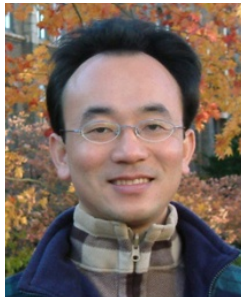


社群網路文字探勘與情感分析 (Text Mining and Sentiment Analytics on Social Media)

Time: 2016/05/10 (Tue) (12:00-14:00)

Place: 淡水校園覺生綜合大樓I501室



Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

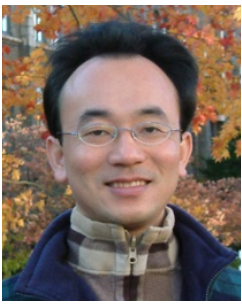
Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

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

2016-05-10





戴敏育 博士 (Min-Yuh Day, Ph.D.)

淡江大學資管系專任助理教授

中央研究院資訊科學研究所訪問學人

國立台灣大學資訊管理博士

Publications Co-Chairs, IEEE/ACM International Conference on
Advances in Social Networks Analysis and Mining (ASONAM 2013-)

Program Co-Chair, IEEE International Workshop on
Empirical Methods for Recognizing Inference in Text (IEEE EM-RITE 2012-)

Workshop Chair, The IEEE International Conference on
Information Reuse and Integration (IEEE IRI)



Outline

- 文字探勘技術
(Text Mining Technologies)
- 情感分析架構
(Architectures of Sentiment Analytics)

Data Scientist

資料科學家

What makes a data scientist?

The big data phenomenon trained a bright spotlight on those who perform deep information analysis and can combine quantitative and statistical modeling expertise with business acumen and a talent for finding hidden patterns. Here's a closer look.

Data scientists rely on analytics, predictive models, statistical analysis and modeling, data mining, sentiment and what-if analysis, and more to do their jobs. Cleansing raw data and building models is detailed work, and the right tools make the process much more efficient.

The IBM® BigInsights™ Data Scientist module accelerates data science with advanced analytics to extract valuable insights from Hadoop. Stable machine learning algorithms are optimized for Hadoop. Text analytics extract insight from unstructured data with existing tooling so analytic applications don't have to be developed from scratch. Big R statistical analysis and distributed frames allow data scientists to use the entire Hadoop cluster, not just a limited sample.

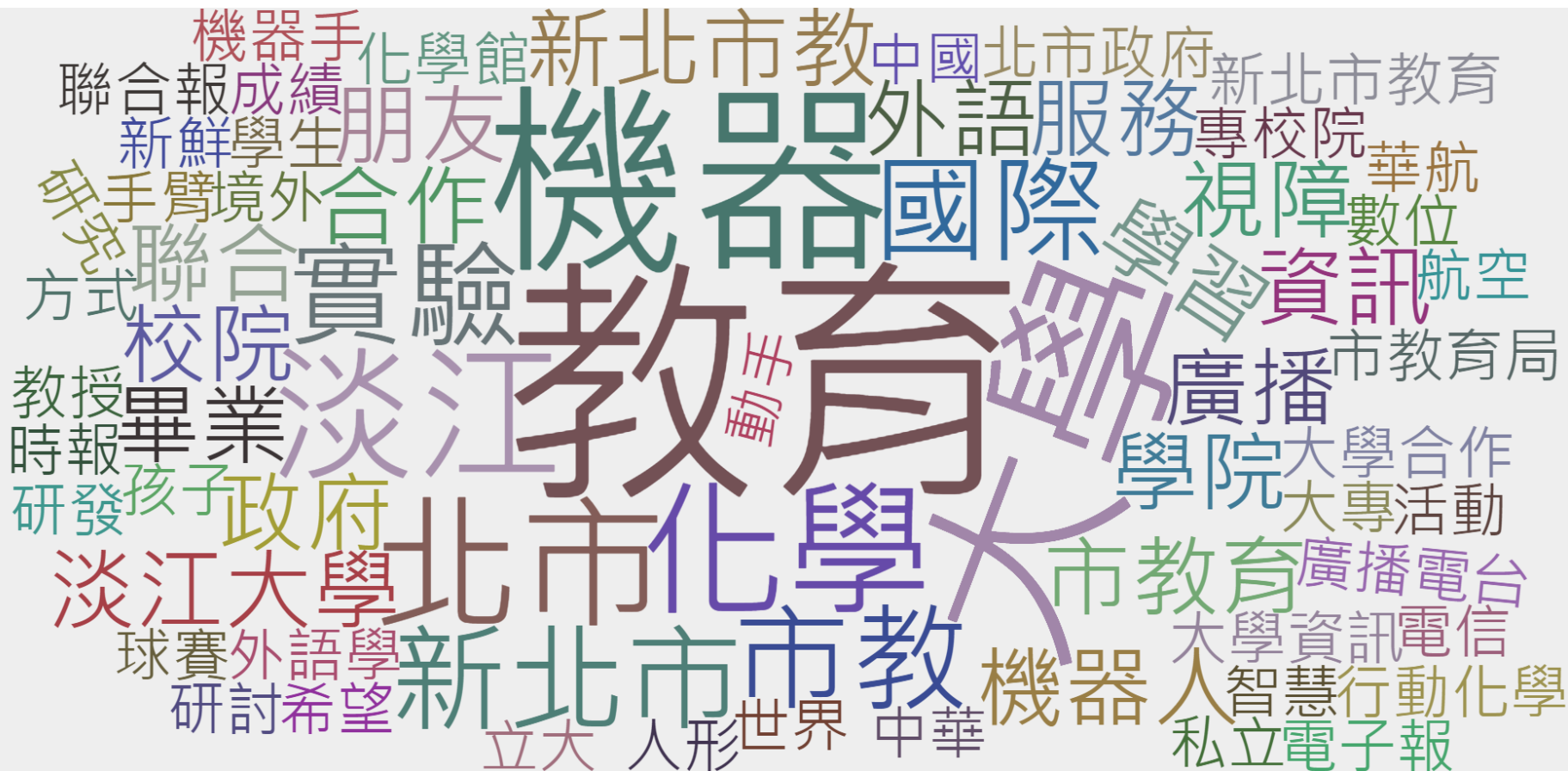


Good data scientists select and address the business problems that have the most value to the organization. Armed with data and analytical results, they must present their informed conclusions and recommendations to technical and nontechnical stakeholders.

The BigInsights Analyst module lets data scientists use their existing skills to find data across the organization and visualize it without extra coding. IBM BigSheets is a spreadsheet-style data manipulation and visualization tool that gives business users direct access to data through a recognizable interface. IBM-designed Big SQL offers HDFS caching and high availability benefits as well as query optimization—without forcing data scientists to learn a new skill set.

淡江大學
媒體報導文字雲

媒體報導文字雲



淡江大學媒體報導

TKU NEWS TODAY
淡江大
大小事

校園焦點
校園公告
媒體報導

RSS訂閱

淡江媒體報導

最新媒體報導

媒體	標題	時間
SSU大專學生運動網	榮譽心促使團結 淡江崛起奪金盃	2016/05/08
經濟日報	廣信益群 研討電商稅務	2016/05/05
工商時報	藝術文創－青春不設限 淡江大傳來「卅野」	2016/05/02
自由時報	檢測食用油 國中生實驗真油趣	2016/04/28
國語日報	分辨好油壞油 新北30校師生做實驗	2016/04/28
中國時報	拗客動物！大學生製動畫妙喻心聲	2016/04/28
真晨報(東方日報)	新北動手做科遊 認識食用油安全	2016/04/28
動腦雜誌Brain	青春不設限 淡江大傳來「卅野」	2016/04/27
自立晚報	新北五科星 全民一起玩科學 千人做實驗	2016/04/27
中國時報	聽障朋友看得見聲音	2016/04/27
中國時報	淡大生設計手語單字貼圖譜曲 聽障朋友看得見聲音	2016/04/27

媒體搜尋

起始日 2016-04-09

結束日 2016-06-09

媒體

關鍵字
查詢

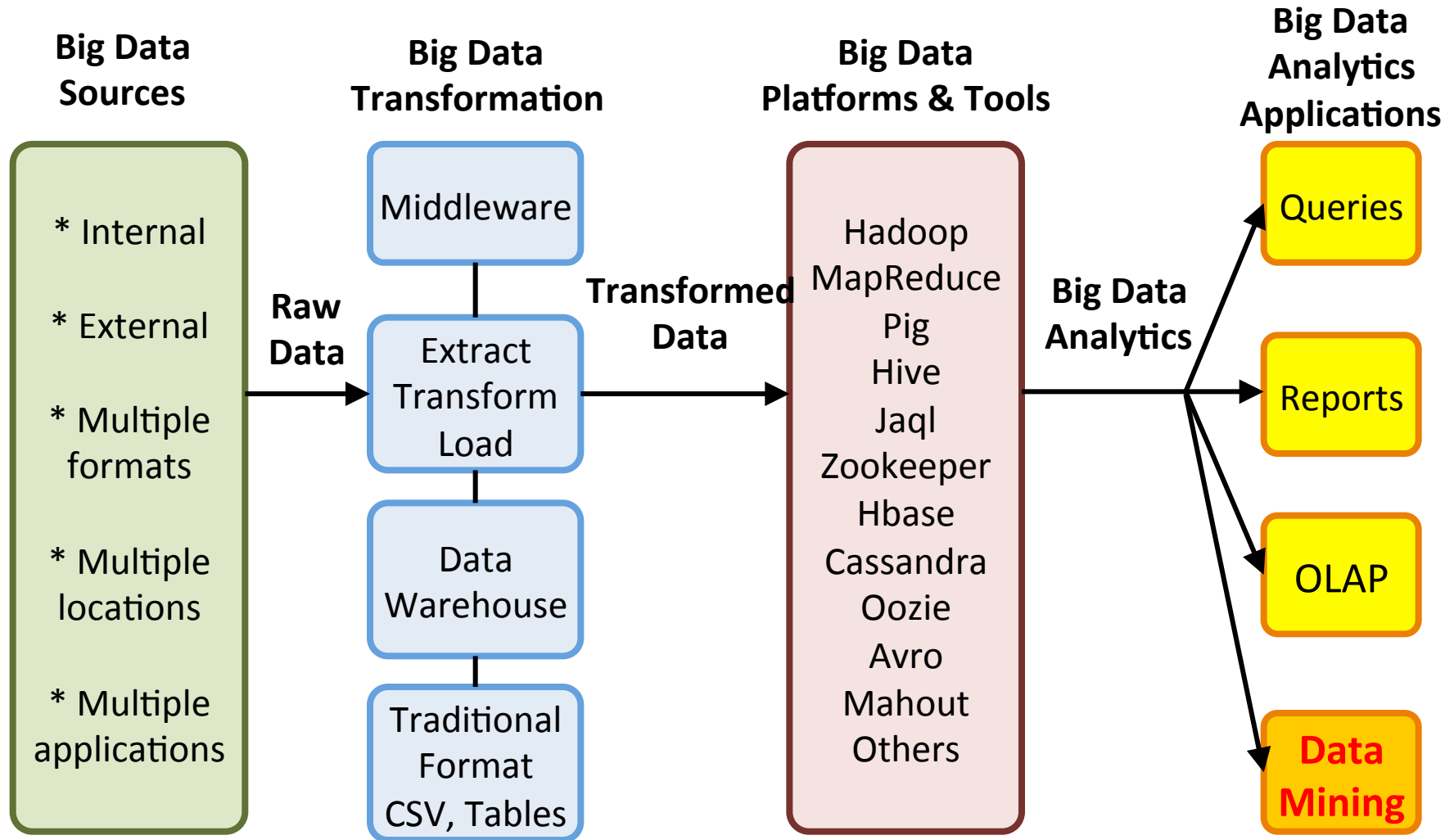
最近媒體

- 榮譽心促使團結 淡江崛起奪金盃
- 廣信益群 研討電商稅務
- 藝術文創－青春不設限 淡江大傳來「卅野」
- 檢測食用油 國中生實驗真油趣
- 分辨好油壞油 新北30校師生做實驗
- 拗客動物！大學生製動畫妙喻心聲
- 新北動手做科遊 認識食用油安全
- 青春不設限 淡江大傳來「卅野」
- 新北五科星 全民一起玩科學 千人做實驗
- 聽障朋友看得見聲音

淡江大學媒體報導關鍵字

- 淡江大學 共出現1173次
- 化學 共出現749次
- 台灣 共出現740次
- 教育 共出現606次
- 活動 共出現575次
- 學校 共出現550次
- 淡大 共出現522次
- 國際 共出現465次

Architecture of Big Data Analytics

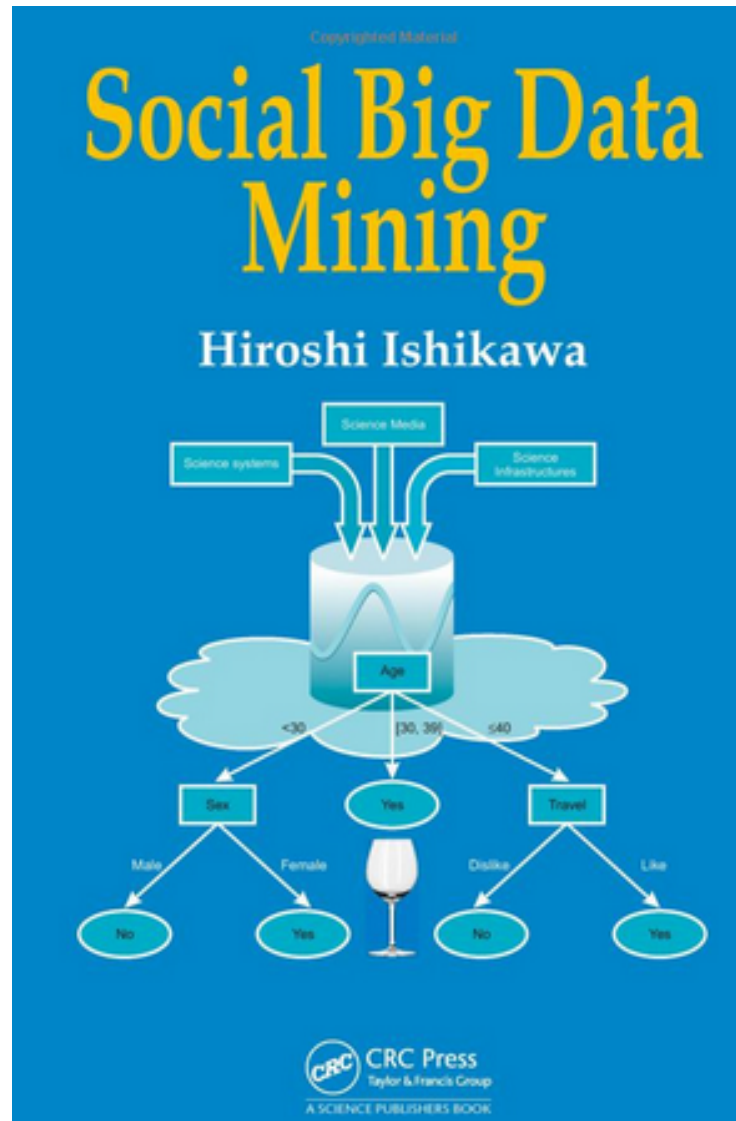


Architecture of Big Data Analytics



Social Big Data Mining

(Hiroshi Ishikawa, 2015)



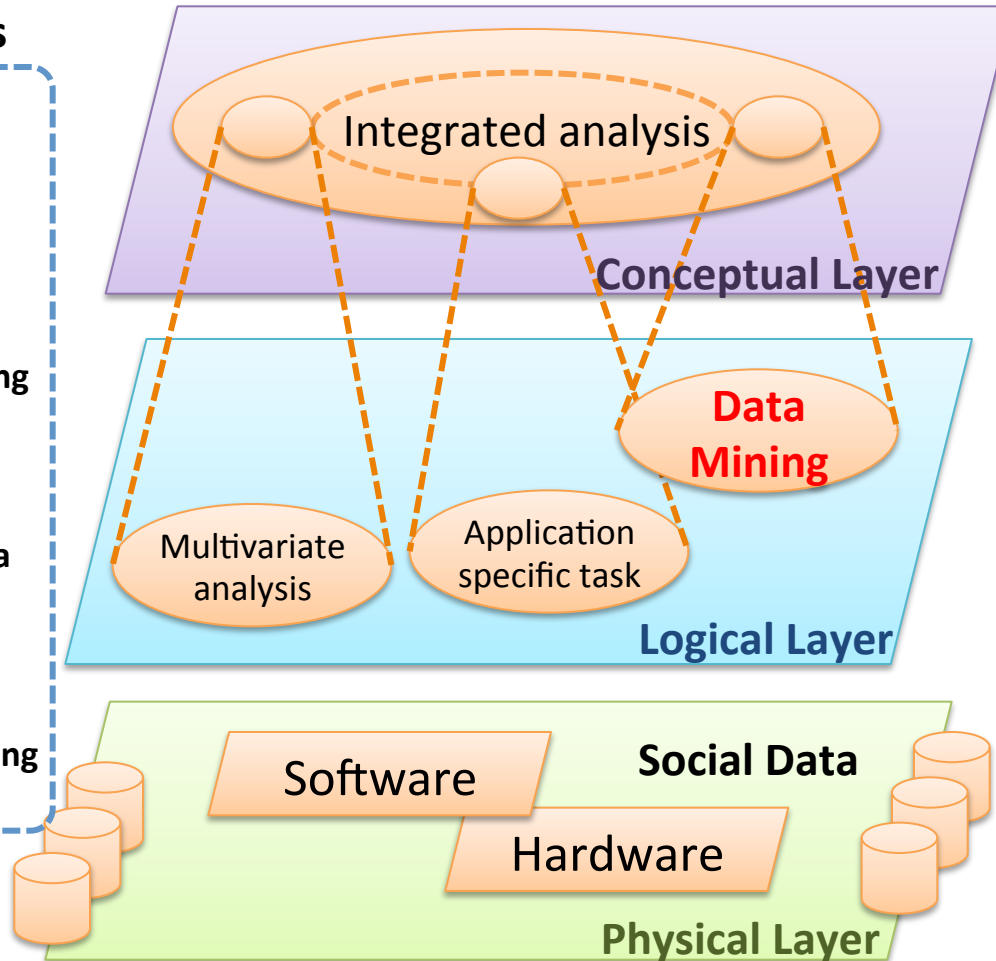
Source: <http://www.amazon.com/Social-Data-Mining-Hiroshi-Ishikawa/dp/149871093X>

Architecture for Social Big Data Mining

(Hiroshi Ishikawa, 2015)

Enabling Technologies

- Integrated analysis model
- Natural Language Processing
- Information Extraction
- Anomaly Detection
- Discovery of relationships among heterogeneous data
- Large-scale visualization
- Parallel distrusted processing



Analysts

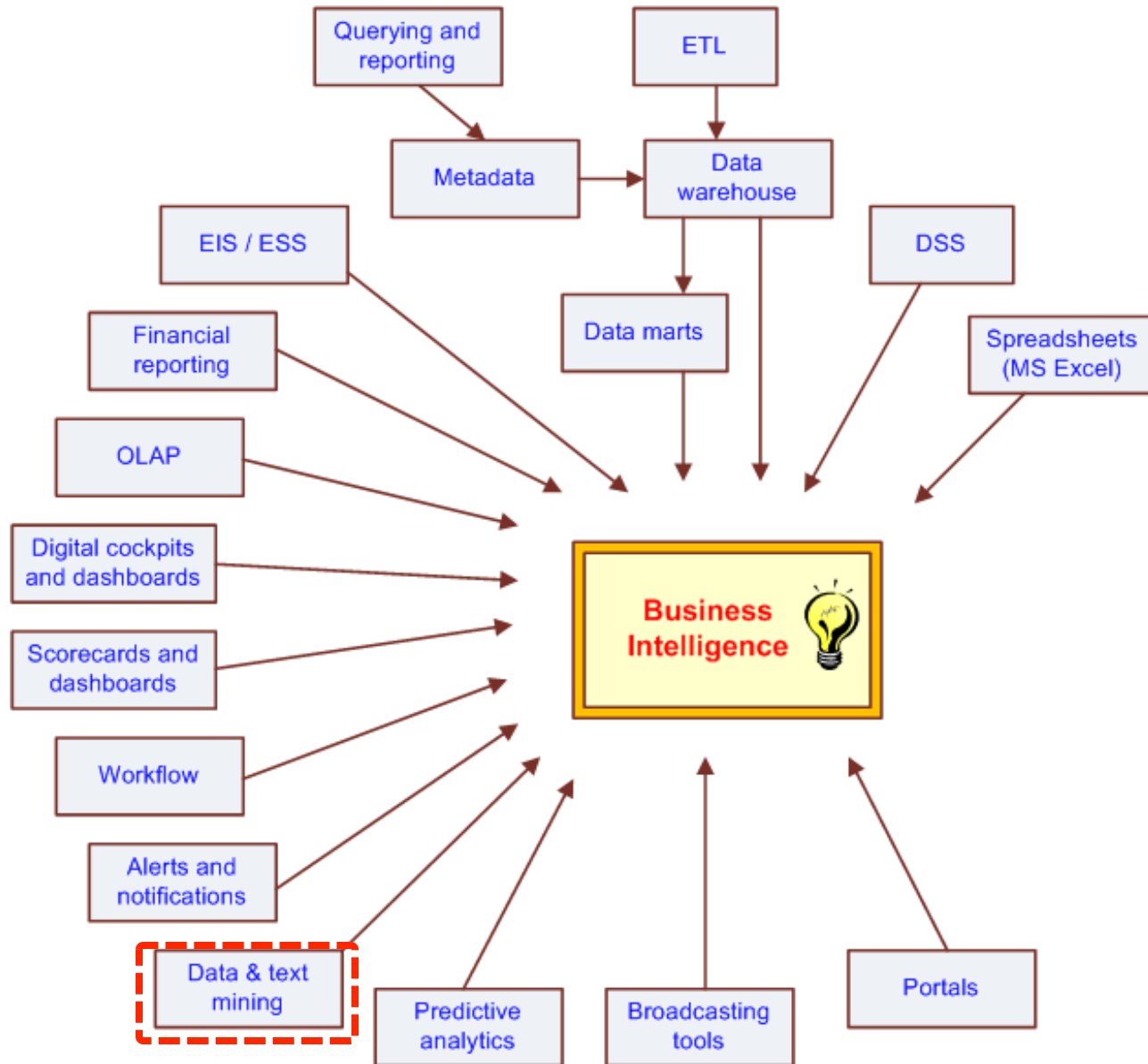
- Model Construction
- Explanation by Model
- Construction and confirmation of individual hypothesis
- Description and execution of application-specific task

Deep Learning

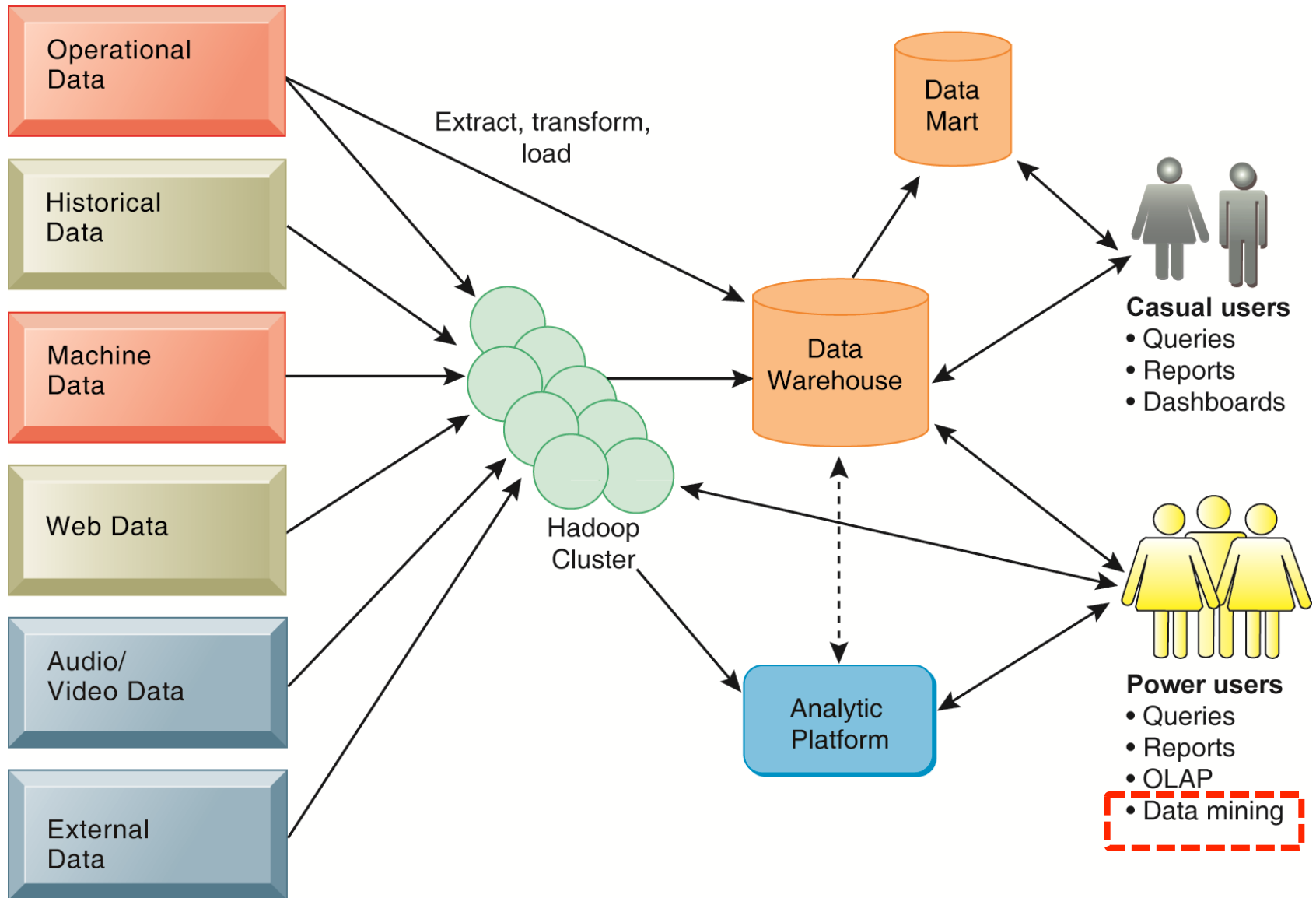
Intelligence from Big Data



The Evolution of BI Capabilities



Business Intelligence (BI) Infrastructure



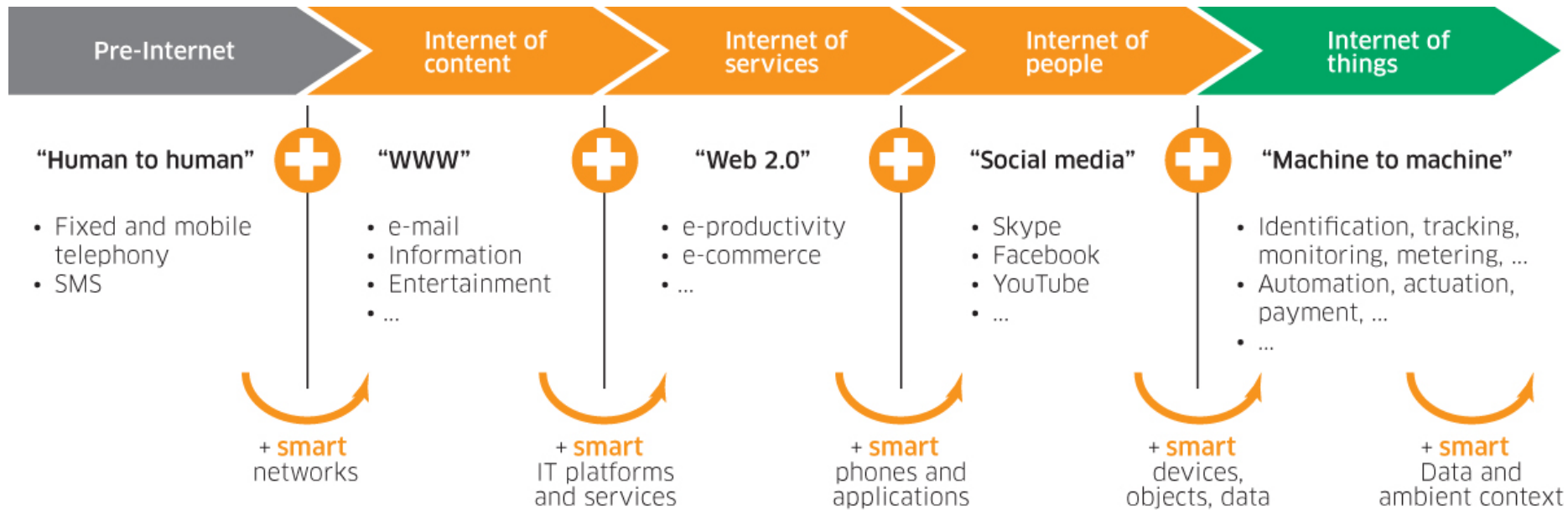
Social Media



Internet Evolution

Internet of People (IoP): Social Media

Internet of Things (IoT): Machine to Machine



Source: Marc Jadoul (2015), The IoT: The next step in internet evolution, March 11, 2015

<http://www2.alcatel-lucent.com/techzine/iot-internet-of-things-next-step-evolution/>

Emotions



Love

Anger

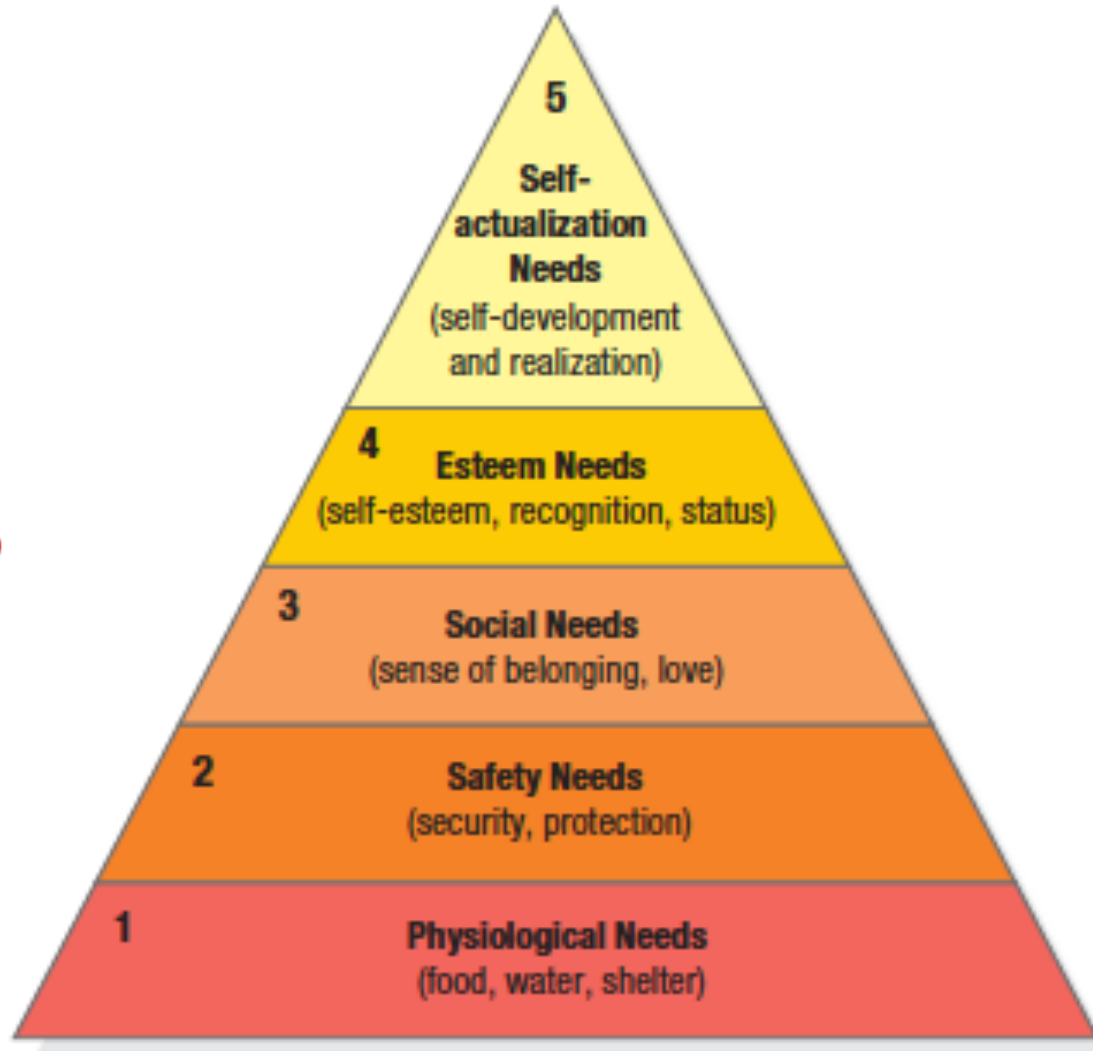
Joy

Sadness

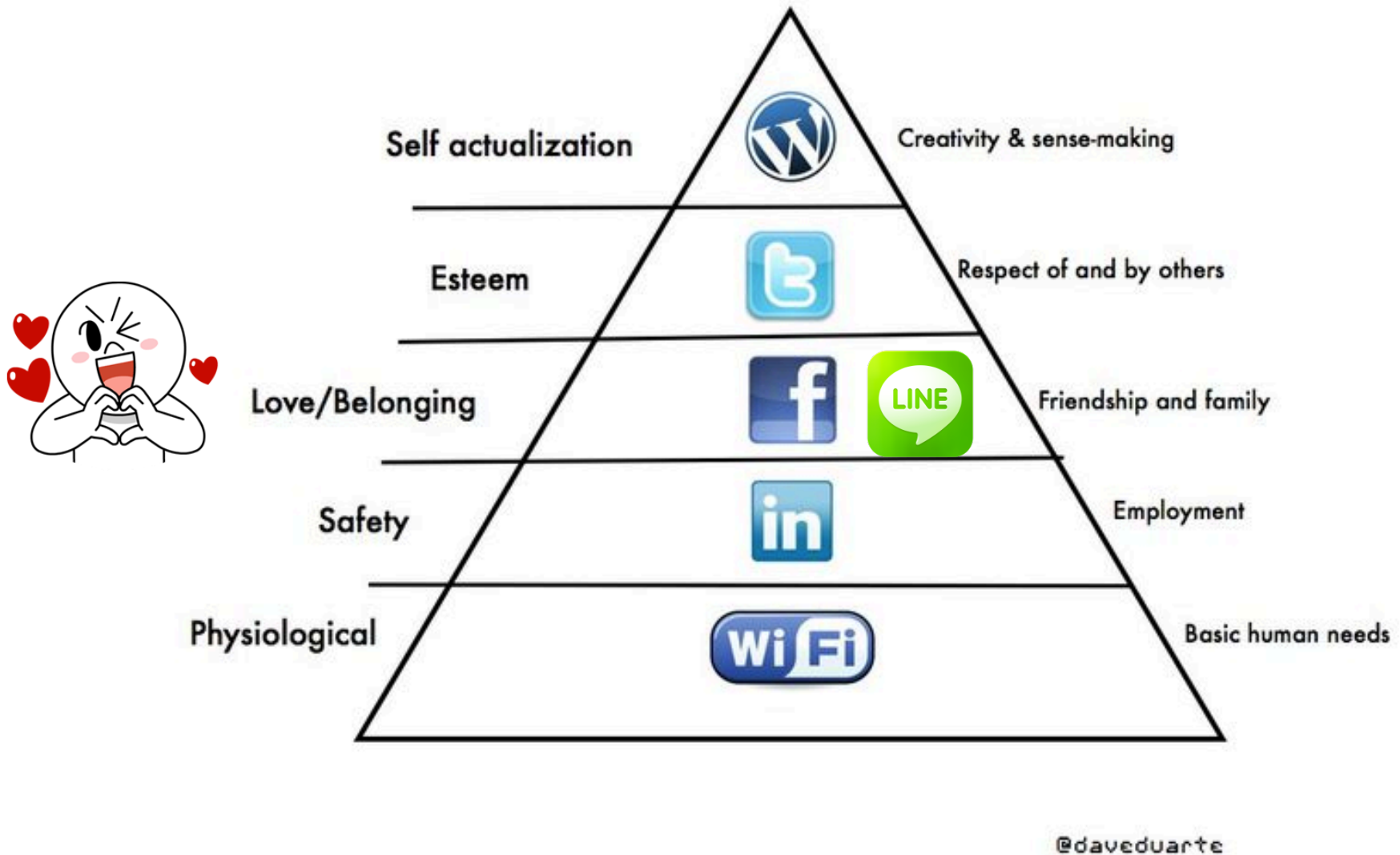
Surprise

Fear

Maslow's Hierarchy of Needs

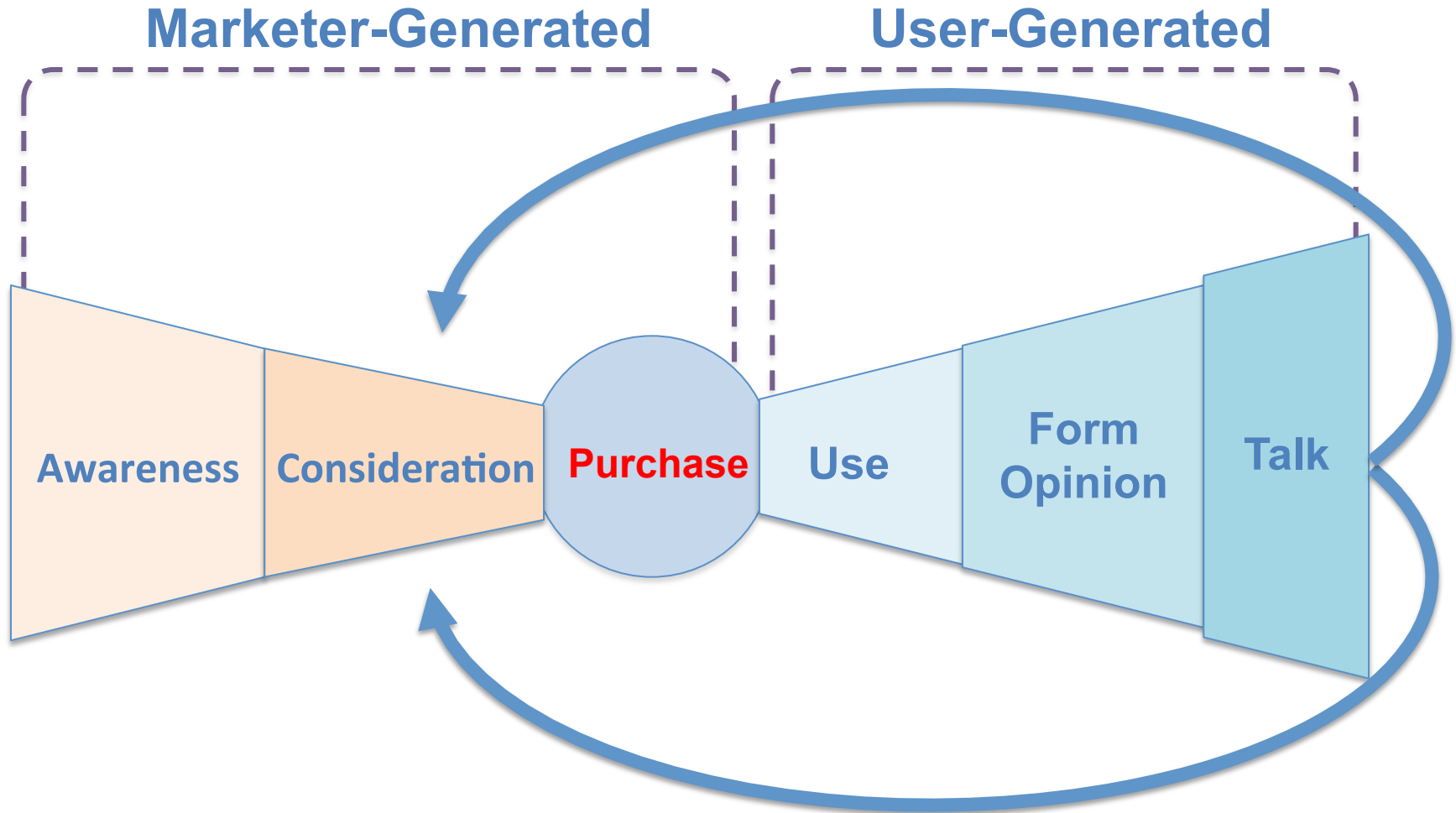


Social Media Hierarchy of Needs

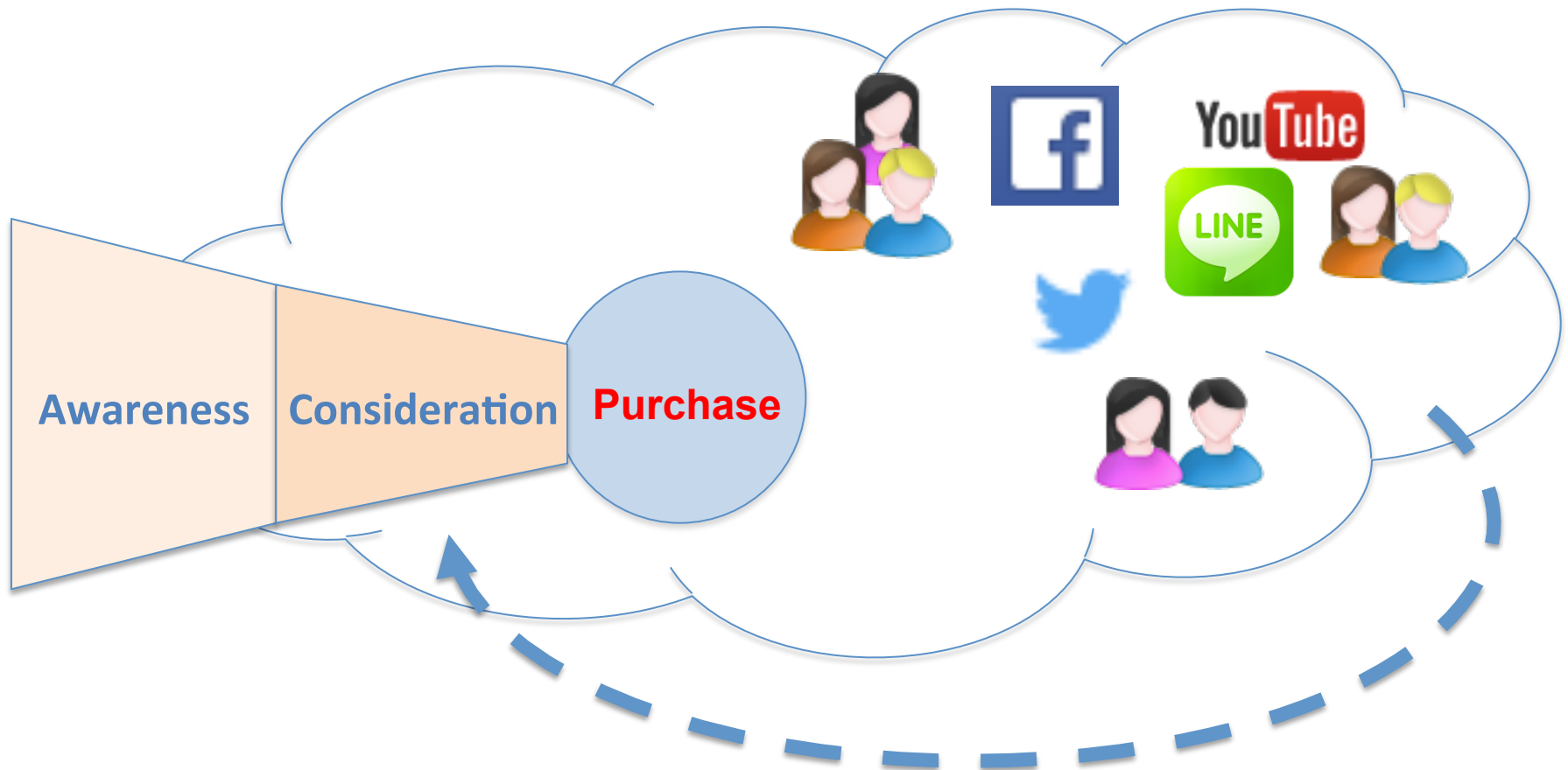


The Social Feedback Cycle

Consumer Behavior on Social Media



The New Customer Influence Path





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**.

(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. ...”



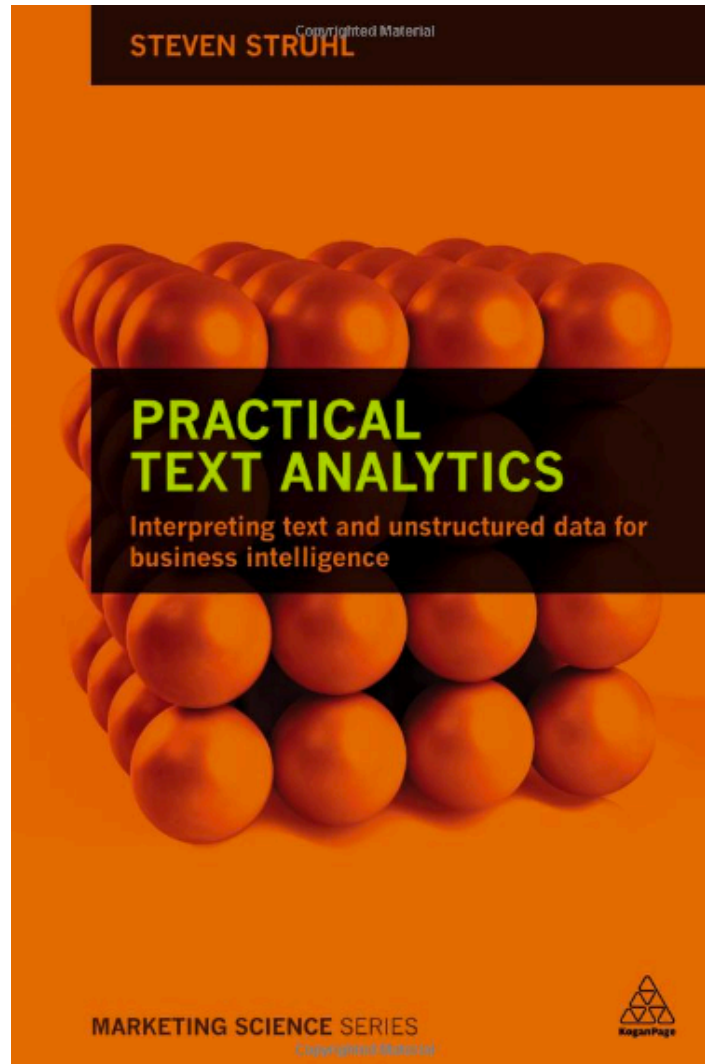
**+Positive
Opinion**



**-Negative
Opinion**

Text Mining Technologies

**Steven Struhl (2015),
Practical Text Analytics:
Interpreting Text and Unstructured Data for Business Intelligence
(Marketing Science), Kogan Page**



Text Mining Concepts

- 85-90 percent of all corporate data is in some kind of unstructured form (e.g., text)
- Unstructured corporate data is doubling in size every 18 months
- Tapping into these information sources is not an option, but a need to stay competitive
- Answer: text mining
 - A semi-automated process of extracting knowledge from unstructured data sources
 - a.k.a. text data mining or knowledge discovery in textual databases

Text mining

Text Data Mining

Intelligent Text Analysis

Knowledge-Discovery in Text (KDT)

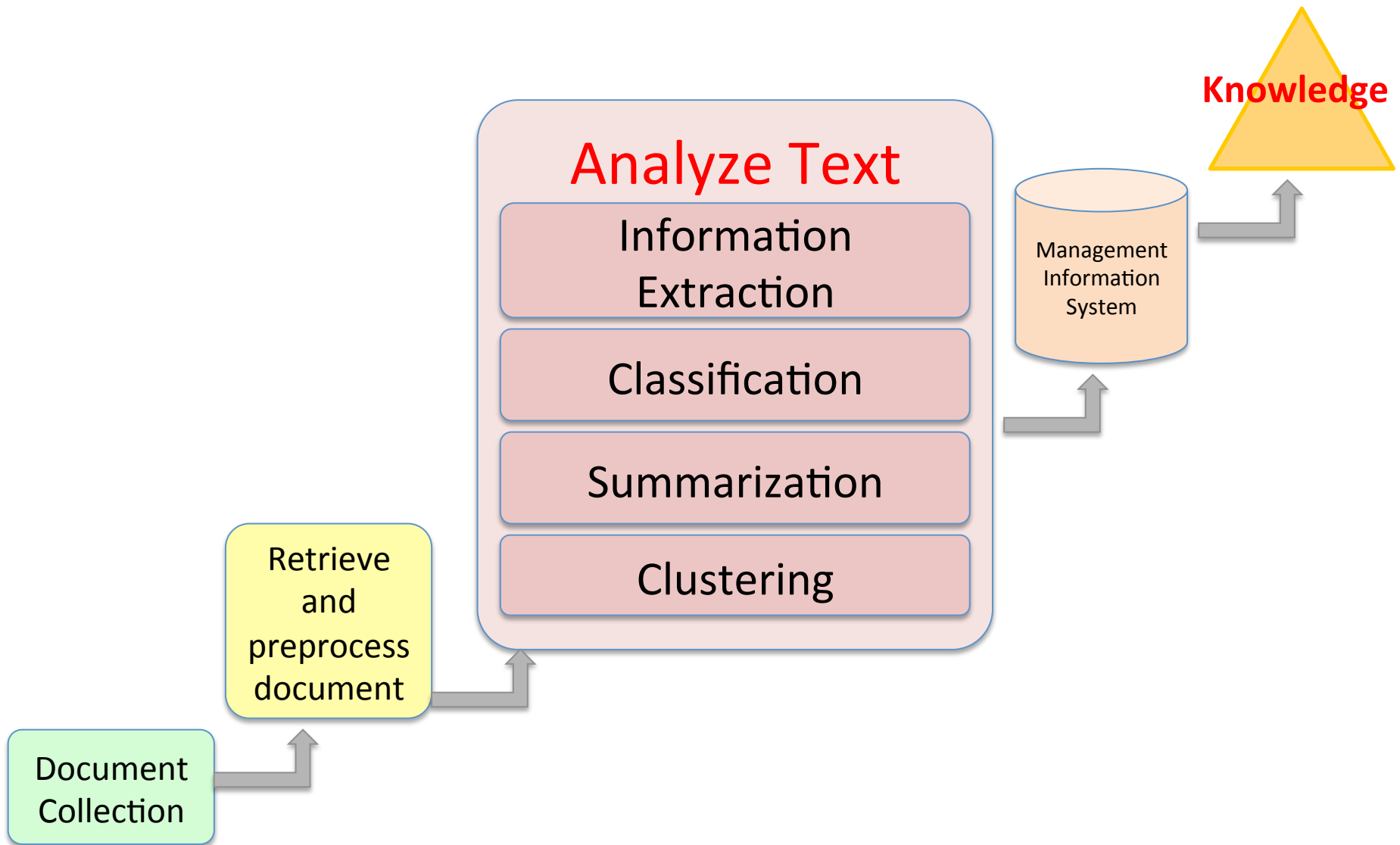
Text Mining:
the process of extracting
interesting and non-trivial
information and knowledge
from unstructured text.

Text Mining:
discovery by computer of
new, previously
unknown information,
by automatically
extracting information
from different written resources.

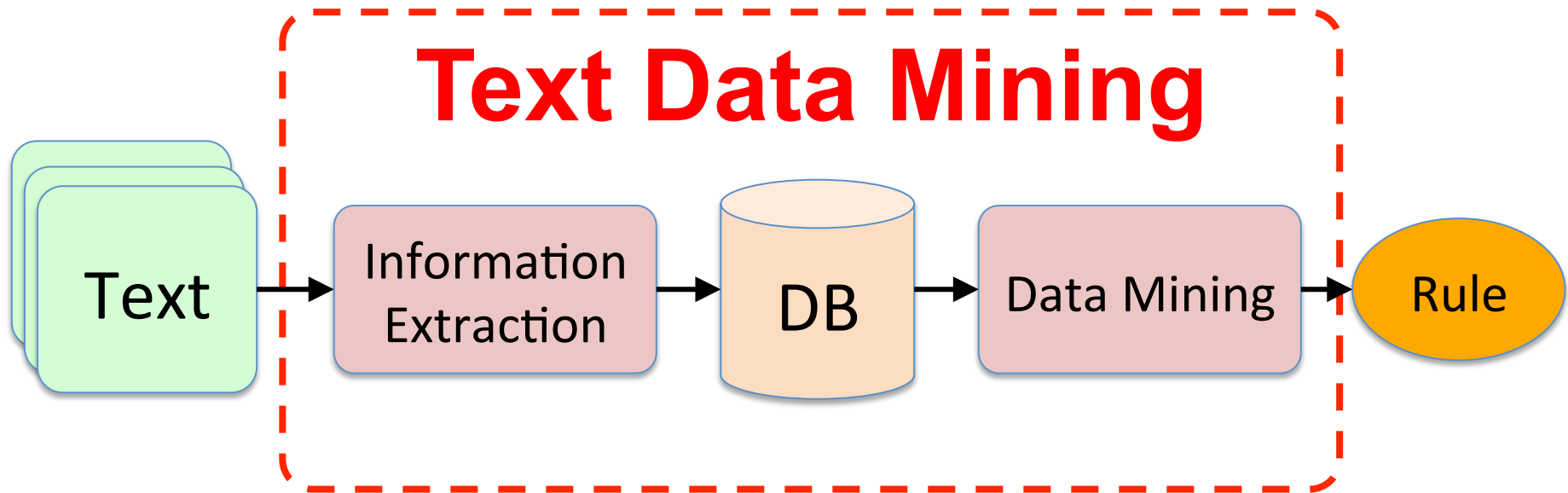
Text Mining (TM)

**Natural Language Processing
(NLP)**

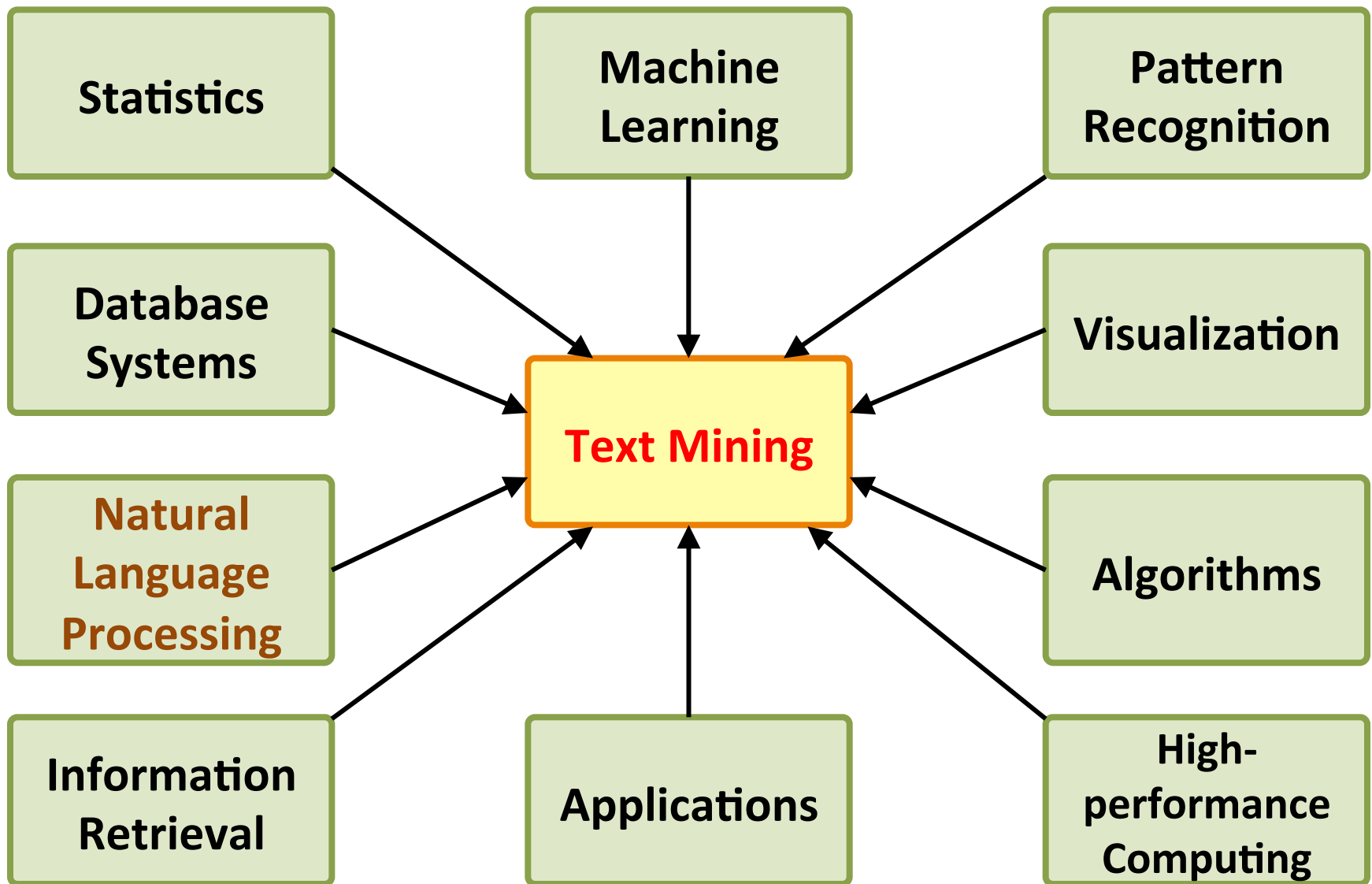
An example of Text Mining



Overview of Information Extraction based Text Mining Framework



Text Mining Technologies



Data Mining versus Text Mining

- Both seek for novel and useful patterns
- Both are semi-automated processes
- Difference is the nature of the data:
 - Structured versus unstructured data
 - **Structured data:** in databases
 - **Unstructured data:** Word documents, PDF files, text excerpts, XML files, and so on
- Text mining – first, impose structure to the data, then mine the structured data



Data Mining:

Core **Analytics** Process

The **KDD** Process for
Extracting Useful **Knowledge**
from Volumes of **Data**

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996).

The **KDD Process** for Extracting Useful **Knowledge** from Volumes of **Data**.

Communications of the ACM, 39(11), 27-34.

Knowledge Discovery in Databases creates the context for developing the tools needed to control the flood of data facing organizations that depend on ever-growing databases of business, manufacturing, scientific, and personal information.

The KDD Process for Extracting Useful Knowledge from Volumes of Data

AS WE MARCH INTO THE AGE of digital information, the problem of data overload looms ominously ahead. Our ability to analyze and understand massive datasets lags far behind our ability to gather and store the data. A new generation of computational techniques and tools is required to support the extraction of useful knowledge from the rapidly growing volumes of data. These techniques and tools are the subject of the emerging field of knowledge discovery in databases (KDD) and data mining.

Large databases of digital information are ubiquitous. Data from the neighborhood store's checkout register, your bank's credit card authorization device, records in your doctor's office, patterns in your telephone calls,

Usama Fayyad,

Gregory Piatetsky-Shapiro,

and Padhraic Smyth

and many more applications generate streams of digital records archived in huge databases, sometimes in so-called data warehouses.

Current hardware and database technology allow efficient and inexpensive reliable data storage and access. However, whether the context is business, medicine, science, or government, the datasets themselves (in raw form) are of little direct value. What is of value is the knowledge that can be inferred from the data and put to use. For example, the marketing database of a consumer

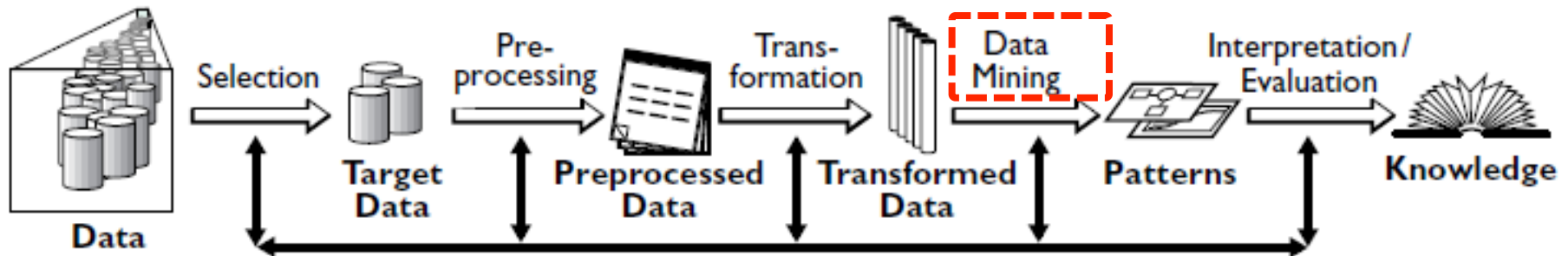


USAMA FAYYAD

Data Mining

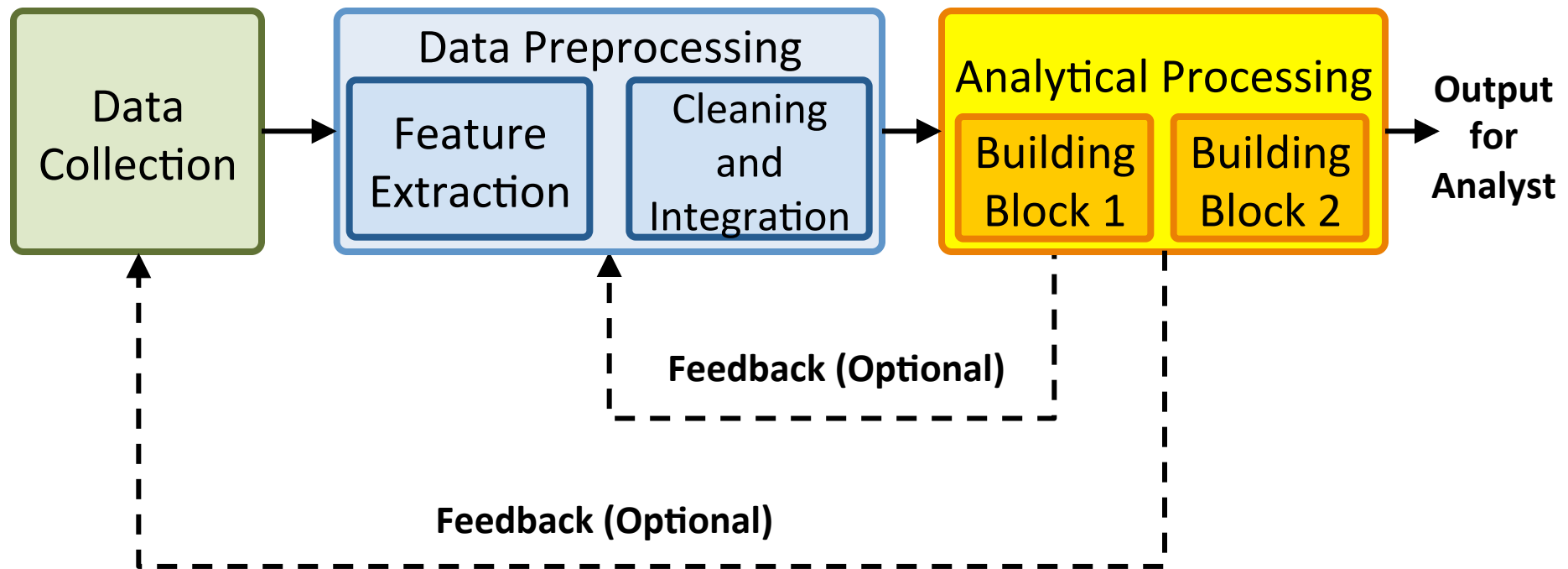
Knowledge Discovery in Databases (KDD) Process

(Fayyad et al., 1996)



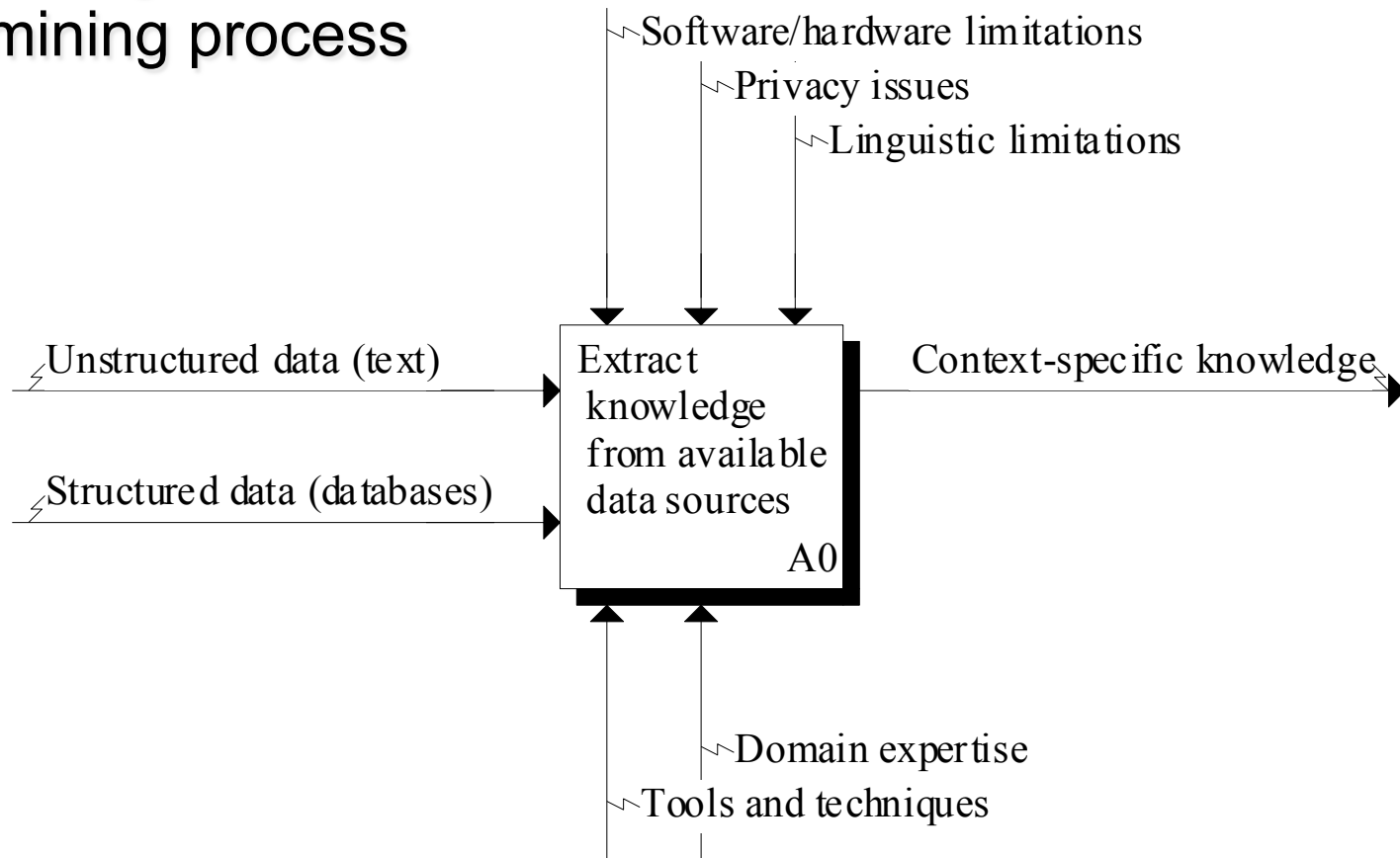
Data Mining Processing Pipeline

(Charu Aggarwal, 2015)

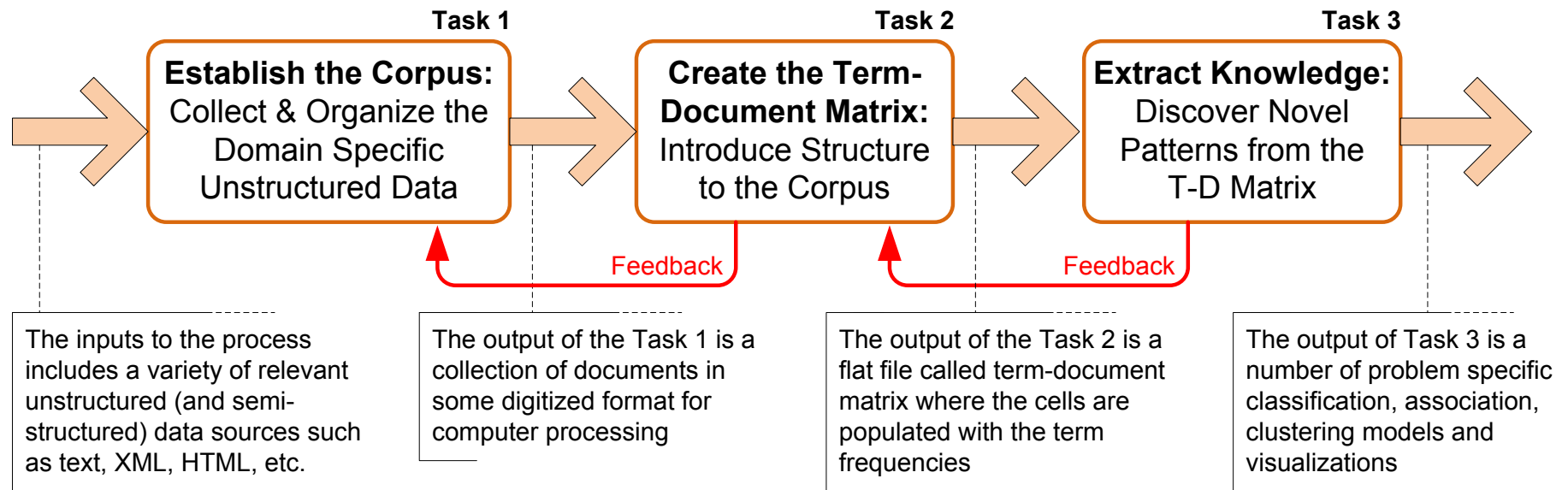


Text Mining Process

Context diagram for the text mining process



Text Mining Process



The three-step text mining process

Text Mining Process

- **Step 1:** Establish the corpus
 - Collect all relevant unstructured data (e.g., textual documents, XML files, emails, Web pages, short notes, voice recordings...)
 - Digitize, standardize the collection (e.g., all in ASCII text files)
 - Place the collection in a common place (e.g., in a flat file, or in a directory as separate files)

Text Mining Process

- **Step 2:** Create the Term-by-Document Matrix

<div>Terms</div> <div>Documents</div>	investment risk	project management	software engineering	development	SAP	...
Document 1	1			1		
Document 2		1				
Document 3			3		1	
Document 4		1				
Document 5			2	1		
Document 6	1			1		
...						

Text Mining Process

- **Step 2:** Create the Term-by-Documents Matrix (TDM), cont.
 - Should all terms be included?
 - Stop words, include words
 - Synonyms, homonyms
 - Stemming
 - What is the best representation of the indices (values in cells)?
 - Row counts; binary frequencies; log frequencies;
 - Inverse document frequency

Text Mining Process

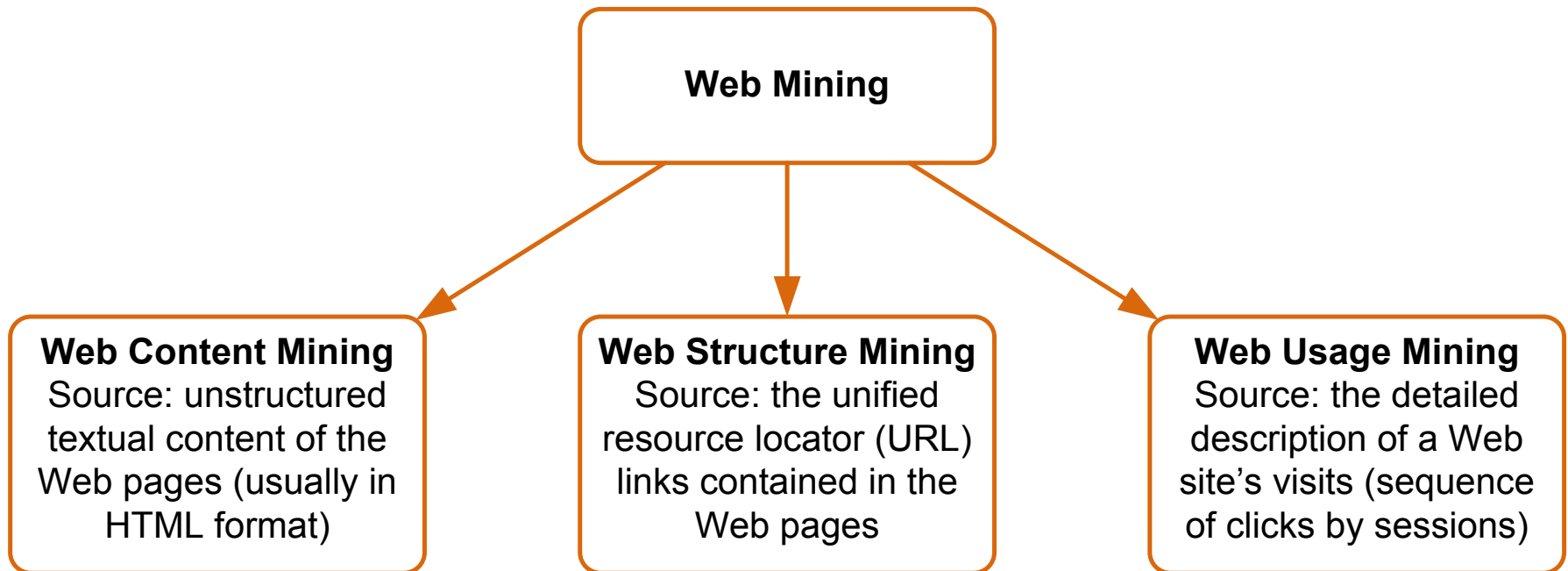
- **Step 2:** Create the Term-by-Document Matrix (TDM), cont.
 - TDM is a sparse matrix. How can we reduce the dimensionality of the TDM?
 - Manual - a domain expert goes through it
 - Eliminate terms with very few occurrences in very few documents (?)
 - Transform the matrix using singular value decomposition (SVD)
 - SVD is similar to principle component analysis

Text Mining Process

- **Step 3:** Extract patterns/knowledge
 - Classification (text categorization)
 - Clustering (natural groupings of text)
 - Improve search recall
 - Improve search precision
 - Scatter/gather
 - Query-specific clustering
 - Association
 - Trend Analysis (...)

Web Mining

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



Text Mining Concepts

- Benefits of text mining are obvious especially in text-rich data environments
 - e.g., law (court orders), academic research (research articles), finance (quarterly reports), medicine (discharge summaries), biology (molecular interactions), technology (patent files), marketing (customer comments), etc.
- Electronic communication records (e.g., Email)
 - Spam filtering
 - Email prioritization and categorization
 - Automatic response generation

Text Mining Application Area

- Information extraction
- Topic tracking
- Summarization
- Categorization
- Clustering
- Concept linking
- Question answering

Text Mining Terminology

- Unstructured or semistructured data
- Corpus (and corpora)
- Terms
- Concepts
- Stemming
- Stop words (and include words)
- Synonyms (and polysemes)
- Tokenizing

Text Mining Terminology

- Term dictionary
- Word frequency
- Part-of-speech tagging (POS)
- Morphology
- Term-by-document matrix (TDM)
 - Occurrence matrix
- Singular Value Decomposition (SVD)
 - Latent Semantic Indexing (LSI)

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

Natural Language Processing (NLP)

- What is “Understanding” ?
 - Human understands, what about computers?
 - Natural language is vague, context driven
 - True understanding requires extensive knowledge of a topic
 - Can/will computers ever understand natural language the same/accurate way we do?

Natural Language Processing (NLP)

- Challenges in NLP
 - Part-of-speech tagging
 - Text segmentation
 - Word sense disambiguation
 - Syntax ambiguity
 - Imperfect or irregular input
 - Speech acts
- Dream of AI community
 - to have algorithms that are capable of automatically reading and obtaining knowledge from text

Natural Language Processing (NLP)

- WordNet
 - A laboriously hand-coded database of English words, their definitions, sets of synonyms, and various semantic relations between synonym sets
 - A major resource for NLP
 - Need automation to be completed
- Sentiment Analysis
 - A technique used to detect favorable and unfavorable opinions toward specific products and services
 - CRM application

NLP Task Categories

- Information retrieval (IR)
- Information extraction (IE)
- Named-entity recognition (NER)
- Question answering (QA)
- Automatic summarization
- Natural language generation and understanding (NLU)
- Machine translation (ML)
- Foreign language reading and writing
- Speech recognition
- Text proofing
- Optical character recognition (OCR)

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](#)

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 28270134 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](#)

[隱私權聲明](#) | [版權聲明](#)



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

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

即時 要聞 娛樂 運動 全球 社會 產經 股市 健康 生活 文教 評論 地方 兩岸

莎士比亞在淡江 遇見賽萬提斯

f 分享

G+ 分享

留言

列印

存新聞

A- A+

2016-04-26 02:27 聯合報 記者徐葳倫／淡水報導

f 讚

分享

20

傳送

G+1

0



淡江大學舉辦「當莎士比亞遇見賽萬提斯」系列活動，讓師生幫莎士比亞、賽萬提斯著色，畫出五彩繽紛的「文學大師」。記者徐葳倫／攝影

4月23日是「世界閱讀日」，也是英國大文豪莎士比亞的生日與忌日，及「唐吉訶德」作

莎士比亞在淡江 遇見賽萬提斯

2016-04-26 02:27 聯合報 記者徐葳倫／淡水報導

分享4月23日是「世界閱讀日」，也是英國大文豪莎士比亞的生日與忌日，及「唐吉訶德」作者賽萬提斯逝世之日。英專起家的淡江大學舉辦「當莎士比亞遇見賽萬提斯」活動，規畫主題書展、彩繪活動，並添購新書，拉近學生與經典文學的距離。

首波登場的「主題書展」，展出2大文豪經典作品的原著、各種譯本以及DVD、電子書等數位化資料，校方也添購許多新書，吸引學生「搶鮮」閱讀經典名作。現場還規畫「彩繪大師」，讓學生發揮創意，畫出五彩繽紛的莎士比亞和賽萬提斯人像。英語系四年級學生陳彥伶說，讀英語系接觸莎士比亞作品，但過去沒有舉辦書展時，這些作品都放在圖書館8樓，現在搬到1樓大廳陳列，不僅有很多莎士比亞、賽萬提斯的經典新書，還可藉由電子書、電影理解兩位作家，是以前沒有過的體驗。

英語系四年級學生鄭少淮表示，莎士比亞的「馬克白」、「羅密歐與茱麗葉」都已經讀過很多次，從經典文學中理解不同城市、國家的文化。

日文系學生賴喬郁說，原本只是喜歡塗鴉才來參加活動，後來才知道畫的是2個大文豪，接觸他們的作品，文學經典「原來離我這麼近」。

淡江大學外語學院院長陳小雀表示，莎士比亞的「to be, or not to be; that is the question」，賽萬提斯的「看得越多，行得越遠；書讀得越多，知識就越廣博」，都是來自文學的名言，校方希望用最簡單的方式，讓學生知道「文學不難」，就在你我身邊。

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.

自 2014/01/06 起，本斷詞系統已經處理過 28270134 篇文章

送出

清除

莎士比亞在淡江 遇見賽萬提斯

2016-04-26 02:27 聯合報 記者徐葳倫 / 淡水報導

分享4月23日是「世界閱讀日」，也是英國大文豪莎士比亞的生日與忌日，及「唐吉訶德」作者賽萬提斯逝世之日。英專起家的淡江大學舉辦「當莎士比亞遇見賽萬提斯」活動，規畫主題書展、彩繪活動，並添購新書，拉近學生與經典文學的距離。

首波登場的「主題書展」，展出2大文豪經典作品的原著、各種譯本以及DVD、電子書等數位化資料，校方也添購許多新書，吸引學生「搶鮮」閱讀經典名作。現場還規畫「彩繪大師」，讓學生發揮創意，畫出五彩繽紛的莎士比亞和賽萬提斯人像。英語系四年級學生陳彥伶說，讀英語系接觸莎士比亞作品，但過去沒有舉辦書展時，這些作品都放在圖書館8樓，現在搬到1樓大廳陳列，不僅有很多莎士比亞、賽萬提斯的經典新書，還可藉由電子書、電影理解兩位作家，是以前沒有過的體驗。

英語系四年級學生鄭少淮表示，莎士比亞的「馬克白」、「羅密歐與茱麗葉」都已經讀過很多次，從經典文學中理解不同城市、國家的文化。

日文系學生賴喬郁說，原本只是喜歡塗鴉才來參加活動，後來才知道畫的是2個大文豪，接觸他們的作品，文學經典「原來離我這麼近」。

淡江大學外語學院院長陳小雀表示，莎士比亞的「to be, or not to be; that is the question」，賽萬提斯的「看得越多，行得越遠；書讀得越多，知識就越廣博」，都是來自文學的名言，校方希望用最簡單的方式，讓學生知道「文學不難」，就在你我身邊。

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.

莎士比亞(Nb) 在(P) 淡江(Nb) 遇見(Vc) 賽萬提(Nb) 斯(Nep) 2016(Neu) -(FW) 04(Neu) -(FW) 2602(Neu) :(COLONCATEGORY)
27(Neu) 聯合報(Nb) 記者(Na) 徐葳倫(Nb) 淡水(Nc) 報導(Na) 分享(VJ) 4月(Nd) 23日(Nd) 是(SHI) 「(PARENTHESISCATEGORY)
也(D) 是(SHI) 英國(Nc) 大(VH) 文豪(Na) 莎士比亞(Nb) 的(DE) 生日(Na) 與(Caa) 忌日(Na) ,(COMMACATEGORY)
及(Caa) 「(PARENTHESISCATEGORY) 唐吉訶德(Nb) 」(PARENTHESISCATEGORY) 作者(Na) 賽萬提(Nb) 斯(Nep) 逝世(VH) 之(DE) 日(Na)
英(Nc) 專(D) 起家(VA) 的(DE) 淡江(Nb) 大學(Nc) 舉辦(Vc) 「(PARENTHESISCATEGORY) 當(P) 莎士比亞(Nb) 遇見(Vc) 賽萬提(Nb) ！
規畫(Vc) 主題(Na) 書展(Na) 、(PAUSECATEGORY) 彩繪(Vc) 活動(Na) ,(COMMACATEGORY)
並(Cbb) 添購(Vc) 新書(Na) ,(COMMACATEGORY)
拉近(Vc) 學生(Na) 與(Caa) 經典(Na) 文學(Na) 的(DE) 距離(Na) 。(PERIODCATEGORY)
首(Nes) 波(Nf) 登場(VA) 的(T) 「(PARENTHESISCATEGORY) 主題(Na) 書展(Na) 」(PARENTHESISCATEGORY) ,(COMMACATEGORY)
展出(Vc) 2(Neu) 大(VH) 文豪(Na) 經典(Na) 作品(Na) 的(DE) 原著(Na) 、(PAUSECATEGORY) 各(Nes) 種(Nf) 譯本(Na) 以及(Caa)
校方(Na) 也(D) 添購(Vc) 許多(Nega) 新書(Na) ,(COMMACATEGORY)
吸引(VJ) 學生(Na) 「(PARENTHESISCATEGORY) 搶鮮(Na) 」(PARENTHESISCATEGORY) 閱讀(Vc) 經典(Na) 名作(Na) 。(PERIODCATEGORY)
現場(Nc) 還(D) 規畫(Vc) 「(PARENTHESISCATEGORY) 彩繪(Vc) 大師(Na) 」(PARENTHESISCATEGORY) ,(COMMACATEGORY)
讓(VL) 學生(Na) 發揮(VJ) 創意(Na) ,(COMMACATEGORY)
畫出(Vc) 五彩繽紛(VH) 的(DE) 莎士比亞(Nb) 和(Caa) 賽萬提(Nb) 斯人(Na) 像(VG) 。(PERIODCATEGORY)
英語系(Nc) 四年級(Na) 學生(Na) 陳彥伶(Nb) 說(VE) ,(COMMACATEGORY)
讀(Vc) 英語系(Nc) 接觸(Vc) 莎士比亞(Nb) 作品(Na) ,(COMMACATEGORY)
但(Cbb) 過去(Nd) 沒有(D) 舉辦(Vc) 書展(Na) 時(Ng) ,(COMMACATEGORY)
這些(Nega) 作品(Na) 都(D) 放(Vc) 在(P) 圖書館(Nc) 8樓(Nc) ,(COMMACATEGORY)

CKIP 中研院中文斷詞系統

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

莎士比亞在淡江 遇見賽萬提斯

2016-04-26 02:27 聯合報 記者徐葳倫／淡水報導

分享4月23日是「世界閱讀日」，也是英國大文豪莎士比亞的生日與忌日，及「唐吉訶德」作者賽萬提斯逝世之日。英專起家的淡江大學舉辦「當莎士比亞遇見賽萬提斯」活動，規畫主題書展、彩繪活動，並添購新書，拉近學生與經典文學的距離。

莎士比亞(Nb) 在(P) 淡江(Nb) 遇見(VC) 賽萬提(Nb) 斯(Nep) 2016(Neu) -(FW) 04(Neu) -(FW) 2602(Neu) :(COLONCATEGORY)

27(Neu) 聯合報(Nb) 記者(Na) 徐葳倫(Nb) 淡水(Nc) 報導(Na) 分享(VJ) 4月(Nd) 23日(Nd) 是(SHI) 「(PARENTHESISCATEGORY) 世界(Nc) 閱讀日(Na) 」(PARENTHESISCATEGORY) , (COMMACATEGORY)

也(D) 是(SHI) 英國(Nc) 大(VH) 文豪(Na) 莎士比亞(Nb) 的(DE) 生日(Na) 與(Caa) 忌日(Na) , (COMMACATEGORY)

及(Caa) 「(PARENTHESISCATEGORY) 唐吉訶德(Nb) 」(PARENTHESISCATEGORY) 作者(Na) 賽萬提(Nb) 斯(Nep) 逝世(VH) 之(DE) 日(Na) 。(PERIODCATEGORY)

英(Nc) 專(D) 起家(VA) 的(DE) 淡江(Nb) 大學(Nc) 舉辦(VC) 「(PARENTHESISCATEGORY) 當(P) 莎士比亞(Nb) 遇見(VC) 賽萬提(Nb) 斯(Nep) 」(PARENTHESISCATEGORY) 活動(Na) , (COMMACATEGORY)

規畫(VC) 主題(Na) 書展(Na) 、(PAUSECATEGORY) 彩繪(VC) 活動(Na) , (COMMACATEGORY) 並(Cbb) 添購(VC) 新書(Na) , (COMMACATEGORY)

拉近(VC) 學生(Na) 與(Caa) 經典(Na) 文學(Na) 的(DE) 距離(Na) 。(PERIODCATEGORY)



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

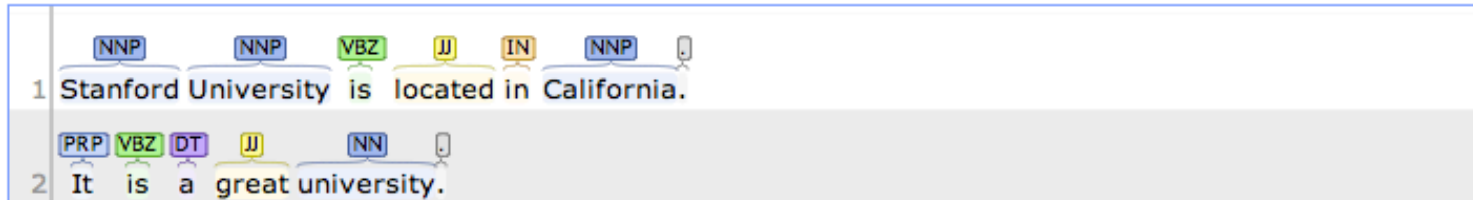
Stanford CoreNLP

Output format:

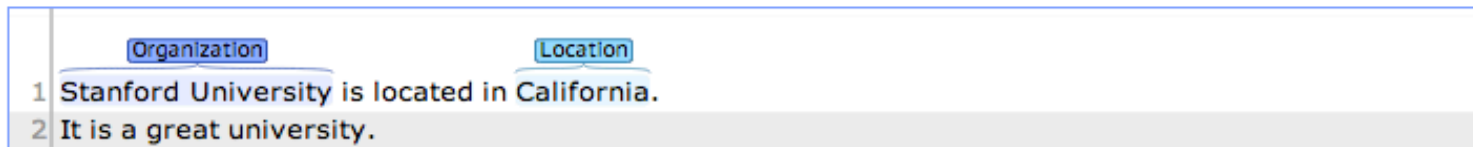
Please enter your text here:

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

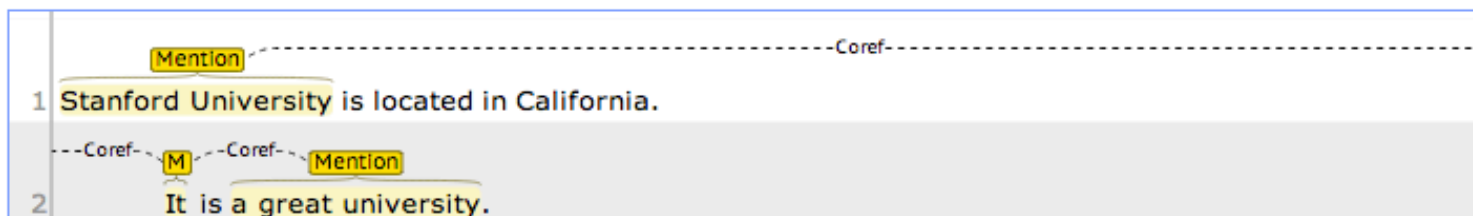
Part-of-Speech:



Named Entity Recognition:



Coreference:

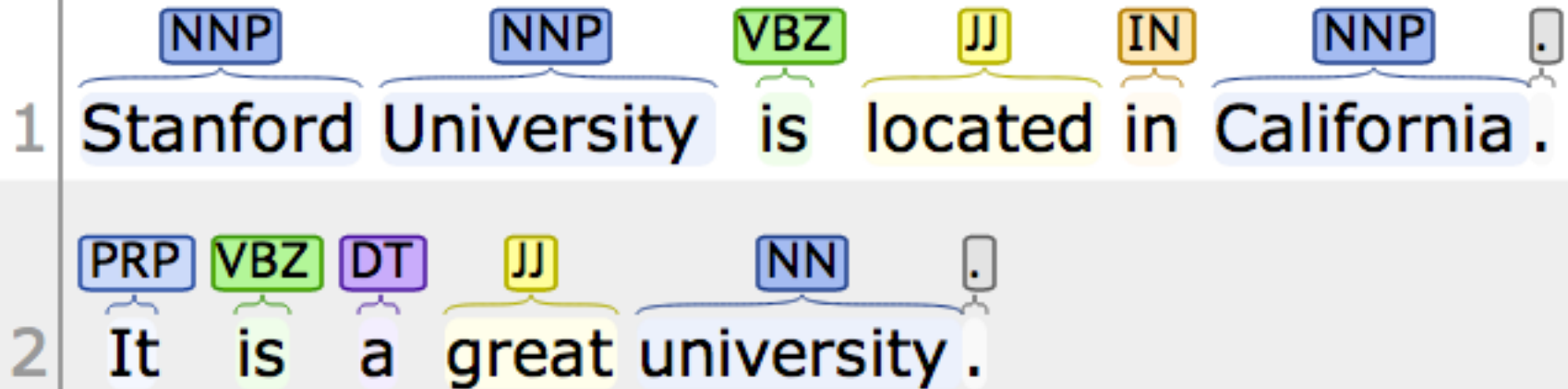


Stanford CoreNLP

<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:

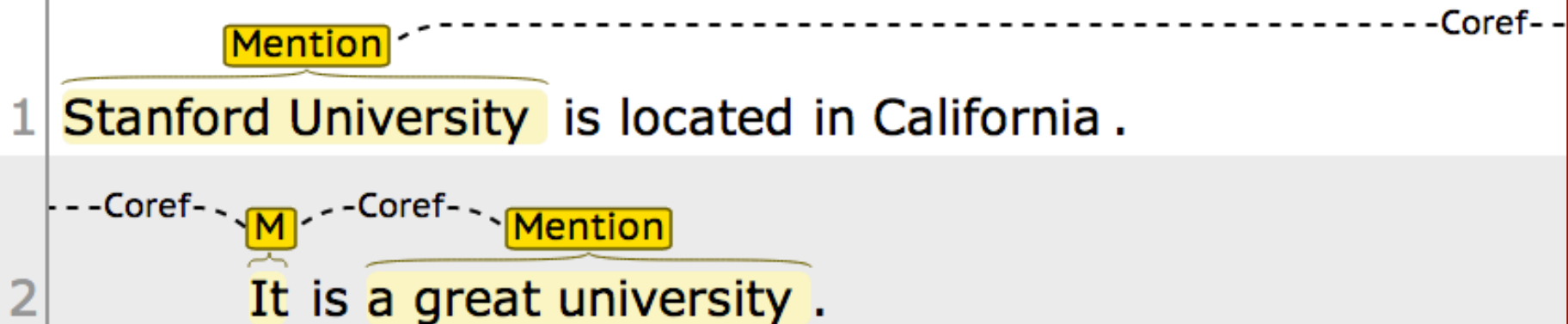
	Organization		Location
1	Stanford University	is located in	California .
2	It is a great university .		

Stanford CoreNLP

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

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

Coreference:

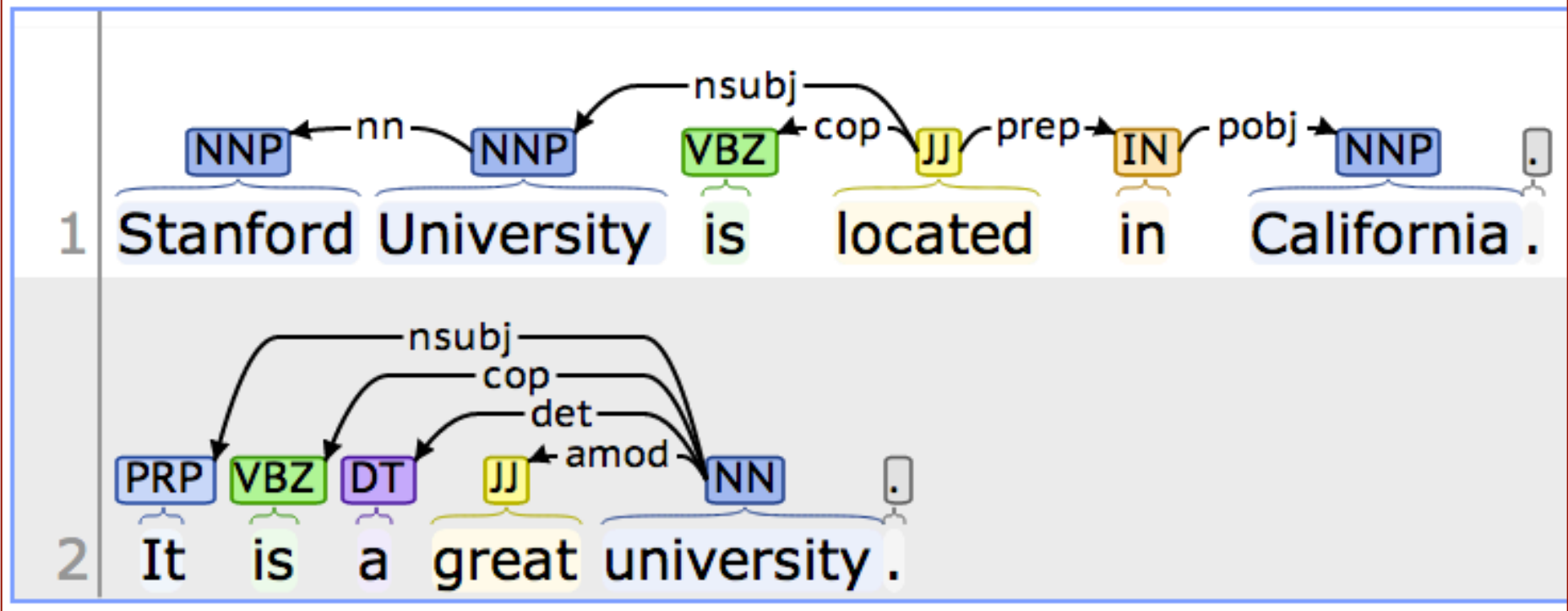


Stanford CoreNLP

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

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

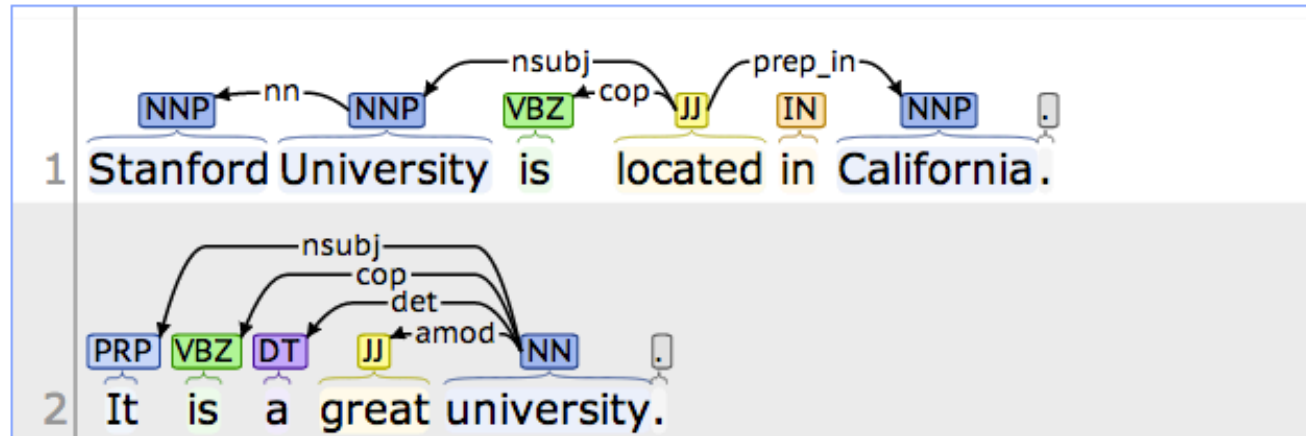
Basic dependencies:



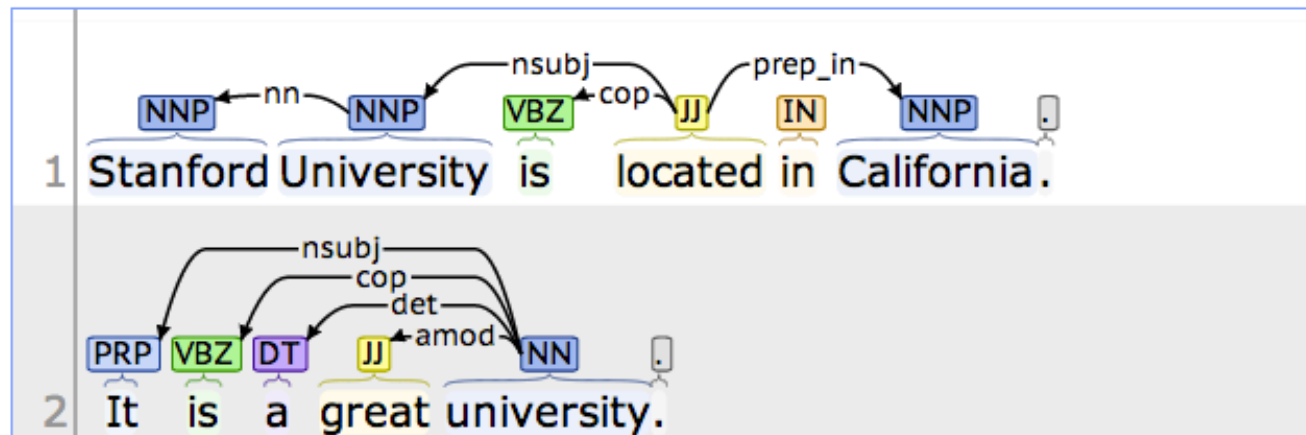
Stanford CoreNLP

<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

Sentence #1

Tokens

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

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 #1

Tokens

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

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

Id	Word	Lemma	Char begin	Char end	POS	NER	Normalized NER	Speaker
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: XML

Please enter your text here:

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

Submit

Clear

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet href="CoreNLP-to-HTML.xsl" 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>PERO</Speaker>
          </token>
          <token id="2">
            <word>University</word>
            <lemma>University</lemma>
            <CharacterOffsetBegin>9</CharacterOffsetBegin>
            <CharacterOffsetEnd>19</CharacterOffsetEnd>
            <POS>NNP</POS>
            <NER>ORGANIZATION</NER>
            <Speaker>PERO</Speaker>
          </token>
```

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



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)

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>.

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

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

Output Format:

Preserve Spacing:

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) —

Submit Clear

<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">'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">'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">years</wi> <wi num="5" entity="O">Gates</wi> <wi num="6" entity="O">has</wi> <wi num="7" entity="O">sold</wi>

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 (CNNTech) —

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** (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

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** (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

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

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)

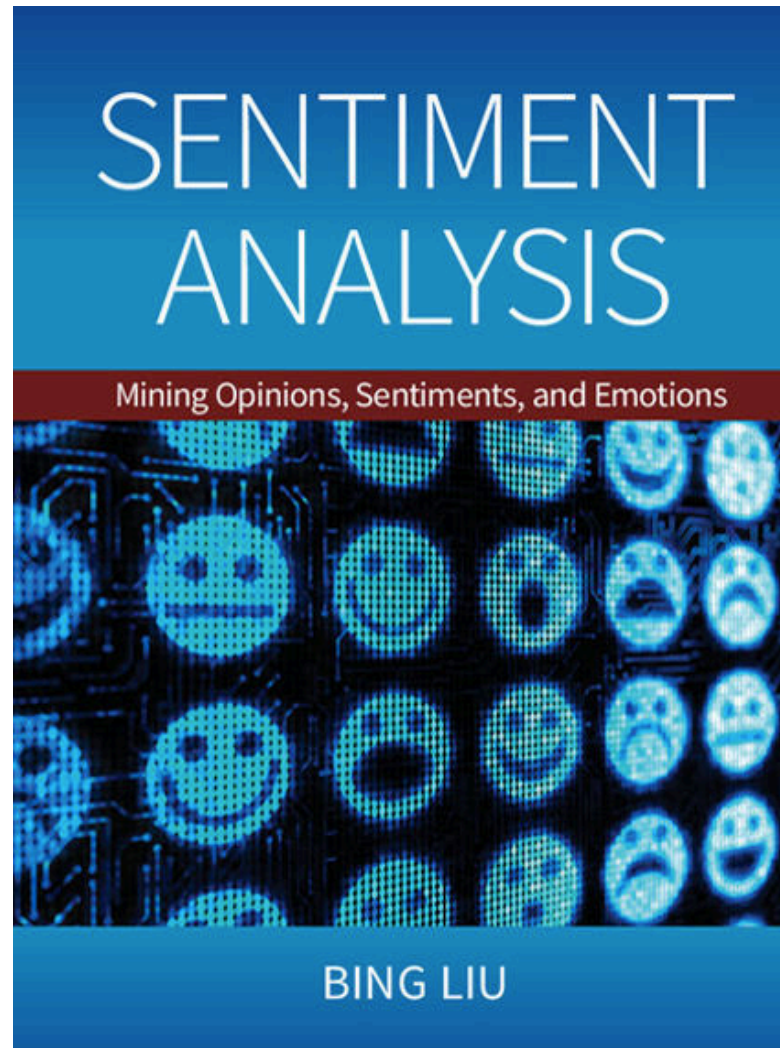
授權金額: 免費授權學術使用。

授權日期: 2011.05.03。

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

Architectures of Sentiment Analytics

Bing Liu (2015),
Sentiment Analysis:
Mining Opinions, Sentiments, and Emotions,
Cambridge University Press



Sentiment Analysis and Opinion Mining

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

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

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**.

(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. ...”

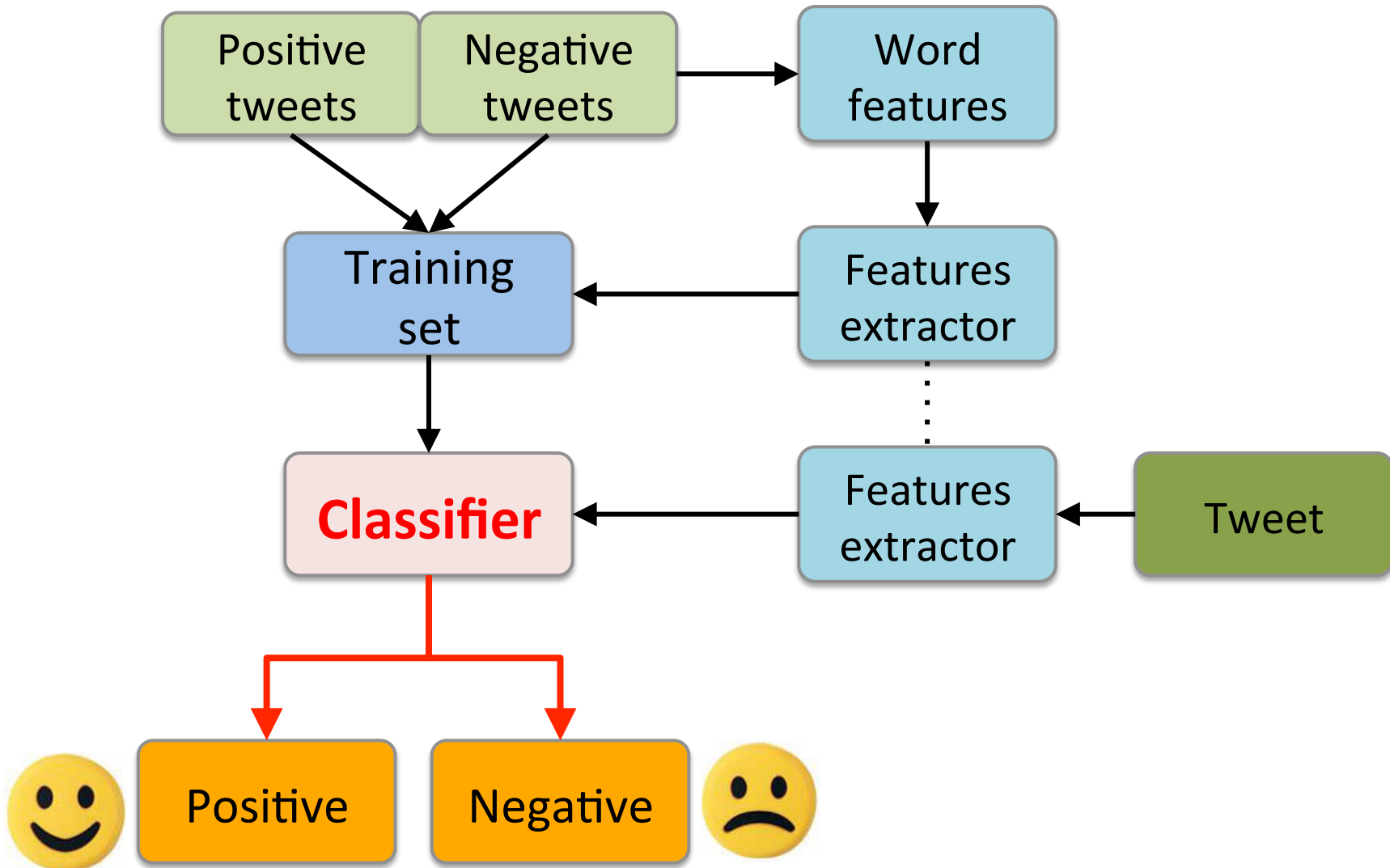


**+Positive
Opinion**

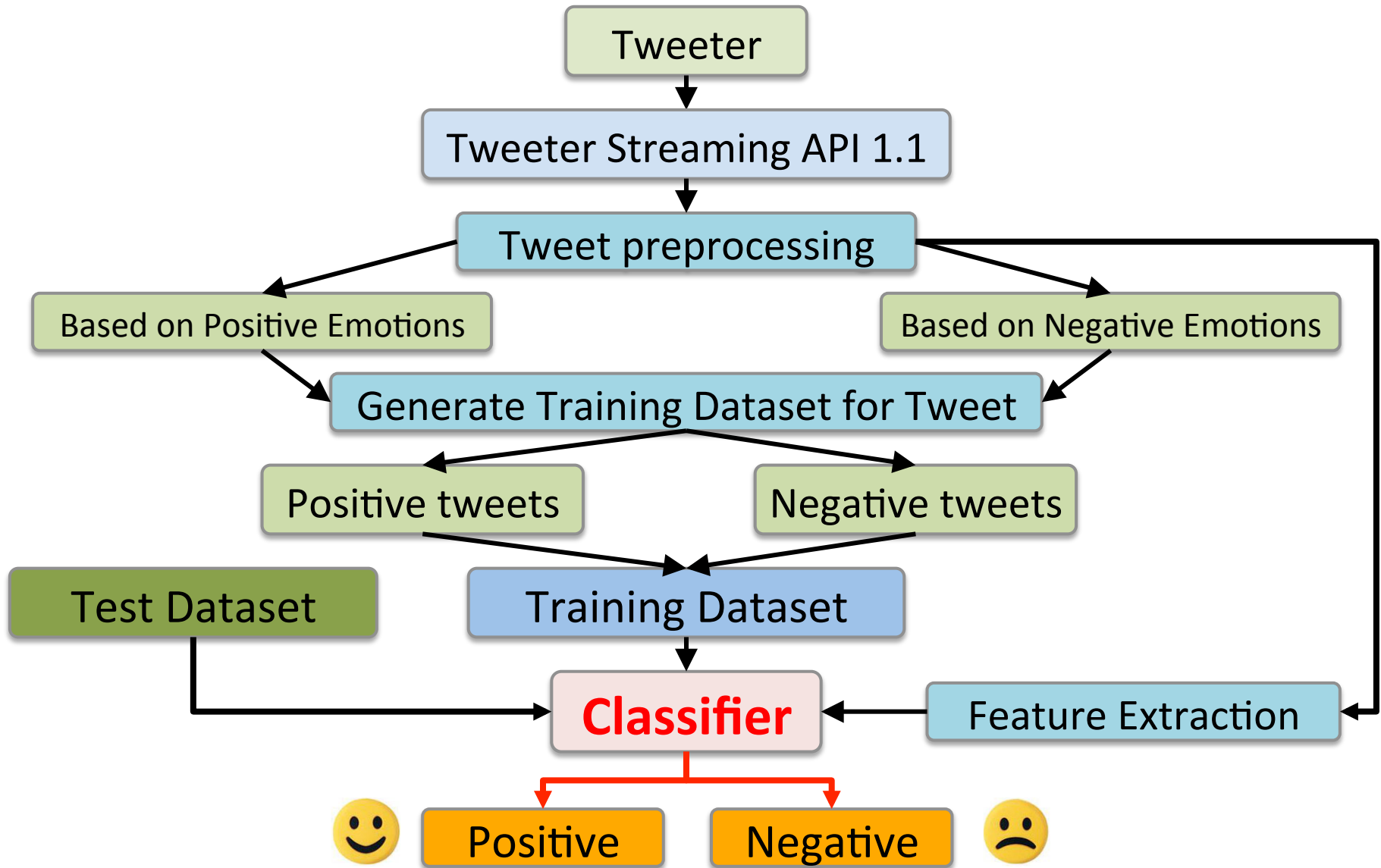


**-Negative
Opinion**

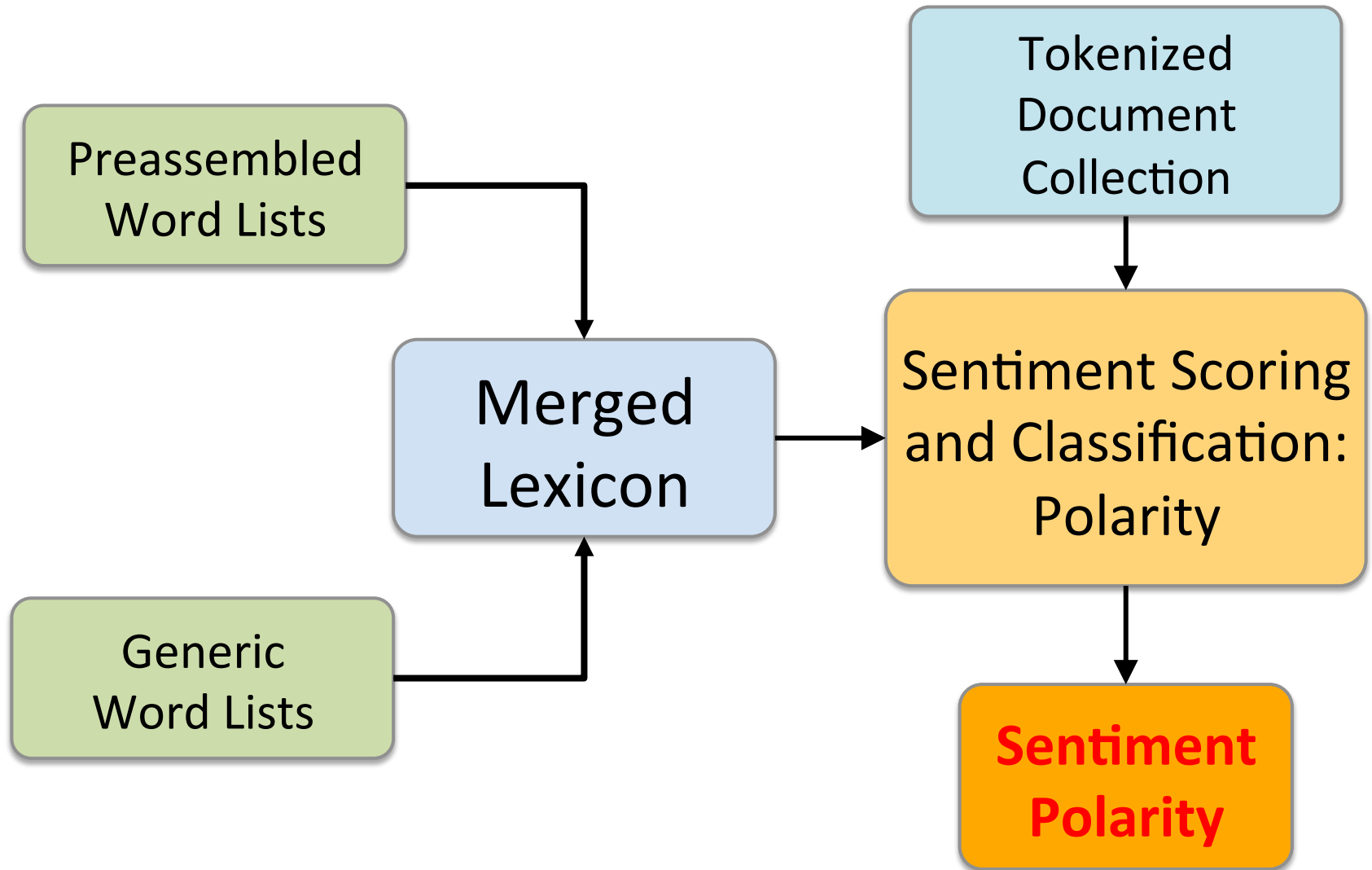
Sentiment Analysis Architecture



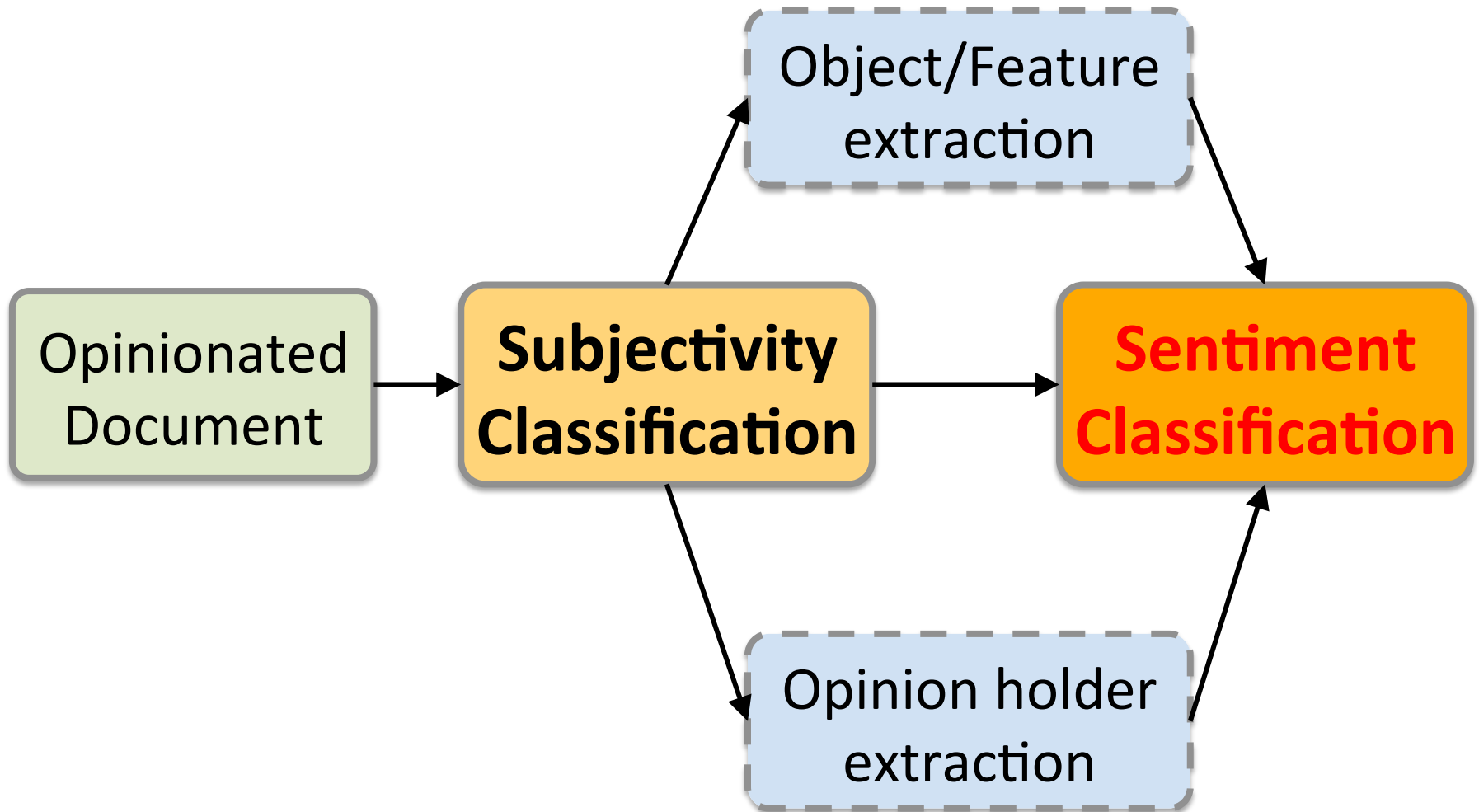
Sentiment Classification Based on Emoticons



Lexicon-Based Model



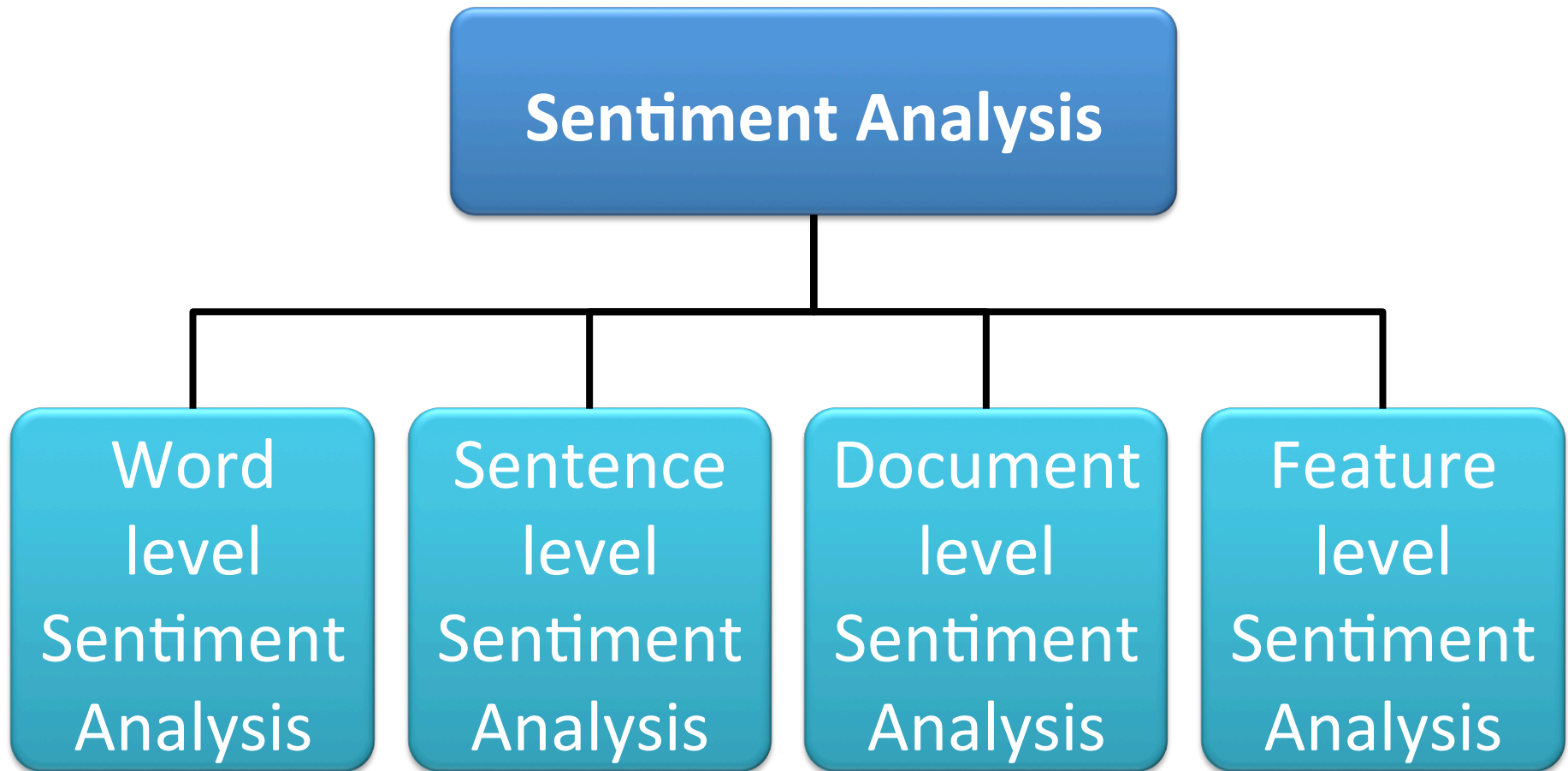
Sentiment Analysis Tasks



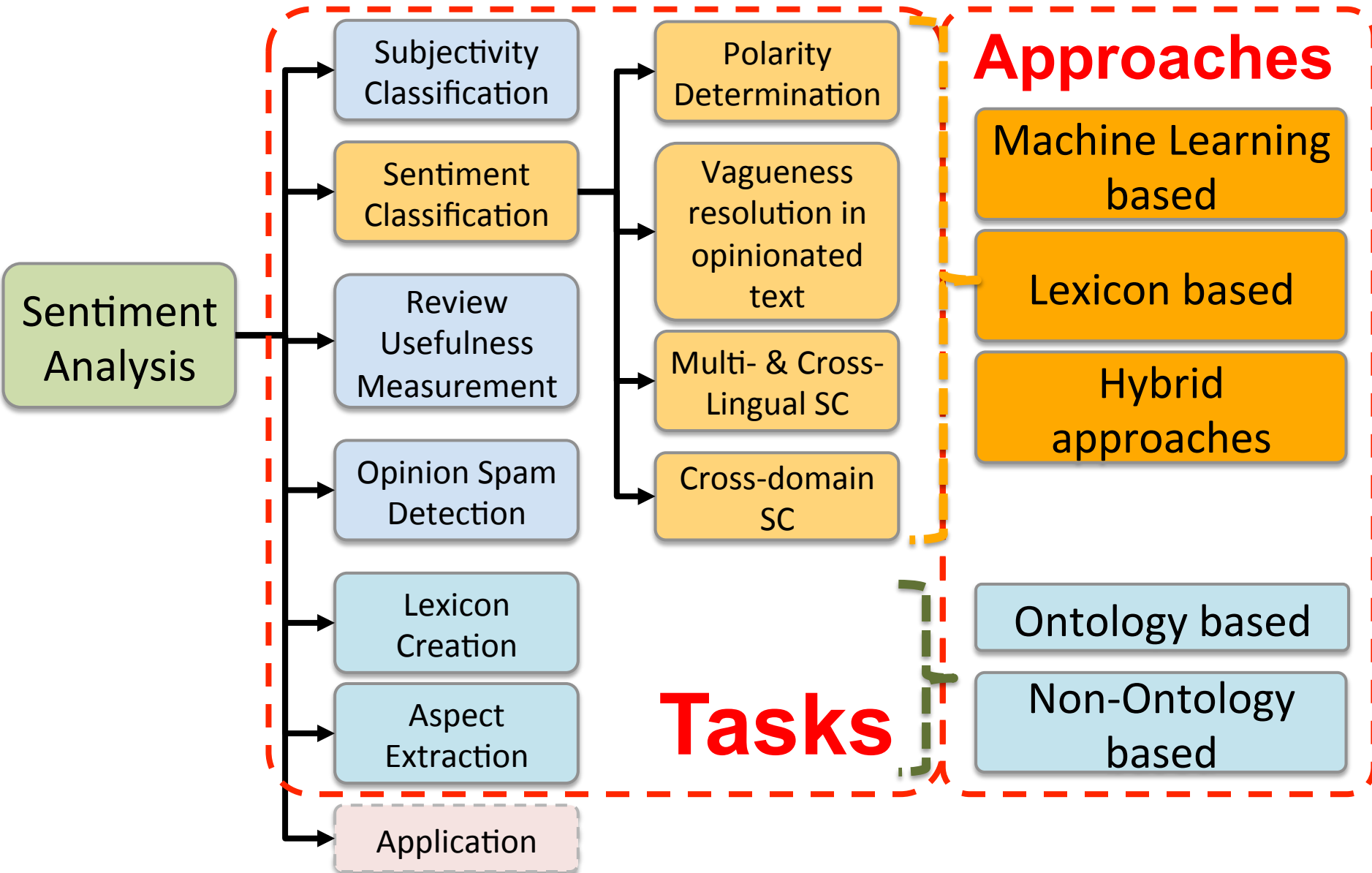
Sentiment Analysis vs. Subjectivity Analysis

Sentiment Analysis	Subjectivity Analysis
Positive	Subjective
Negative	
Neutral	Objective

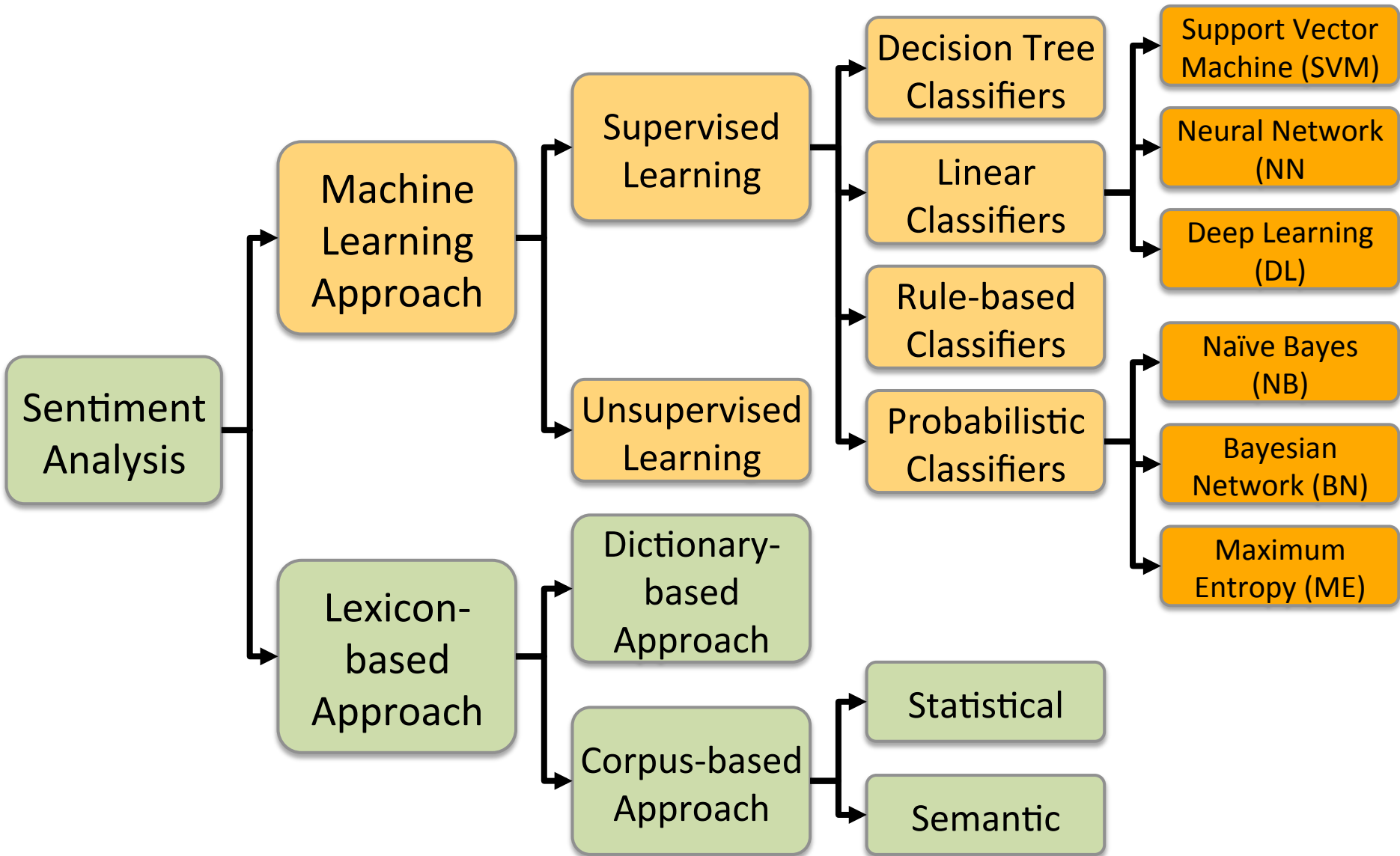
Levels of Sentiment Analysis



Sentiment Analysis



Sentiment Classification Techniques



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"	

Evaluation of Text Mining and Sentiment Analysis

- Evaluation of Information Retrieval
- Evaluation of Classification Model (Prediction)
 - Accuracy
 - Precision
 - Recall
 - F-score

Business Insights with Social Analytics

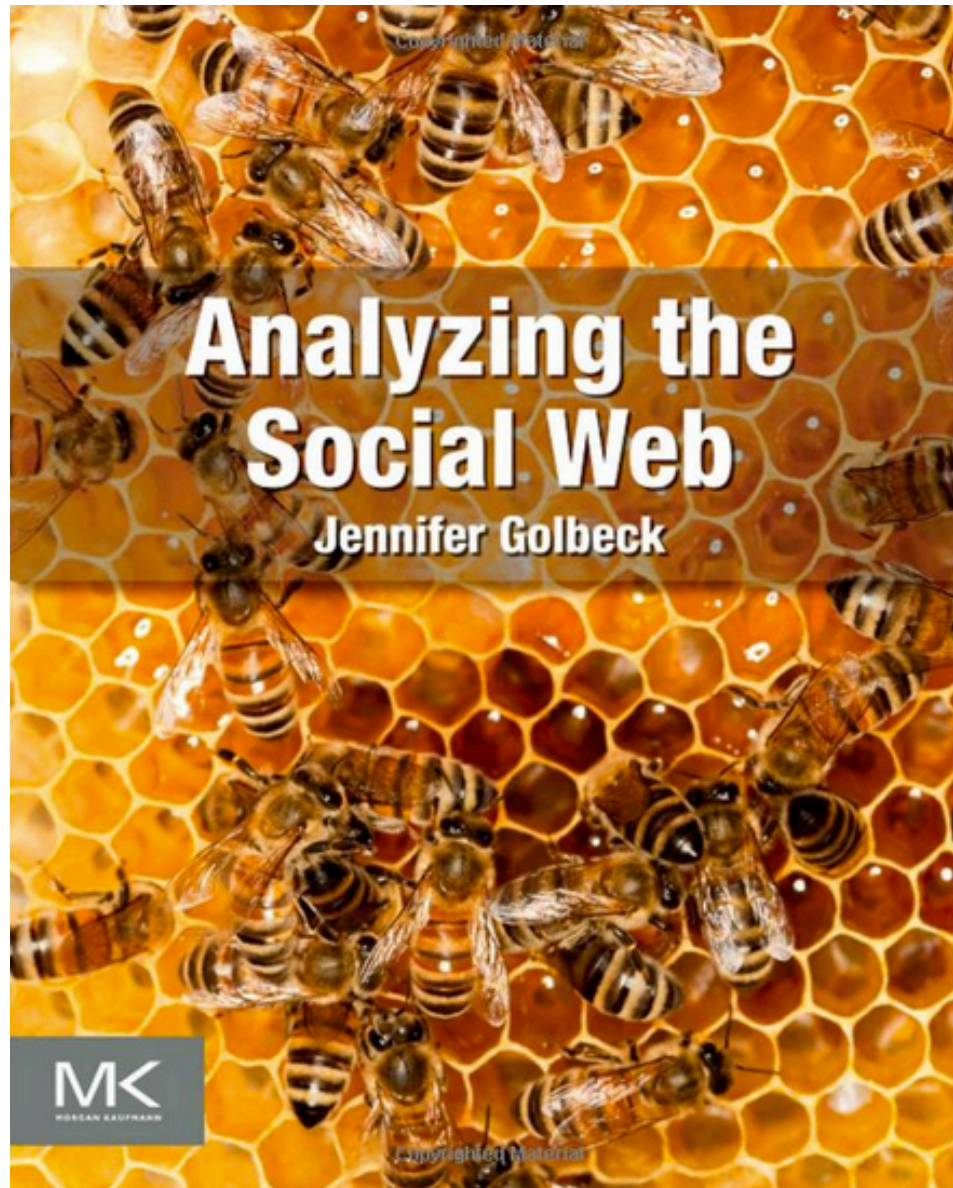
Social Computing

- Social Network Analysis
- Link mining
- Community Detection
- Social Recommendation

Analyzing the Social Web:

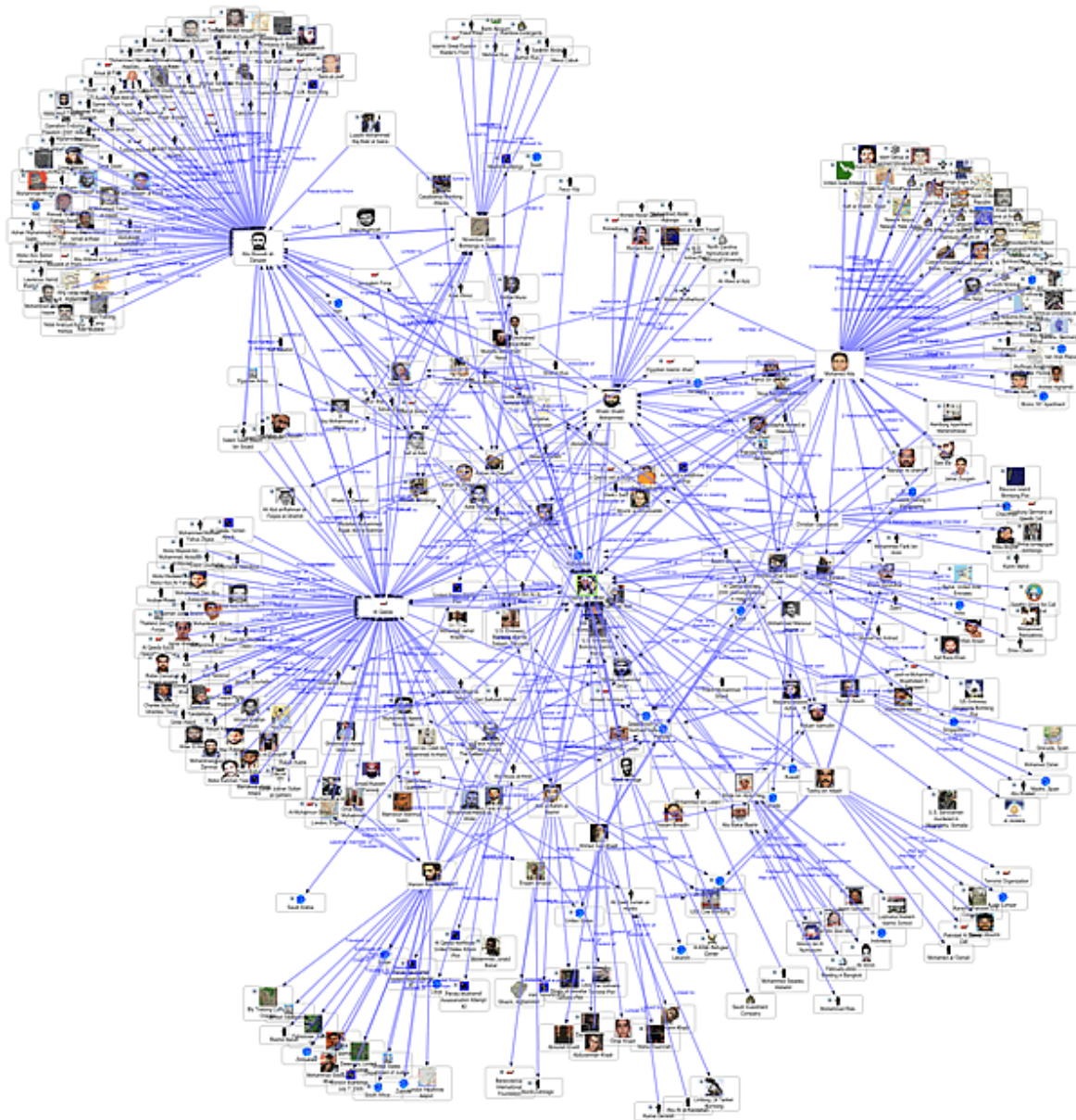
Social Network Analysis

Jennifer Golbeck (2013), **Analyzing the Social Web**, Morgan Kaufmann



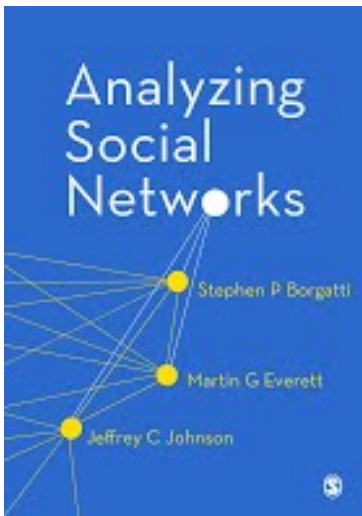
Source: <http://www.amazon.com/Analyzing-Social-Web-Jennifer-Golbeck/dp/0124055311>

Social Network Analysis



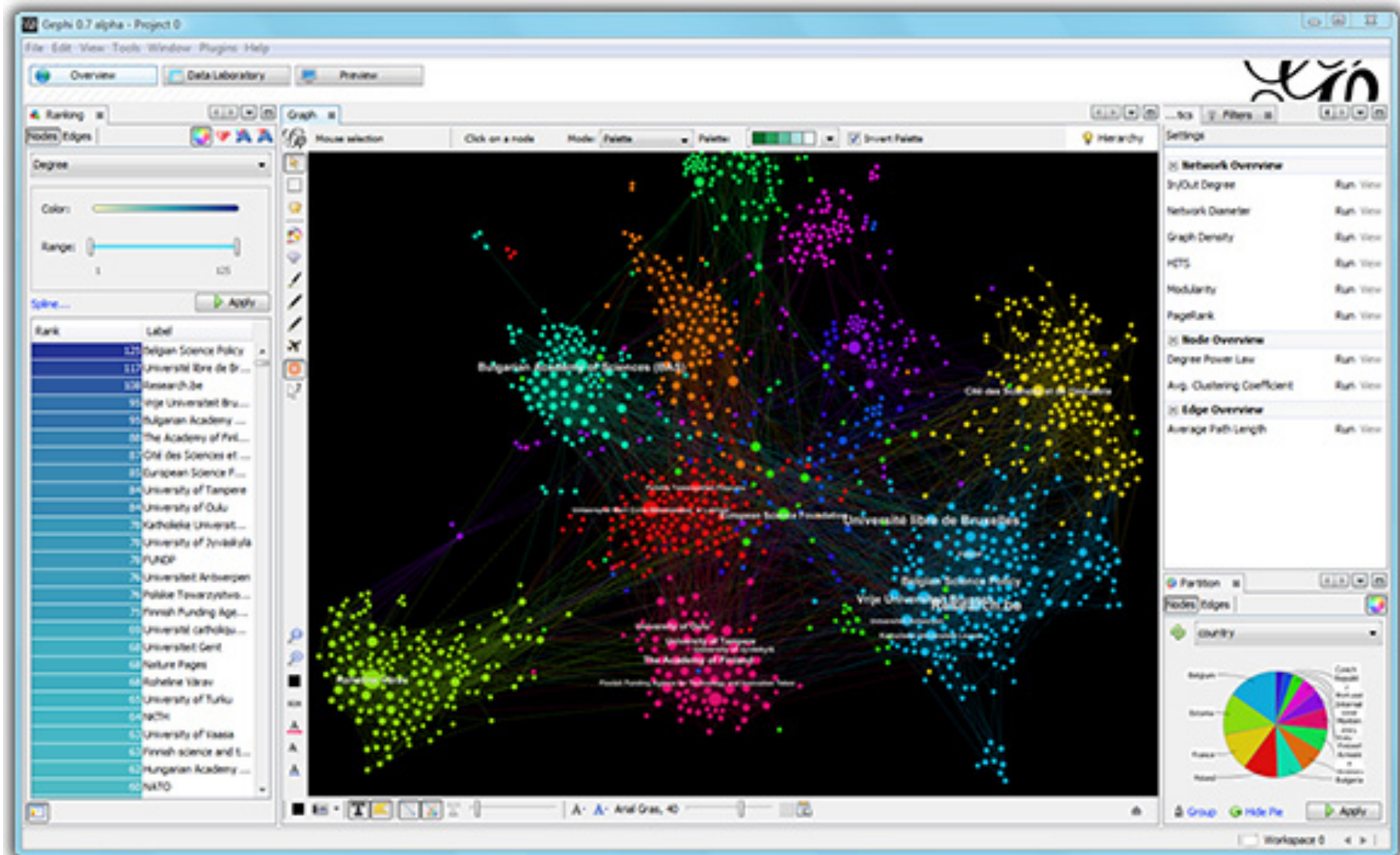
Social Network Analysis (SNA) Tools

- **NetworkX**
- **igraph**
- **Gephi**
- **UCINet**
- **Pajek**



Gephi

The Open Graph Viz Platform



Attensity: Track social sentiment across brands and competitors

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

The screenshot displays the Attensity website homepage. At the top, there's a navigation bar with the Attensity logo, a language selector set to 'English', and links for 'Contact', 'Resources', 'Support', 'Blog', and a search bar. Below this is a secondary navigation bar with 'Products', 'Solutions', 'Services', 'Customers', and 'Partners'. The main content area features a large central banner with the headline 'Your real-time window into the social web.' and a quote from Yahoo! praising Attensity's analytics. To the left of the banner is a sidebar with links to 'Social Analytics', 'Social Response', 'Customer Analytics', 'Industry Solutions', and 'Why Attensity'. To the right, there are several smaller panels showing various analytics: a 'Comparison of Feedback Over Different Time Periods' bar chart, a 'Net Promoter Score' gauge, a 'Sentiment Analysis' pie chart, and a 'Twitter Accounts' list. At the bottom, there are four distinct sections: 'Attensity for Marketing' and 'Attensity for Customer Service' (both with sub-sections for IT and effectiveness), a 'Success Story' for JetBlue Airways with a 'DOWNLOAD NOW' button, an 'About Attensity' section stating it's the leading provider of social analytics, and a 'Watch Video' section for the 'Command Center Video'.

Attensity Home Page | Attensity

www.attensity.com/home/

Select your language **English**

ATTENSITY

Products Solutions Services Customers Partners

Your real-time window into the social web.

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

– Yahoo!

Learn More

Social Analytics

Social Response

Customer Analytics

Industry Solutions

Why Attensity

Attensity for Marketing

Attensity for Customer Service

Attensity for IT

Effectiveness of your social marketing strategies:

Success Story

JetBlue Airways

Listen.

About Attensity

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

Watch Video

Command Center Video

ATTENSITY

Voice of the Customer. Done Right.

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

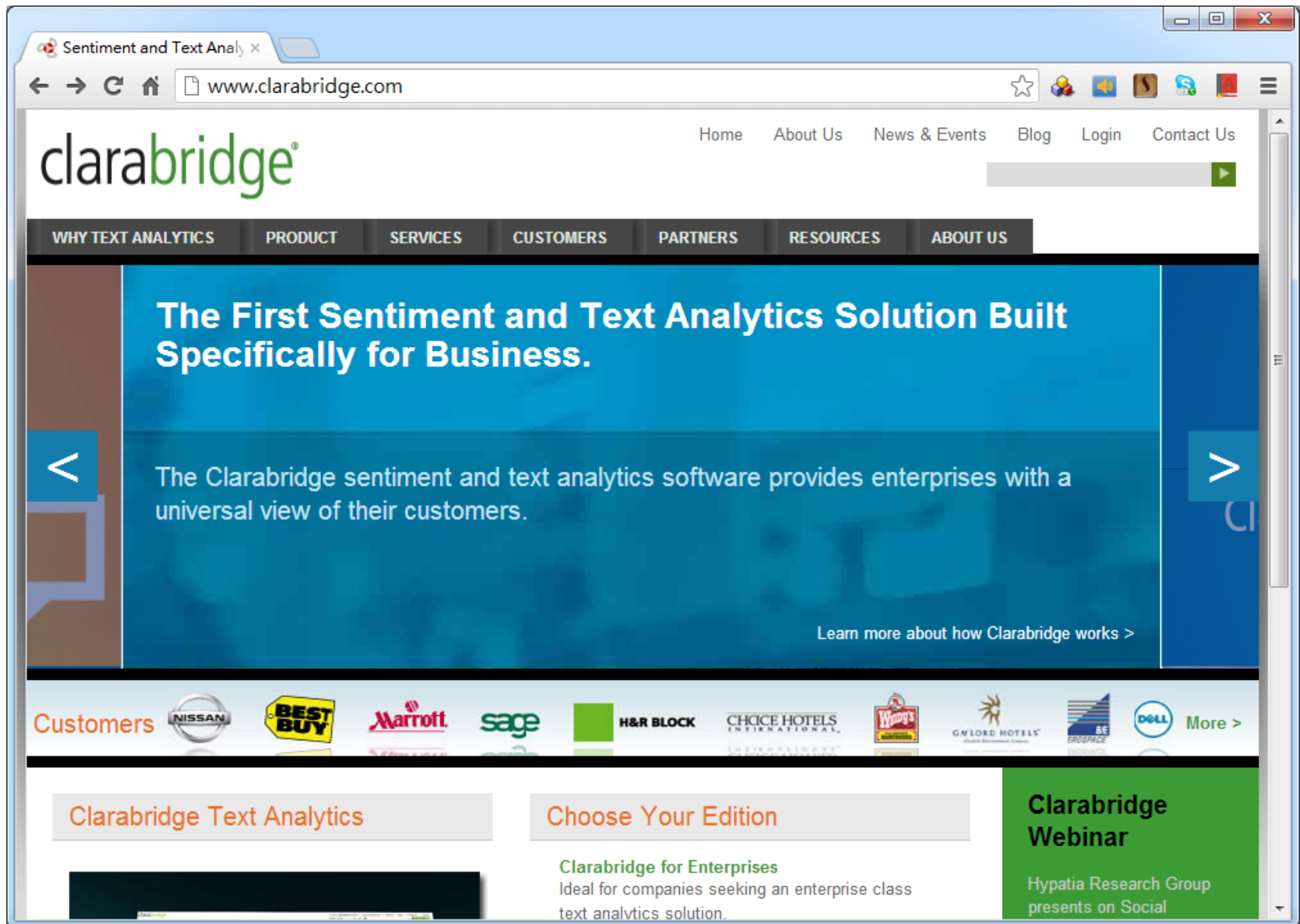
ence.

DOWNLOAD NOW

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

Clarabridge: Sentiment and Text Analytics Software

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



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

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

Social Media Monitoring x

www.radian6.com

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

salesforce **radian6**

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

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

Have Us Contact You

Live Demo

Free Trial

Chat & find out more.

Offline. Leave us a message

f t YouTube g+ radian6 Community

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

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



THE POWER TO KNOW.
Providing software solutions since 1976

[Home](#)
[Products & Solutions](#)
[Customer Success](#)
[Partners](#)
[Company](#)
[Support & Training](#)

[Log In](#)
[Worldwide Sites](#)
[Contact Us](#)
[Follow Us](#)

[NEWS](#)
[EVENTS](#)
[CONSULTING](#)
[CAREERS](#)
[RESOURCE CENTER](#)

PRODUCTS & SOLUTIONS / SOCIAL MEDIA ANALYTICS

Products and Solutions

- Industries
- Small and Midsize Business
- Nonprofit Organizations
- Analytics
 - Business Analytics
 - Business Intelligence
 - Customer Intelligence
 - Strategy & Planning
 - Information & Analytics
 - Orchestration & Interaction
 - Customer Experience
 - Customer Experience Analytics
 - Social Media Analytics
 - Web Analytics
- Financial Intelligence
- Foundation Tools
- Fraud & Financial Crimes
- Governance, Risk & Compliance
- High-Performance Analytics
- Human Capital Intelligence
- Information Management
- IT & CIO Enablement

SAS® Social Media Analytics

Integrate, archive, analyze and act on online conversations

[Overview](#)
[Benefits](#)
[Features](#)
[Demos & Screenshots](#)
[System Requirements](#)


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

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

Benefits

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

Product Demo






“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](#)

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](#)

What do tweeples think ab x

www.tweetfeel.com/index.php#iPhone4s

FAQ | Contact Us

tweetfeel

|| iPhone4s Search

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

 40  41 = 51%

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

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

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

 This 12 year old has an iPhone4s wtf

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

 iPhone4s is funny ass a bitch

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

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

Follow us
Email us


Brought to you by
conversion

Powered by
twitter

eLand

<http://www.eland.com.tw/>



關於意藍 產品與雲端服務 ▾ 新聞與活動 聯絡資訊 

< 巨量搜尋。語意分析。社群大數據 >



OpView社群口碑資料庫

OpView

<http://www.opview.com.tw/>



OpView 介紹 ▾

產業應用 ▾

新聞與活動

分析報告

資源與課程 ▾

聯絡資訊



社群大數據

觀測 · 分析 · 探索 · 預警





首頁

i-Buzz系統

產業數據

時事議題

社群研究

解決方案

FOCUS

i-Buzz

VOC口碑分析平台

自動化海量資料分析
迅速掌握網路口碑動態



母親節好禮大比拼 聽聽網友怎麼說

這個周末就是母親節了，大家有想好要如何慶祝了嗎？吃大餐、送好禮已成了節慶的基本盤，再加上百貨針對母親節紛紛推出特賣優惠，不僅讓孝子孝女省下荷包，也讓平常有在觀望檔期活動的網友殺紅了眼，更增添了其口碑豐富性...

i-Buzz

專業口碑客服團隊

公關危機處理，扭轉話題關鍵
提供具有科學性的策略方針

熱門文章



Summary

- 文字探勘技術
(Text Mining Technologies)
- 情感分析架構
(Architectures of Sentiment Analytics)

References

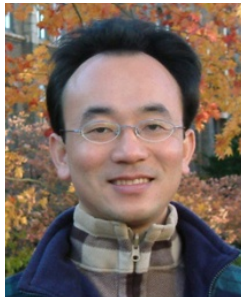
- Kumar Ravi and Vadlamani Ravi (2015), "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems, 89, pp.14-46.
- Vishal Kharde and Sheetal Sonawane (2016), "Sentiment Analysis of Twitter Data: A Survey of Techniques," International Journal of Computer Applications, vol 139, no. 11, 2016. pp.5-15.
- Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, and Enrique Herrera-Viedma (2015), "Sentiment analysis: A review and comparative analysis of web services," Information Sciences, 311, pp. 18-38.
- Vishal Gupta and Gurpreet S. Lehal (2009), "A survey of text mining techniques and applications," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60-76.
- Steven Struhl (2015), Practical Text Analytics: Interpreting Text and Unstructured Data for Business Intelligence (Marketing Science), Kogan Page
- Bing Liu (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press
- Jiawei Han and Micheline Kamber (2011), Data Mining: Concepts and Techniques, Third Edition, Elsevier
- Jennifer Golbeck (2013), Analyzing the Social Web, Morgan Kaufmann
- Stephan Kudyba (2014), Big Data, Mining, and Analytics: Components of Strategic Decision Making, Auerbach Publications
- Hiroshi Ishikawa (2015), Social Big Data Mining, CRC Press
- Efraim Turban, Ramesh Sharda, Dursun Delen, Decision Support and Business Intelligence Systems, Ninth Edition, 2011, Pearson.
- Michael W. Berry and Jacob Kogan, Text Mining: Applications and Theory, 2010, Wiley
- Guandong Xu, Yanchun Zhang, Lin Li, Web Mining and Social Networking: Techniques and Applications, 2011, Springer
- Matthew A. Russell, Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites, 2011, O'Reilly Media
- Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, 2009, Springer

Q & A

社群網路文字探勘與情感分析 (Text Mining and Sentiment Analytics on Social Media)

Time: 2016/05/10 (Tue) (12:00-14:00)

Place: 淡水校園覺生綜合大樓I501室



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/>

2016-05-10

