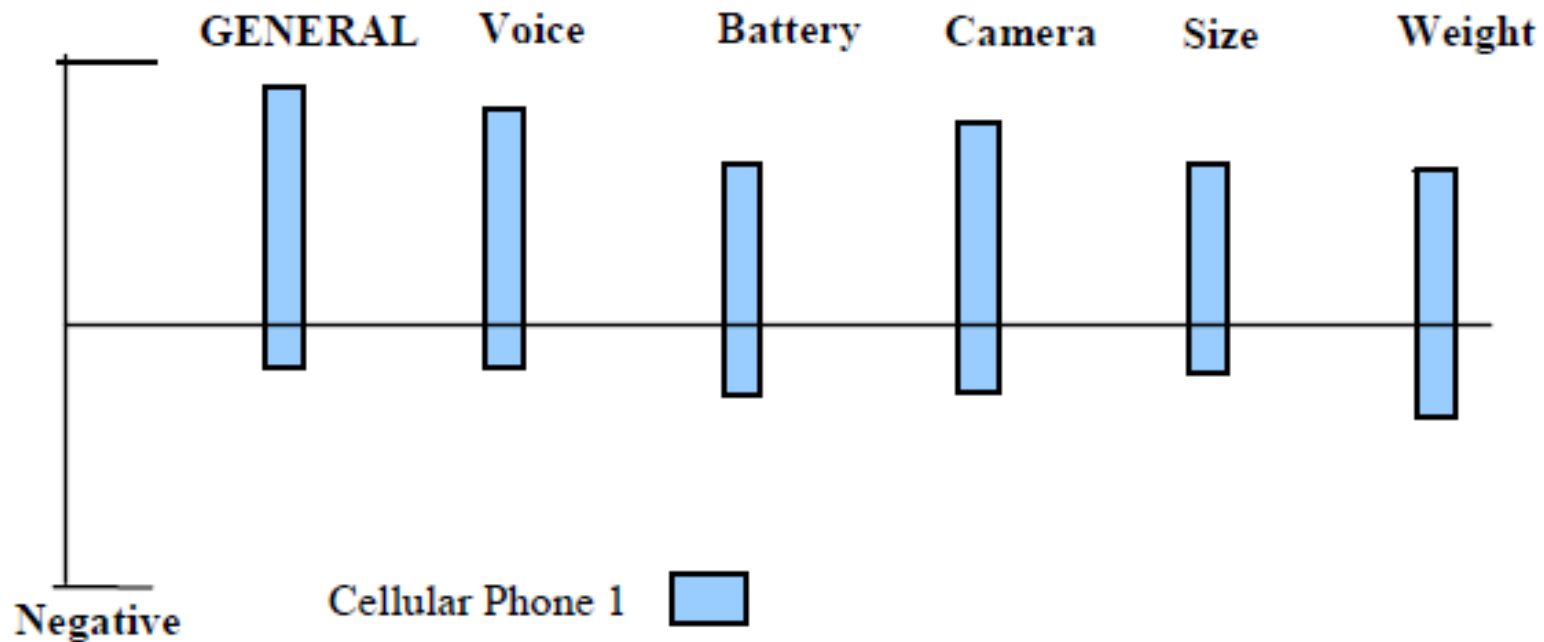
Aspect: **Battery**

Positive: 80 <individual review sentences>

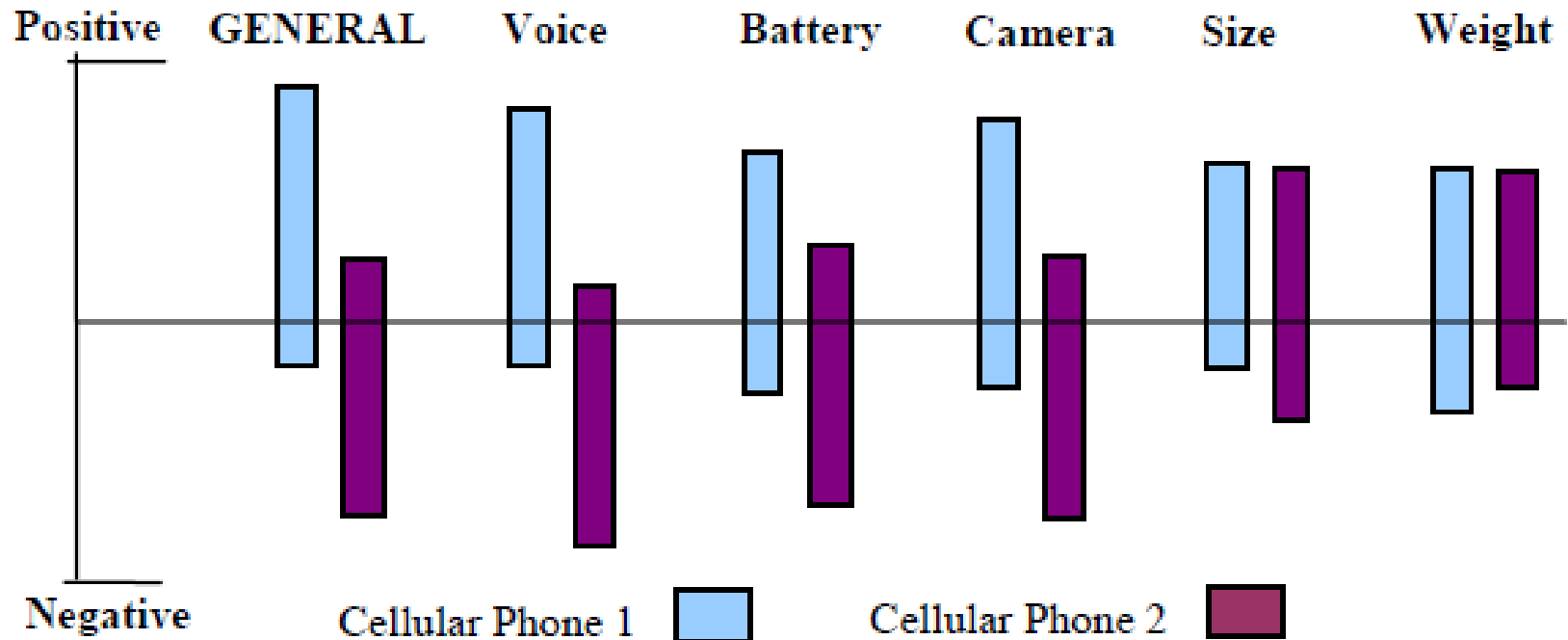
Negative: 12 <individual review sentences>

...

# 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

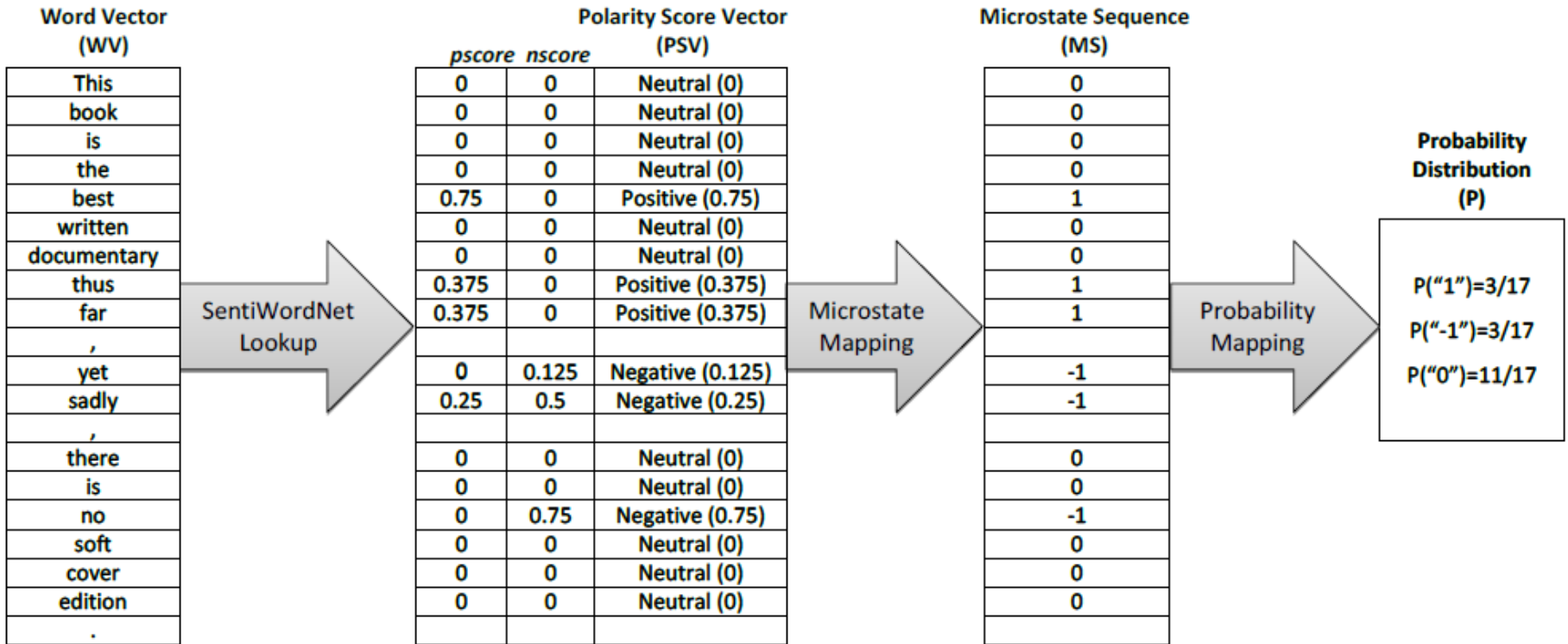
(Both = Subjectivity and Polarity; ML= Machine Learning; Lexicon/Rule= Lexicon enhanced by linguistic rules)

# 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	This	DT
book	book	NN
is	is	VBZ
the	the	DT
best	best	JJS
written	written	VCN
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



# Resources of Opinion Mining



# Datasets of Opinion Mining

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



# CKIP 中研院中文斷詞系統

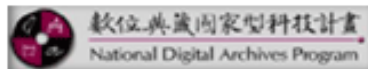
<http://ckipsvr.iis.sinica.edu.tw/>

## 中文斷詞系統

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

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

[隱私權聲明](#) | [版權聲明](#)



Copyright © National Digital Archives Program, Taiwan.  
All Rights Reserved.

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

自 2014/01/06 起，本斷詞系統已經處理過 929135 篇文章

[送出](#)

[清除](#)

歐巴馬是美國的一位總統

歐巴馬是美國的一位總統

[文章的文字檔](#)

[擷取未知詞過程](#)


[包含未知詞的斷詞標記結果](#)

[未知詞列表](#)

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

# 中文文字處理：中文斷詞

## 抗氣候變遷 白宮籲採緊急行動

 中央社 – 2014年5月6日 下午10:58

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

這份為期4年的調查警告，極端氣候事件將對住家、基礎設施及產業帶來嚴重威脅。

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

但歐巴馬在任上一直未能說服美國國會採取重大行動。

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

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

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

海平面上升也將吞噬密西西比等低窪地區。

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

抗氣候變遷 白宮籲採緊急行動

中央社中央社 – 2014年5月6日 下午10:58

（中央社華盛頓6日綜合外電報導）白宮今天公布全球暖化對全美及美國經濟關鍵產業造成何種衝擊的新報告，呼籲採取緊急行動對抗氣候變遷。這份為期4年的調查警告，極端氣候事件將對住家、基礎設施及產業帶來嚴重威脅。

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

但歐巴馬在任上一直未能說服美國國會採取重大行動。

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

美國數百名來自政府與民間的頂尖氣候科學家及技術專家，共同投入這項研究，檢視氣候變遷對當今帶來的衝擊並預測將對下個世紀帶來何種影響。研究人員警告，加州可能發生旱災、奧克拉荷馬州發生草原大火，東岸則可能遭遇海平面上升，尤其佛羅里達，而這些事件多為人類造成。

海平面上升也將吞噬密西西比等低窪地區。

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

報告並說：「過去被認為是遙遠未來議題的氣候變遷，已著實成為當前議題。」（譯者：中央社蔡佳伶）1030506

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

# CKIP 中研院中文斷詞系統

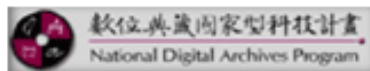
<http://ckipsvr.iis.sinica.edu.tw/>

## 中文斷詞系統

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

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

[隱私權聲明](#) | [版權聲明](#)



Copyright © National  
Digital Archives Program,  
Taiwan.  
All Rights Reserved.

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

自 2014/01/06 起，本斷詞系統已經處理過 929136 篇文章

送出

清除

抗氣候變遷 白宮籲採緊急行動

中央社中央社 - 2014年5月6日 下午10:58

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

這份為期4年的調查警告，極端氣候事件將對住家、基礎設施及產業帶來嚴重威脅。

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

但歐巴馬在任上一直未能說服美國國會採取重大行動。

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

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

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

海平面上升也將吞噬密西西比等低窪地區。

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

報告並說：「過去被認為是遙遠未來議題的氣候變遷，已著實成為當前議題。」(譯者：中央社蔡佳伶) 1030506

# CKIP 中研院中文斷詞系統

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

## 中文斷詞系統

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

抗(VJ) 氣候(Na) 變遷(VH) 白宮(Nc) 籲(VE) 採(VC) 緊急(VH) 行動(Na) 中央社(Nc) 中央社(Nc) 2014年(Nd) 5月(Nd) 6日(Nd) 下午(Nd) 1  
58(Neu) ((PARENTHESISCATEGORY) 中央社(Nc) 華盛頓(Nc) 6日(Nd) 綜合(A) 外電(Na) 報導(VE) ) (PARENTHESISCATEGORY) 白宮(Nc) 今天(Nd)  
呼籲(VE) 採取(VC) 緊急(VH) 行動(Na) 對抗(VC) 氣候(Na) 變遷(VH) 。(PERIODCATEGORY)  
這(Nep) 份(Nf) 為期(VH) 4年(Nd) 的(DE) 調查(VE) 警告(VE) 。(COMMACATEGORY)  
極端(VE) 氣候(Na) 事件(Na) 將(D) 對(P) 住家(Na) 、(PAUSECATEGORY) 基礎(VH) 設施(Na) 及(Caa) 產業(Na) 帶來(VC) 嚴重(VH) 威脅(Na) 。  
美國(Nc) 總統(Na) 歐巴馬(Nb) 2008年(Nd) 當選(VG) 總統(Na) 時(Ng) 普(D) 在(P) 競選(VC) 造勢(VB) 時(Ng) 誓言(VE) 。(COMMACATEGORY)  
要(D) 讓(VL) 美國(Nc) 成為(VG) 對抗(VC) 氣候(Na) 變遷(VH) 與(Caa) 相關(VH) 「(PARENTHESISCATEGORY) 安全(VH) 威脅(Na) 」(PARENTHESISCATEGORY)  
但(Cbb) 歐巴馬(Nb) 在任(VH) 上(Ng) 一直(D) 未(D) 能(D) 說服(VF) 美國(Nc) 國會(Nc) 採取(VC) 重大(VH) 行動(Na) 。(PERIODCATEGORY)  
在(P) 本(Nes) 週(Nf) 對(P) 這(Nep) 項(Nf) 議題(Na) 採取(VC) 的(DE) 新作(Na) 為(P) 中(Ncd) 。(COMMACATEGORY)  
歐巴馬(Nb) 今天(Nd) 將(D) 與(P) 數(Neu) 名(Nf) 氣象學家(Na) 接受(VC) 電視(Na) 訪問(VC) 。(COMMACATEGORY)  
討論(VE) 美國(Nc) 全國(Nc) 氣候(Na) 評估(VE) 第3(Neu) 版(Na) 調查(VE) 結果(Dk) 。(PERIODCATEGORY)  
美國(Nc) 數百(Neu) 名(Nf) 來自(VJ) 政府(Na) 與(Caa) 民間(Nc) 的(DE) 頂尖(VH) 氣候(Na) 科學家(Na) 及(Caa) 技術(Na) 專家(Na) 。(COMMACATEGORY)  
共同(A) 投入(VC) 這(Nep) 項(Nf) 研究(Na) 。(COMMACATEGORY)  
檢視(VC) 氣候(Na) 變遷(VH) 對(P) 當今(Nd) 帶來(VC) 的(DE) 衝擊(Na) 並(D) 預測(VE) 將(D) 對(P) 下(Nes) 個(Nf) 世紀(Na) 帶來(VC) 何  
研究(Na) 人員(Na) 警告(VE) 。(COMMACATEGORY)  
加州(Nc) 可能(D) 發生(VJ) 旱災(Na) 、(PAUSECATEGORY) 奧克拉荷馬州(Nc) 發生(VJ) 草原(Na) 大火(Na) 。(COMMACATEGORY)  
東岸(Nc) 則(D) 可能(D) 遭遇(VJ) 海平面(Na) 上升(VA) 。(COMMACATEGORY)  
尤其(D) 佛羅里達(Nc) 。(COMMACATEGORY)  
而(Cbb) 這些(Nega) 事件(Na) 多(D) 為(VG) 人類(Na) 造成(VK) 。(PERIODCATEGORY)  
海平面(Na) 上升(VA) 也(D) 將(D) 吞噬(VC) 密西西比(Nb) 等(Cab) 低窪(VH) 地區(Nc) 。(PERIODCATEGORY)  
至於(P) 超過(VJ) 8000萬(Neu) 人(Na) 居住(VA) 且(Cbb) 擁有(VJ) 全美(Nb) 部分(Nega) 成長(VH) 最(Dfa) 快(VH) 都會區(Nc) 的(DE) 東  
「(PARENTHESISCATEGORY) 海平面(Na) 上升(VA) 加上(VC) 其他(Nega) 與(Caa) 氣候(Na) 變遷(VH) 有關(VJ) 的(DE) 衝擊(Na) 。(COMMACATEGORY)

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

[隱私權聲明](#) | [版權聲明](#)



Copyright © National Digital Archives Program, Taiwan. All Rights Reserved.



## The Stanford Natural Language Processing Group

[home](#) · [people](#) · [teaching](#) · [research](#) · [publications](#) · [software](#) · [events](#) · [local](#)

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

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

### Supported software distributions

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

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

#### Stanford CoreNLP

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

#### Stanford Parser

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

#### Stanford POS Tagger

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



# Stanford NLP Software

## Stanford CoreNLP

Output format: Visualise

Please enter your text here:

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

Submit

Clear

### Part-of-Speech:

	NP	NP	VBZ	JJ	IN	NP	.
1	Stanford	University	is	located	in	California	.
2	It	is	a	great	university	.	

### Named Entity Recognition:

	Organization		Location
1	Stanford University	is located in	California
2	It is a great university	.	

### Coreference:

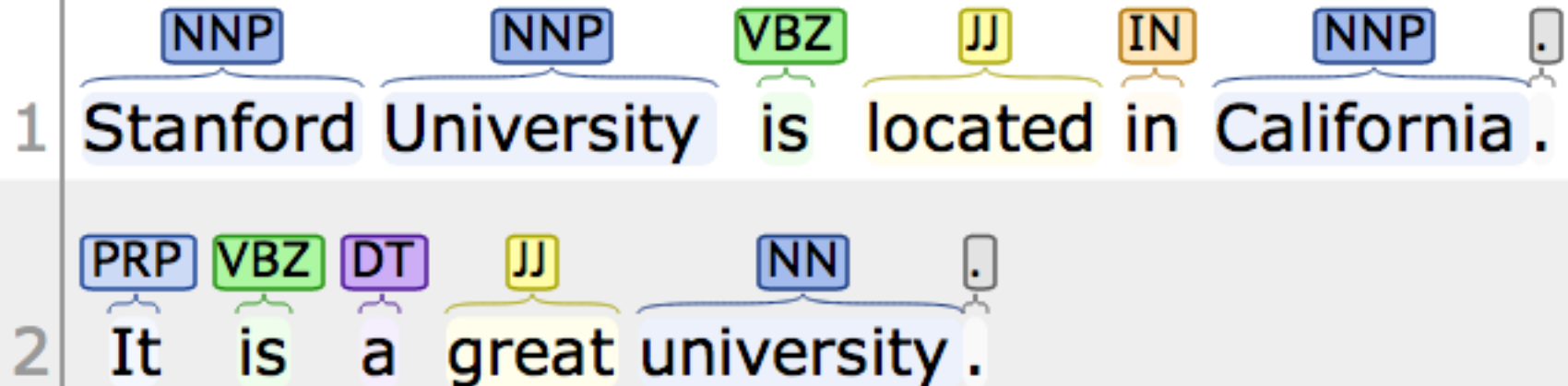
	Mention	----- Coref -----
1	Stanford University	is located in California.
2	It	is a great university.

# Stanford CoreNLP

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

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

## Part-of-Speech:



# Stanford CoreNLP

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

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

## Named Entity Recognition:

	<b>Organization</b>		<b>Location</b>
1	Stanford University	is located in	California .
2	It is a great university .		

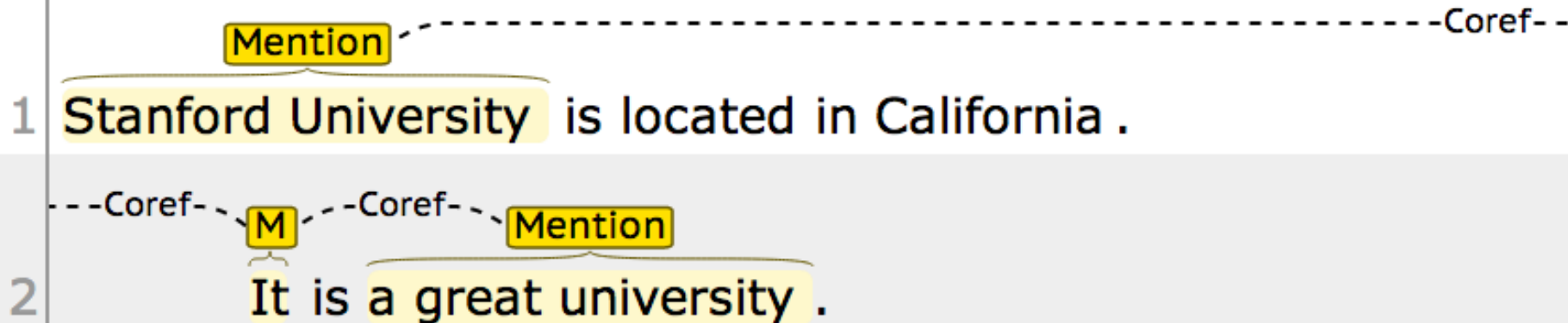


# Stanford CoreNLP

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

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

## Coreference:

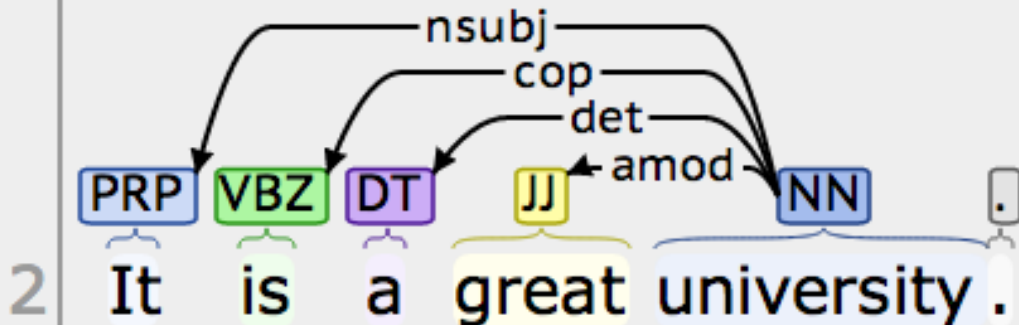
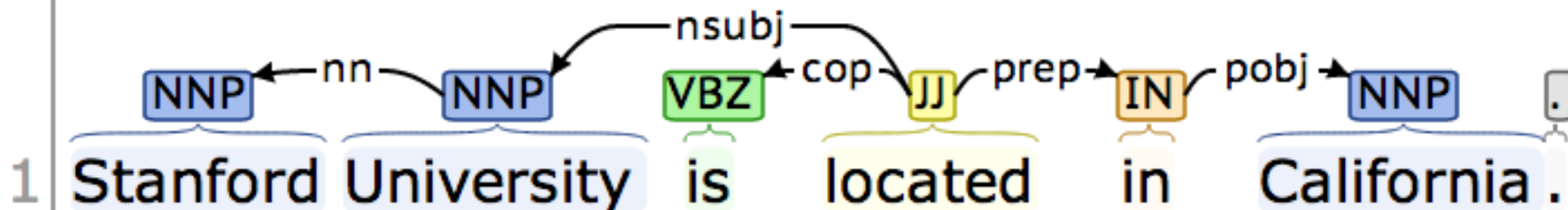


# Stanford CoreNLP

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

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

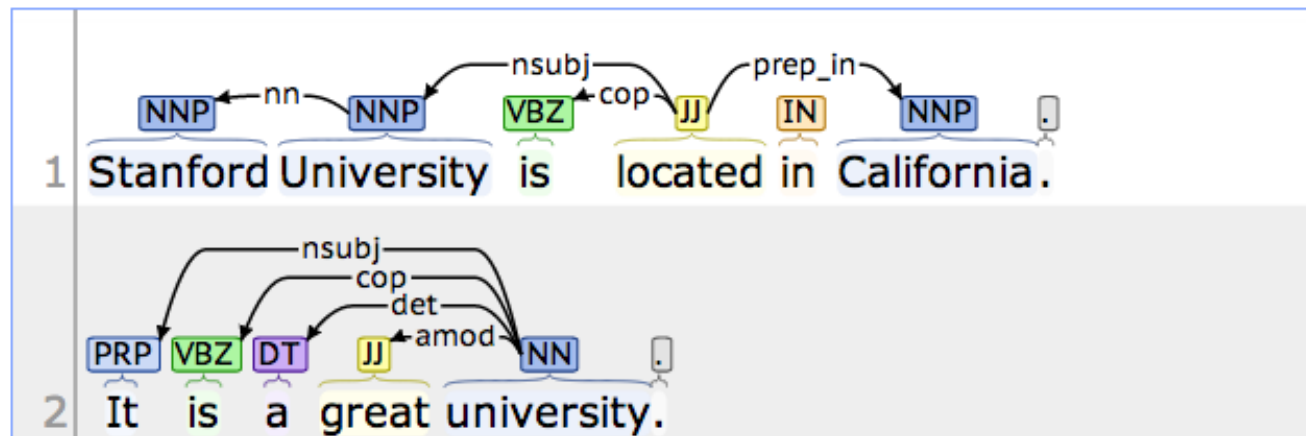
## Basic dependencies:



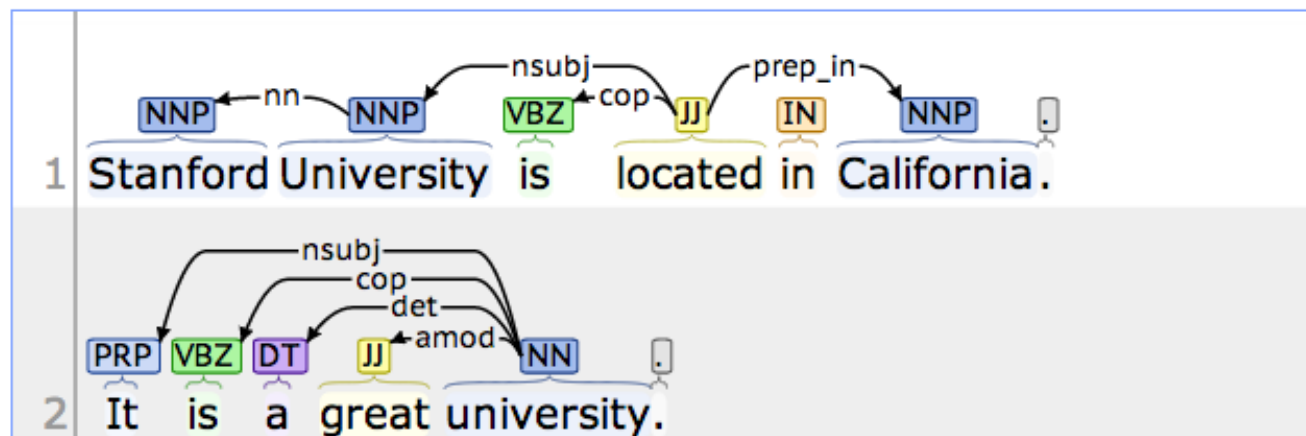
# Stanford CoreNLP

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

## Collapsed dependencies:



## Collapsed CC-processed dependencies:



Visualisation provided using the [brat visualisation/annotation software](#).  
Copyright © 2011, Stanford University, All Rights Reserved.

Output format:  ↕

Please enter your text here:

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

### Stanford CoreNLP XML Output

Document								
Document Info								
Sentences								
<b>Sentence #1</b>								
<i>Tokens</i>								
Id	Word	Lemma	Char begin	Char end	POS	NER	Normalized NER	Speaker
1	Stanford	Stanford	0	8	NNP	ORGANIZATION		PERO
2	University	University	9	19	NNP	ORGANIZATION		PERO
3	is	be	20	22	VBZ	O		PERO
4	located	located	23	30	JJ	O		PERO
5	in	in	31	33	IN	O		PERO
6	California	California	34	44	NNP	LOCATION		PERO
7	.	.	44	45	.	O		PERO
<i>Parse tree</i>								
(ROOT (S (NP (NNP Stanford) (NNP University)) (VP (VBZ is) (ADJP (JJ located) (PP (IN in) (NP (NNP California)))))) (. .)))								

# Stanford CoreNLP

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

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

## Sentence #1

### Tokens

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

### Parse tree

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

# Stanford CoreNLP

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

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

*Sentence #2*

*Tokens*

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

*Parse tree*

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

# Stanford CoreNLP

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

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

## Coreference resolution graph

1.

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

## Tokens

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

## Parse tree

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

## Uncollapsed dependencies

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

root ( ROOT-0 , located-4 )  
nn ( University-2 , Stanford-1 )  
nsubj ( located-4 , University-2 )  
cop ( located-4 , is-3 )  
prep\_in ( located-4 , California-6 )  
Collapsed dependencies with CC processed

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

# Stanford CoreNLP

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

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



Output format: 

Please enter your text here:

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

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

# NER for News Article

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

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

**2K**  
TOTAL SHARES

461

1K

74

25

## Bill Gates no longer Microsoft's biggest shareholder

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

Recommend 1.2k



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

**2K**  
TOTAL SHARES

461 1K 74 25

NEW YORK (CNNMoney)

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

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

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

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

NEW YORK (CNNMoney)

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

That puts him behind Microsoft's former CEO Steve Ballmer who owns 333 million shares.

Related: Gates reclaims title of world's richest billionaire Ballmer, who was Microsoft's CEO until earlier this year, was one of Gates' first hires.

It's a passing of the torch for Gates who has always been the largest single owner of his company's stock. Gates now spends his time and personal fortune helping run the Bill & Melinda Gates foundation.

The foundation has spent \$28.3 billion fighting hunger and poverty since its inception back in 1997.

# Stanford Named Entity Tagger (NER)

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

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

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

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

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

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

Potential tags:

- LOCATION**
- TIME**
- PERSON**
- ORGANIZATION**
- MONEY**
- PERCENT**
- DATE**

# Stanford Named Entity Tagger (NER)

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

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder  
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET  
Bill Gates sold nearly 8 million shares of Microsoft over the past two days.  
NEW YORK (CNNMoney) —

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

# Stanford Named Entity Tagger (NER)

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

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

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

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

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

Copyright © 2011, Stanford University. All Rights Reserved.

# Stanford Named Entity Tagger (NER)

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

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

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

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

NEW YORK (CNNMoney) —

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

# Stanford Named Entity Tagger (NER)

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

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

Bill Gates no longer Microsoft's biggest shareholder  
By Patrick M. Sheridan @CNNTech May 2, 2014: 5:46 PM ET  
  
Bill Gates sold nearly 8 million shares of Microsoft over the past two days.  
  
NEW YORK (CNNMoney)

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

Potential tags:

LOCATION

ORGANIZATION

PERSON

MISC

# Stanford Named Entity Tagger (NER)

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

## Stanford Named Entity Tagger

Classifier:

Output Format:

Preserve Spacing:

Please enter your text here:

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

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

NEW YORK (CNNMoney) —

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

Potential tags:

**LOCATION**

**ORGANIZATION**

**PERSON**



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

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

Potential tags:

**LOCATION**

**TIME**

**PERSON**

**ORGANIZATION**

**MONEY**

**PERCENT**

**DATE**

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

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

Potential tags:

**LOCATION**

**ORGANIZATION**

**PERSON**

# Stanford Named Entity Tagger (NER)

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

## Stanford NER Output Format: inlineXML

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

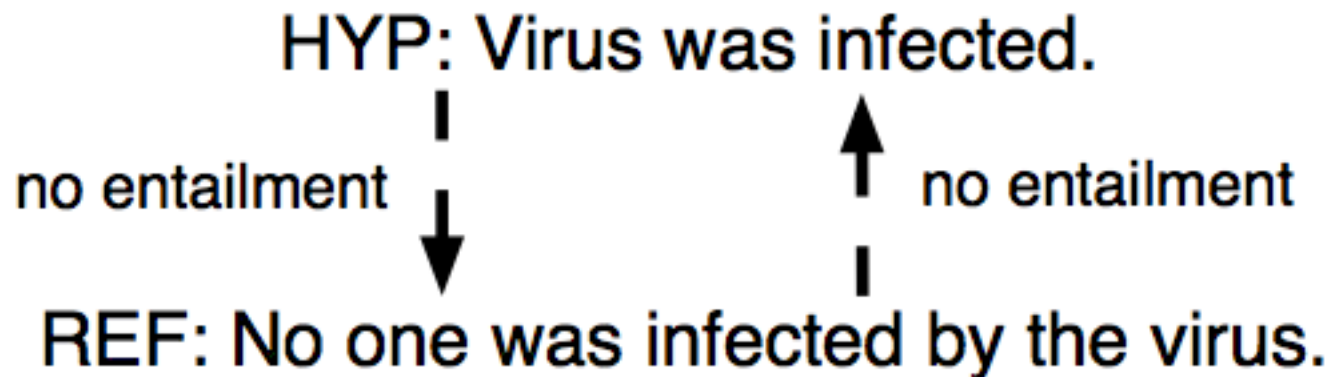
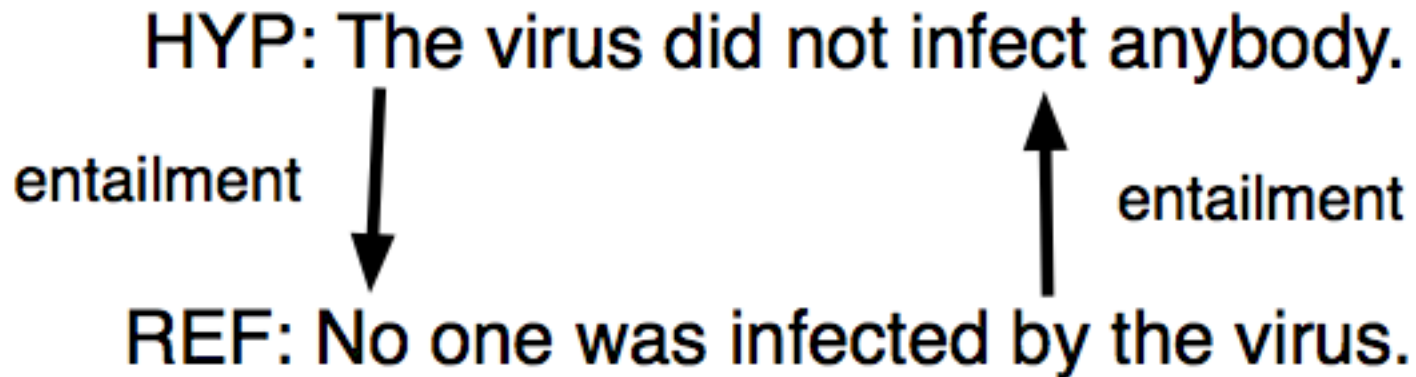
# Stanford Named Entity Tagger (NER)

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

## Stanford NER Output Format: slashTags

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

# Textual Entailment Features for Machine Translation Evaluation



# 自然語言處理與資訊檢索研究資源

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

淡江大學資訊管理學系

(Department of Information Management, Tamkang University)

自然語言處理與資訊檢索研究資源

(Resources of Natural Language Processing and Information Retrieval)

## 1. 中央研究院CKIP中文斷詞系統

授權單位：中央研究院詞庫小組

授權金額：免費授權學術使用。

授權日期：2011.03.31。

CKIP: <http://ckipsvr.iis.sinica.edu.tw/>

## 2. 「中央研究院中英雙語詞網」(The Academia Sinica Bilingual Wordnet)

「中央研究院中英雙語詞網」(The Academia Sinica Bilingual Wordnet)，

授權「淡江大學資訊管理學系」(Department of Information Management, Tamkang University)學術使用。

授權單位：中央研究院，中華民國計算語言學學會

授權金額：「中央研究院中英雙語詞網」(The Academia Sinica Bilingual Wordnet)

國內非營利機構(1-10人使用) 非會員：NT\$61,000元，

授權日期：2011.05.16。

Sinica BOW: <http://bow.ling.sinica.edu.tw/>

# 自然語言處理與資訊檢索研究資源

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

## 3. 開放式中研院專名問答系統 (OpenASQA)

授權單位：中央研究院資訊科學研究所智慧型代理人系統實驗室

授權金額：免費授權學術使用。

授權日期：2011.05.05。

ASQA: <http://asqa.iis.sinica.edu.tw/>

# 自然語言處理與資訊檢索研究資源

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

## 4. 哈工大資訊檢索研究中心(HIT-CIR)語言技術平臺

### 語料資源

哈工大資訊檢索研究中心漢語依存樹庫 [ HIT-CIR Chinese Dependency Treebank ]

哈工大資訊檢索研究中心同義詞詞林擴展版 [ HIT-CIR Tongyici Cilin (Extended) ]

### 語言處理模組

斷句 (SplitSentence: Sentence Splitting)

詞法分析 (IRLAS: Lexical Analysis System)

基於SVMTool的詞性標注 (PosTag: Part-of-speech Tagging)

命名實體識別 (NER: Named Entity Recognition)

基於動態局部優化的依存句法分析 (Parser: Dependency Parsing)

基於圖的依存句法分析 (GParser: Graph-based DP)

全文詞義消歧 (WSD: Word Sense Disambiguation)

淺層語義標注模組 (SRL: shallow Semantics Labeling)

### 資料表示

語言技術置標語言 (LTML: Language Technology Markup Language)

### 視覺化工具

LTML視覺化XSL

授權單位：哈工大資訊檢索研究中心(HIT-CIR)

授權金額：免費授權學術使用。

授權日期：2011.05.03。

HIT IR: <http://ir.hit.edu.cn/>

# Opinion Spam Detection



# Opinion Spam Detection

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

# Opinion Spamming

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

# Forms of Opinion spam

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

# Fake Review Detection

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

# Professional Fake Review Writing Services (some Reputation Management companies)

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

SponsoredReviews.com x

www.sponsoredreviews.com

 **SponsoredReviews.com**  
Bloggers Earn Cash, Advertisers Build Buzz!

[Members Login](#)

Sales Toll-Free (877) 360-3683

[Home](#) [Sign-Up](#) [FAQs](#) [Our Blog](#) [Contact Us](#)

SponsoredReviews connects bloggers with SEO's, Marketers, and Advertisers looking to build Links, Traffic and Buzz.

**Direct Traffic.**

Millions of people read blogs every day. Paying for posts puts the spotlight on your company and will generate tons of targeted traffic.

**Buzz & Branding.**

The more bloggers talk about your site the better. Many blogs syndicate stories they see on other sites. A couple well timed sponsored posts has the potential to generate a flurry of other post being written.

**Search Engine Rankings.**

Every post has links back to your site. Getting links from quality blogs will increase your link popularity and will help your site rank better in the search engines.

**Valuable Feedback.**

Getting Reviewed by bloggers will provide you with valuable feedback that you can use to better understand your audience and customers.

**Advertisers**  
Start Here.



- Announce your products, services, websites, and ideas to the world!
- Tap into the power of the blogosphere to build traffic, links and valuable feedback.

[Free Sign Up](#) [Read More](#)


**Bloggers**  
Earn Cash.



- Earn cash by writing honest posts about our advertiser's products and services.
- Write posts in your own tone and style, and gear them to your audience's interest.

[Free Sign Up](#) [Read More](#)


**How it works:**

 Advertiser  Blogger

PayPerPost : Blog Marke x  
https://payperpost.com

# payperpost


advertisers bloggers ethics about login



## advertisers

Hire bloggers to blog about your company, service or website. PayPerPost gives you access to a diverse pool of bloggers from all over the world. Make offers, negotiate deals and approve posts.

[signup now](#)




## bloggers

Make money blogging! PayPerPost lets you pick your advertisers, name your own price and negotiate your own deals. You can get paid to blog on virtually any subject. Sign up below!


[signup now](#)

### see how it works



[click here and watch the video](#)

### customer testimonial



"PayPerPost has been instrumental in helping our company streamline our various product awareness campaigns."  
-C. Litchfield

1 (877) 916 POST

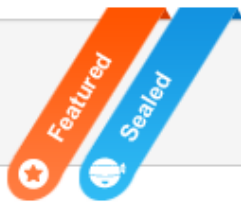


[Post Project](#)
[Find Freelancers](#)
[Browse Projects](#)
[Post Contest](#)

# Need someone to write and post positive reviews

Like 0 Send Tweet 0 +1 0 Share

Bids	Avg Bid (USD)	Project Budget (USD)
10	N/A	\$250 - \$750



CLOSED

### Project Description:

We need someone to write and post positive reviews about our company on websites. Please send an example of a review you would post for any company. We can also send examples of comments our customers have sent us to use and refer too as well

This is a long term project, so if it works out there will be a healthy amount of work. Please reply back with all your experience and how much you would charge per post.

thank you.

### Skills required:

Publicación en foros, Opiniones



### Project posted by:

dvel 5.0 (1 Review)

VERIFIED

Follow

Your ad could **From \$100/week**



# Papers on Opinion Spam Detection

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

# Summary

- Affective Computing and Social Computing
- Opinion Mining and Sentiment Analysis
- Social Media Monitoring/Analysis
- Resources of Opinion Mining
- Opinion Spam Detection

# References

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