

# CORSO DI *WEB MINING E RETRIEVAL* *- INTRODUZIONE AL CORSO -*

---

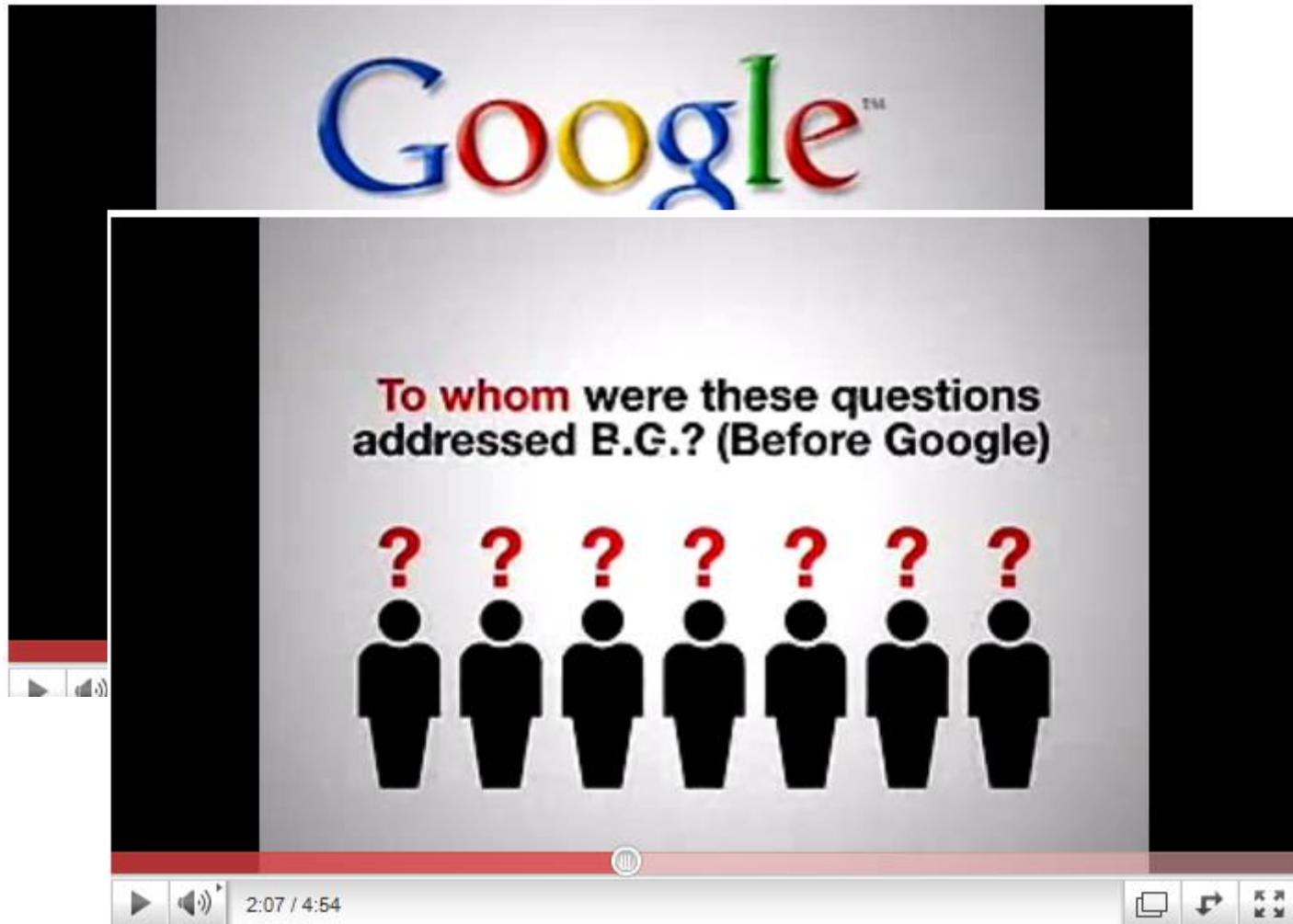
Corso di Laurea in Informatica, Ing. Internet,  
Ing. Informatica, Ing. Gestionale  
(a.a. 2014-2015)

Roberto Basili

# Overview

- WM&R: Motivazioni e prospettive
- Modalità di erogazione del Corso
- Prerequisiti
- Forma e struttura delle prove d'esame

# Did you know?



# Did you know?



# A Web of people and opinions

- **31.7%** of the more than 200 million bloggers worldwide blog about opinions on products and brands (Universal McCann, July 2009)
- **71%** of all active Internet users read blogs.
- 2009 Survey of **25,000** Internet users in **50** countries: **70%** of consumers trust opinions posted online by other consumers (Nielsen Global Online Consumer, 2010).

# Social Media Analytics

- Complex process for Social Media Analytics are necessary whereas ...
- ... Opinion Mining and Sentiment Analysis play a crucial role



# WM&R: Motivazioni

- *Cos'è il Web Mining?*
- *Perché IR?*
- *Perché Apprendimento Automatico?*
- *Quale contributo l'IR fornisce alle tecnologie di sfruttamento delle informazioni del Web?*
- *Quali sono le prospettive per l'impiego di tali tecnologie?*

# Cos'è il Web Mining?

- *Web Mining* attualmente si riferisce ad un insieme di tecnologie necessarie per lo **sfruttamento** delle informazioni pubblicamente disponibili nel Web
  - Contenuti: dati ma anche ... persone, luoghi, eventi, concetti, ...
  - Relazioni:
    - Link strutturali
    - Collegamenti tematici, concettuali e interpersonali
    - Ridondanze/analogie
  - Multilingualità
  - Trend e comportamenti collettivi
  - Opinioni

# Perché IR?

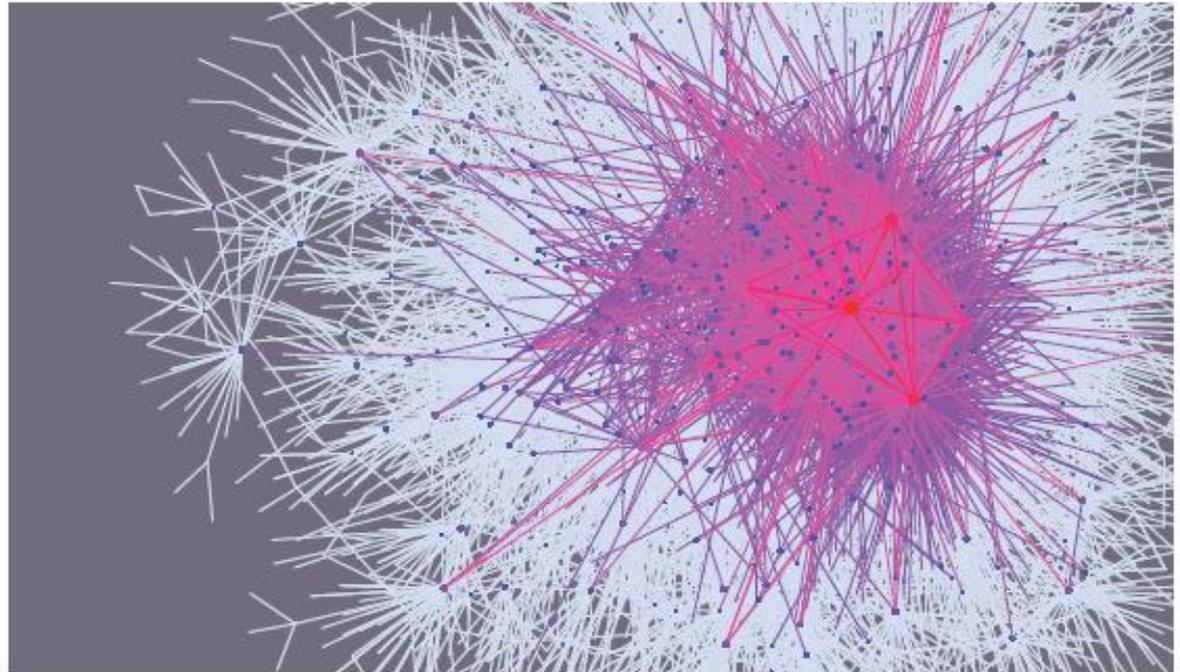
- La taglia delle informazioni in gioco pone il problema della *localizzazione*
- Accedere in modo automatico è possibile solo governando il problema di sapere **dove** si trova una informazione *rilevante*
- La ricerca corrisponde al calcolo di una funzione *aleatoria* di mapping tra requisiti e informazione utile

# Machine Learning vs IR?

- La eterogeneità delle informazioni produce significativi effetti di incertezza nel processo di ricerca
  - Incompletezza della informazione
  - Ricchezza di dati, formati e modalità di accesso
  - Requisiti vaghi
  - Aspetti soggettivi
  - Tempestività

# ML vs. IR

- La pervasività degli elementi di incertezza rende impraticabile la ricerca di soluzioni esaustive (ottimi globali)
- “*Finding diamonds in the rough*”  
(Fan Chung, UCSD)



# ML vs. IR

- Le tecniche di ML propongono una ampia serie di algoritmi, strategie e tecniche per la produzione di soluzioni *sub-ottime* effettive
- Nel processo di *learning* i dati suggeriscono la ipotesi risolutiva per la funzione di *mapping*
- Tale ipotesi è attesa migliorare la prestazione complessiva del sistema di base
  - Accuratezza
  - Efficienza computazionale

# Machine Learning

- (Langley, 2000): l'Apprendimento Automatico si occupa dei meccanismi attraverso i quali un agente intelligente migliora nel tempo le sue prestazioni  $P$  nell'effettuare un compito  $C$ .
- La prova del successo dell'apprendimento è quindi nella capacità di misurare l'incremento  $\Delta P$  delle prestazioni sulla base delle esperienze  $E$  che l'agente è in grado di raccogliere durante il suo ciclo di vita.
- La natura dell'apprendimento è quindi tutta nella caratterizzazione delle nozioni qui primitive di *compito*, *prestazione* ed *esperienza*.

# Esperienza ed Apprendimento

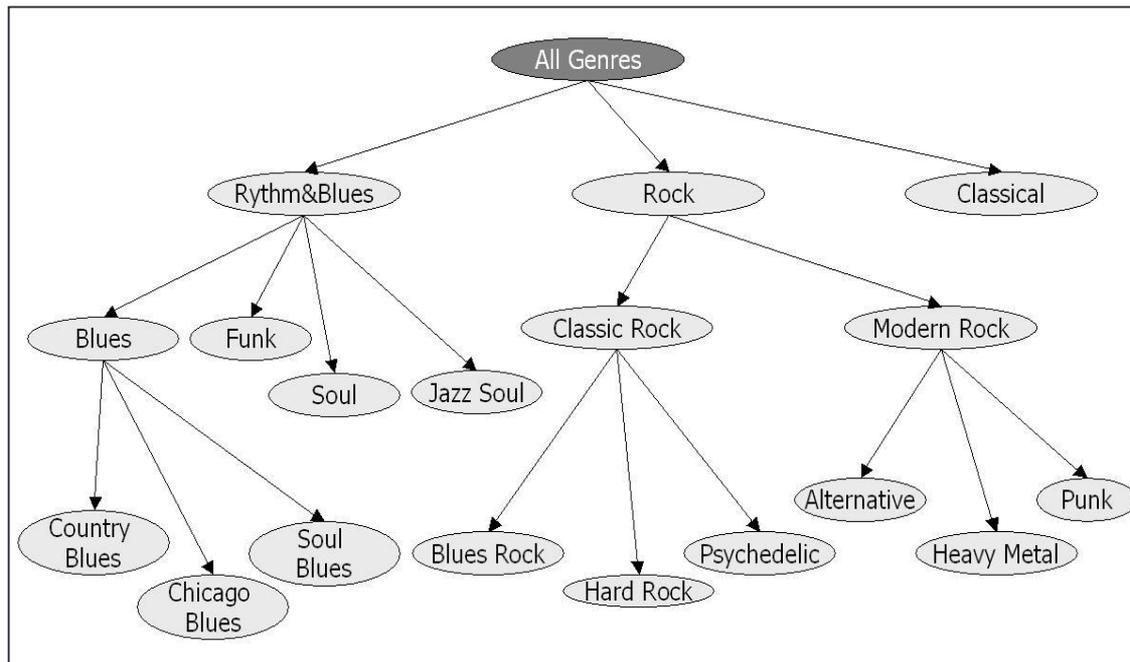
- L'esperienza, per esempio, nel gioco degli scacchi può essere interpretata in diversi modi:
  - i dati sulle vittorie (e sconfitte) pregresse per valutare la bontà (o la inadeguatezza) di strategie e mosse eseguite rispetto all'avversario.
  - valutazione fornita sulle mosse da un docente esterno (oracolo, guida).
  - Adeguatezza dei comportamenti derivata dalla auto-osservazione, cioè dalla capacità di analizzare partite dell'agente contro se stesso secondo un modello esplicito del processo (partita) e della sua evoluzione (comportamento, vantaggi, ...).

# Apprendimento senza supervisione

- In assenza di un oracolo o di conoscenze sul task esistono ancora molti modi di migliorare le proprie prestazioni, ad es.
  - Migliorando il proprio modello del mondo (acquisizione/*discovery* della conoscenza)
  - Migliorando le proprie prestazioni computazionali (ottimizzazione)

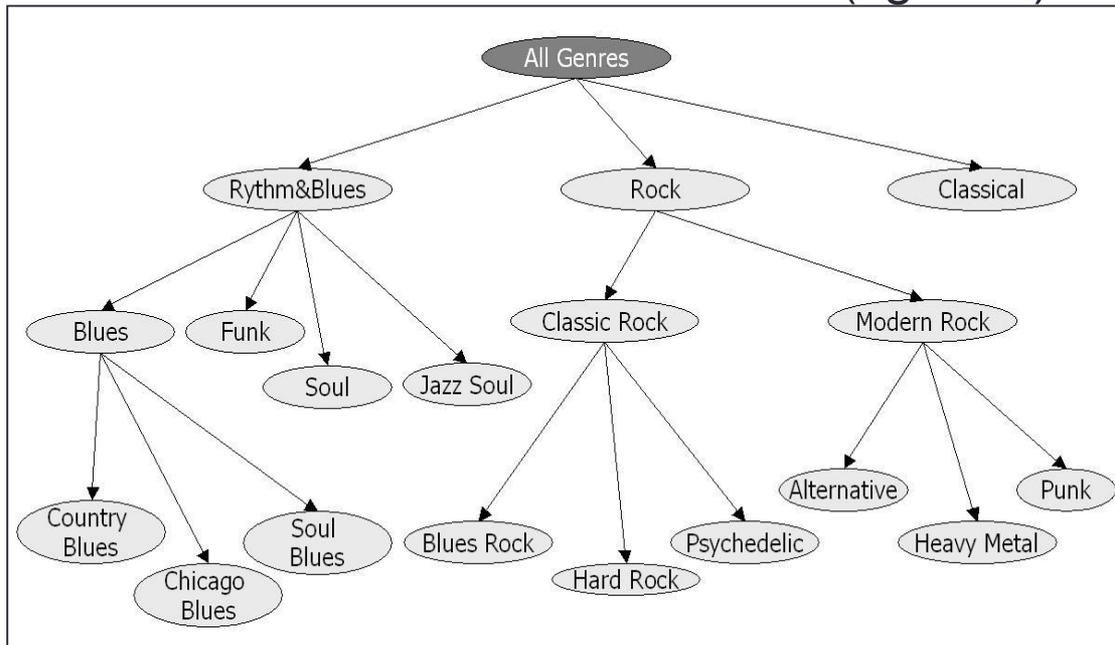
# Apprendimento senza supervisione

- Esempio:
  - una collezione mp3 può essere organizzata in generi attraverso il raggruppamento di brani simili secondo proprietà audio (*clustering*): tale organizzazione è naturalmente gerarchica
  - Il miglioramento avviene quindi almeno rispetto agli algoritmi di ricerca: la organizzazione gerarchica consente di esaminare solo i membri dell'insieme in alcune classi (i generi).



# Apprendimento senza supervisione

- Esempio: Al termine del processo di acquisizione il sistema dispone di un sistema di classi e relazioni indotti che migliora la sua interazione futura con l'ambiente operativo (ad es. l'utente)
- Il miglioramento avviene quindi almeno rispetto agli algoritmi di ricerca: la organizzazione gerarchica consente di esaminare solo i membri dell'insieme in alcune classi (i generi).



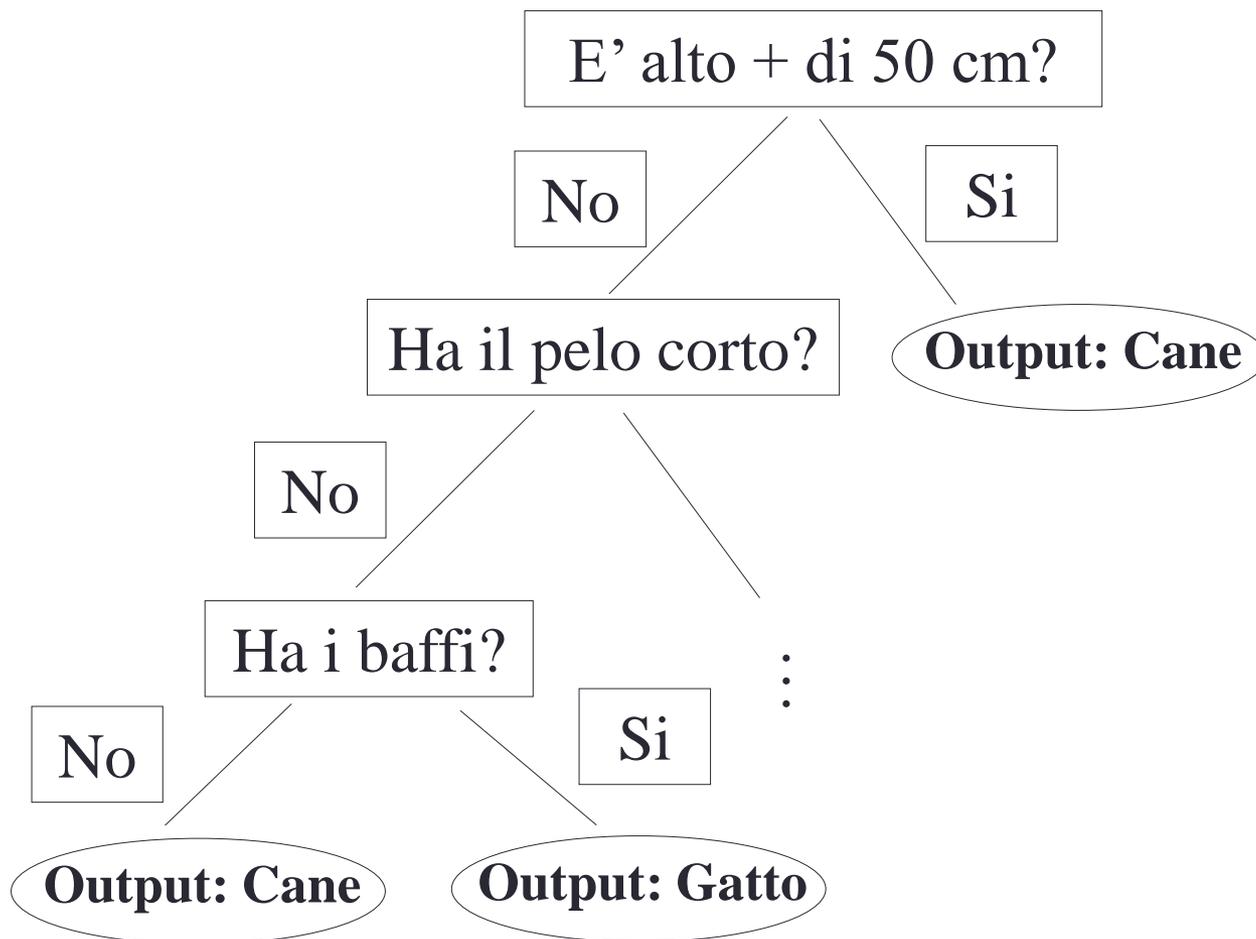
# L'apprendimento automatico

- Apprendere una funzione da esempi:
  - a valori reali, *regression*
  - a valori interi finiti, *classification*
- Supponiamo di volere apprendere una funzione intera:
  - 2 classi, *gatto e cane*
  - $f(x) \rightarrow \{\text{gatto}, \text{cane}\}$
- Dato un insieme di esempi per le due classi
  - Si estraggono le features (*altezza, baffi, tipo di dentatura, numero di zampe*).
- Si applica l'algoritmo di learning per generare  $f$

# Algoritmi di Apprendimento

- Funzioni logiche booleane, (ad es., alberi di decisione).
- Funzione di Probabilità, (ad es., classificatore Bayesiano).
- Funzioni di separazione in spazi vettoriali
  - Non lineari: KNN, reti neurali multi-strato,...
  - Lineari, percettroni, Support Vector Machines,...
- Trasformazioni di spazi: embeddings, analisi spettrale

# Alberi di decisione (Gatti/Cani)



# Web IR?

- I processi di IR studiati in domini antecedenti all'affermarsi del Web debbono essere estesi ed adattati rispetto alla maggiore ricchezza ed ai problemi maggiori che tali scenari presentano
  - Complessità strutturale: contenuti, topologia e uso
  - Affidabilità dell'informazione
  - Multimodalità, Multimedialità
  - Partecipazione (aspetti sociali)

# Web IR

- Processing Web data: content detection, link detection, ...
- Web Crawling
- Web Search: indici, link analysis
- Ranking: weighting contents, links and formats, authority, timeliness
- Meta-search
- Link Analysis

# Prospettive delle tecnologie WM&R

- Crescita esponenziale della taglia dei problemi
- Crescente interesse verso processi di IR agenti su dati complessi (multimediali, sociali)
- Web partecipativo: Web 2.0
- Ruolo crescente della mediazione degli strumenti informatici
  - Software as a Service
  - Personalizzazione
- Big Data challenges:
  - Scala
  - Opacità Semantica

# Information, Web and the language

Web contents, characterized by rich multimedia information, are mostly **opaque from a semantic standpoint**

Today is 2011年11月13日 星期日 顯示器最佳分辨率 1024X768 今日天氣 加入最愛 設為首頁 大公網新版

www.takungpao.com.hk

2011 中国証券金紫荆獎 Golden Bauhinia Awards

首頁 國內 國際 港澳 兩岸 評論 財經 體育 教育 科技 醫學 娛樂 文化 副刊 軍事 生活 旅遊 圖片 博客

關鍵詞: 欄目: 全部 最近三個月 三個月之前 檢索

▶ 手機新聞 ▶ 手機博客 ▶ 漢語學習 ▶ 新聞點擊排行

滾動新聞:

**胡總語特首:防範經濟金融風險**  
胡錦濤在夏威夷會見出席APEC峰會的曾蔭權。他祝賀香港區議會選舉成功,並充分肯定曾蔭權及港府工作,要求做好經濟金融風險防範

**胡連會登場 共同宣示九二共識**  
胡錦濤第四次在APEC峰會期間會見連戰。他強調,認同「九二共識」是兩岸開展對話協商的必要前提,也是兩岸關係和平發展的重要基礎

西藏黨代會高調反「藏獨」 德國作家:外埠雜誌報道西藏  
傳媒入日本福島核電站探險 英國大裁軍 傷兵難雜免  
滇礦難已30死 13人生還 礦工講述內幕 事故並不意外  
范徐麗泰認民望跌最不耐 選委再獲60提名表 累積逾千人  
聖保羅中學本月底截止招 選委再獲60提名表 累積逾千人  
民調逆轉 藍高層:國親吵鬧地 秋門訴求多 向藍綠表不滿  
世界新七奇觀 亞洲景佔四 新奇觀選舉惹爭議

中國實體書店苦苦掙扎求 加入TPP 台密集會談探險  
香港人家/蔡仕榮 人生導師 活出自我香港人家/教導子女...  
債務危機好 港ADR幾全線造 歐元反降 兌美元逼近1.38  
入世十年/充分對接 華強北最 入世十年/挑戰「二次...  
抽身除「雜軍」 工人險生 南亞漢命案 警日籍妻

即時新聞

- 組國/河南全國太極拳錦標賽賽
- 奧巴馬重申美不支持「台灣獨立
- 巴基斯坦西北部兩起襲擊 16人
- 圖文/胡錦濤會見美國總統奧巴
- 兩岸30對愛侶在廈門集體證婚
- 中日韓衛生部長會議在青島舉行
- 面向中國遊客中英雜誌紐約創刊
- 「CEO聖經」成內地官員考試內
- 斯特恩:經紀人是勞資談判的障
- 香港冀成爲人幣國際化關鍵角色
- 日學者提出地核物質形態新假設
- 中國影視機構向國際大師「取經

焦點關注

區議會選舉 香港

2011APEC 港黑金事件 2011

神八天宮對接 第七次陳江會 李

9.1衝擊事件 中國航母試航 辛

http://www.takungpao.com.hk/news/11/11/13/2011\_apec\_xgbd-1423309.htm

# Information, Web and language

Hu meets KMT honorary chairman in Hawaii - People's Daily Online - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Yahoo! Strumenti Aiuto

Hu meets KMT honorary chairman

Indietro Avanti Download

*Chinese President Hu Jintao (R) shakes hands with Honorary Chairman of the Chinese Kuomintang (KMT) Lien Chan, in Honolulu, Hawaii, the U.S., Nov. 11, 2011.*  
*(Xinhua/Huang Jingwen)*

HONOLULU, United States, Nov. 11 (Xinhua) -- Hu Jintao, general secretary of the Central

Latest News: • Indonesia to host European Higher Education Fair

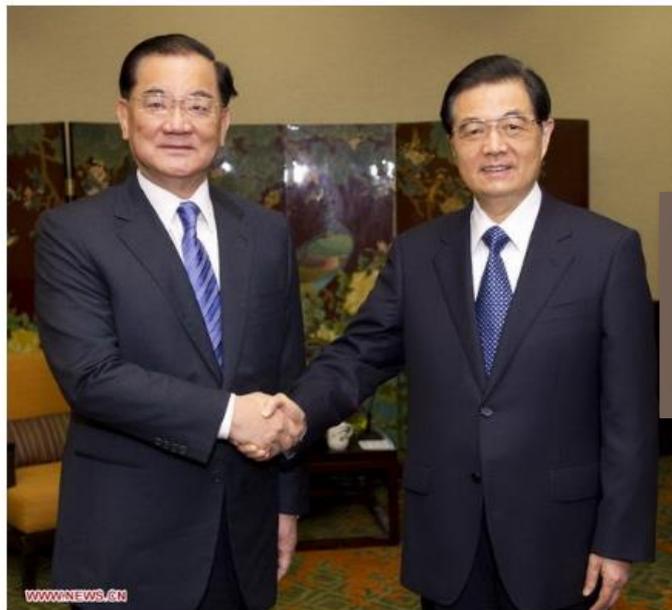
Beijing Sunny 15 / 1 City Forecast

Home >> China Politics

## Hu meets KMT honorary chairman in Hawaii

(Xinhua)

11:10, November 12, 2011 🔍 +-



*Chinese President Hu Jintao (R) shakes hands with Honorary Chairman of the Chinese Kuomintang (KMT) Lien Chan, in Honolulu, Hawaii, the U.S., Nov. 11, 2011.*

Selections for you



Miao ethnic group celebrates Miao's New Year in SW China



World's first Angry Birds exclusive shop opens in Helsinki

## Who is Hu Jintao?

Most Popular

- 1 Hu reaffirms support to Hong Kong's sta...
- 2 Hu meets KMT honorary chairman in Hawaii
- 3 China in APEC: a mutually beneficial en...
- 4 Night life in Shanghai
- 5 China's 2011 foreign trade to grow 20 p...
- 6 Beijing house prices stumble 5.1 pct as...
- 7 Lama students start school in Tibet Col...
- 8 Police in central China crack phoney ca...



Hu Jintao



Ricerca

Circa 725.000 risultati (0,09 secondi)

- Tutto
- Immagini
- Mappe
- Video
- Notizie
- Shopping
- PIÙ conte

Tutti i ri  
Per argomento

Qualsiasi dimensione

- Grandi
- Medie
- Icone
- Maggiori di...
- Dimensioni esatte...

Qualsiasi colore

- A colori
- Bianco e nero



Qualsiasi tipo

- Volti
- Foto
- Clip art
- Disegni

Visual standard

Mostra dimensioni

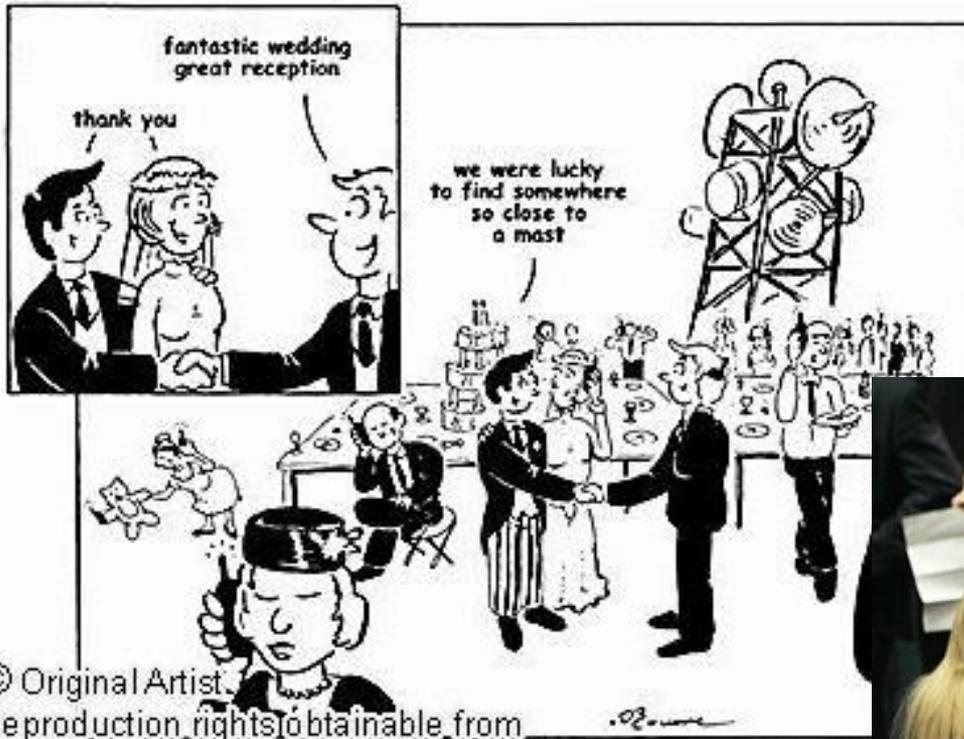


# Content Semantics and Natural Language

- Human languages are the main carrier of the information involved in processes such as *retrieval*, *publication* and *exchange* of knowledge as it is associated to the open Web contents
- Words and NL syntactic structures express concepts, activities, events, abstractions and conceptual relations we usually share through data
- “*Language is parasitic to knowledge representation languages but the viceversa is not true*” (Wilks, 2001)

# NL and Knowledge

Natural Languages are even too successful in modern communication



# Natural Languages: undesirable aspects



Some crucial language understanding problems affect the comprehension of human speakers:

- **Noise**
- **Incompleteness**
- **Ambiguity**

... and for machines and software systems they are even more critical



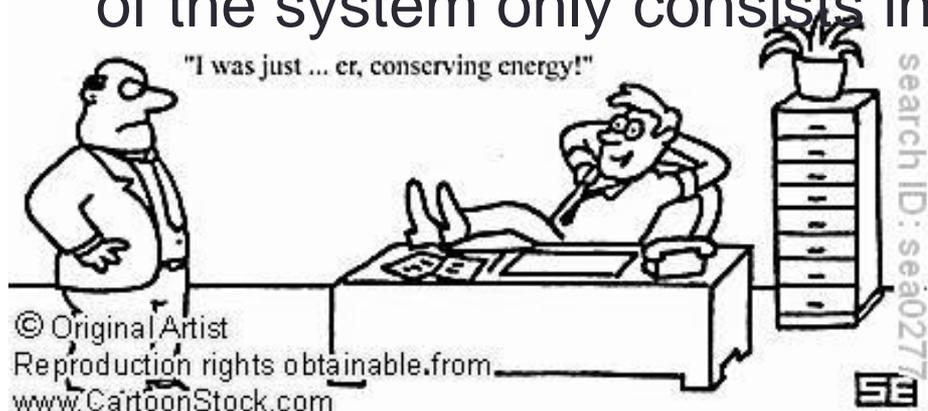
- Since '90s successful methods inspired by data-driven paradigms allow a quantitative (i.e. numerical) approach:
  - Markov models of languages and communication processes
  - Neural nets approaches (annotated examples to induce and optimize robust and large scale models)

# Inductive Pattern Recognition

- The technology: **Data mining**...
  - Exploration and analysis on large amounts of data in order to discover patterns of interest
- ... and **Machine Learning (ML)**
  - algorithms that from sets of well defined classification examples (known cases) induce methods to classify unknown cases
- Machine Learning suggests that **data-driven algorithms** are
  - more **robust**
  - more **accurate** (performing better on the more important cases),
  - **flexible** and **scalable** wrt optimisation, specialization and maintenance
- **Success stories** in Language Processing:
  - **Machine Translation** (see DARPA MT competitions)
  - Syntactic Analysis, Sense disambiguation
  - **Semantic Interpretation** (see SemEval competitions held in US)

# Benefits of a data-driven approach

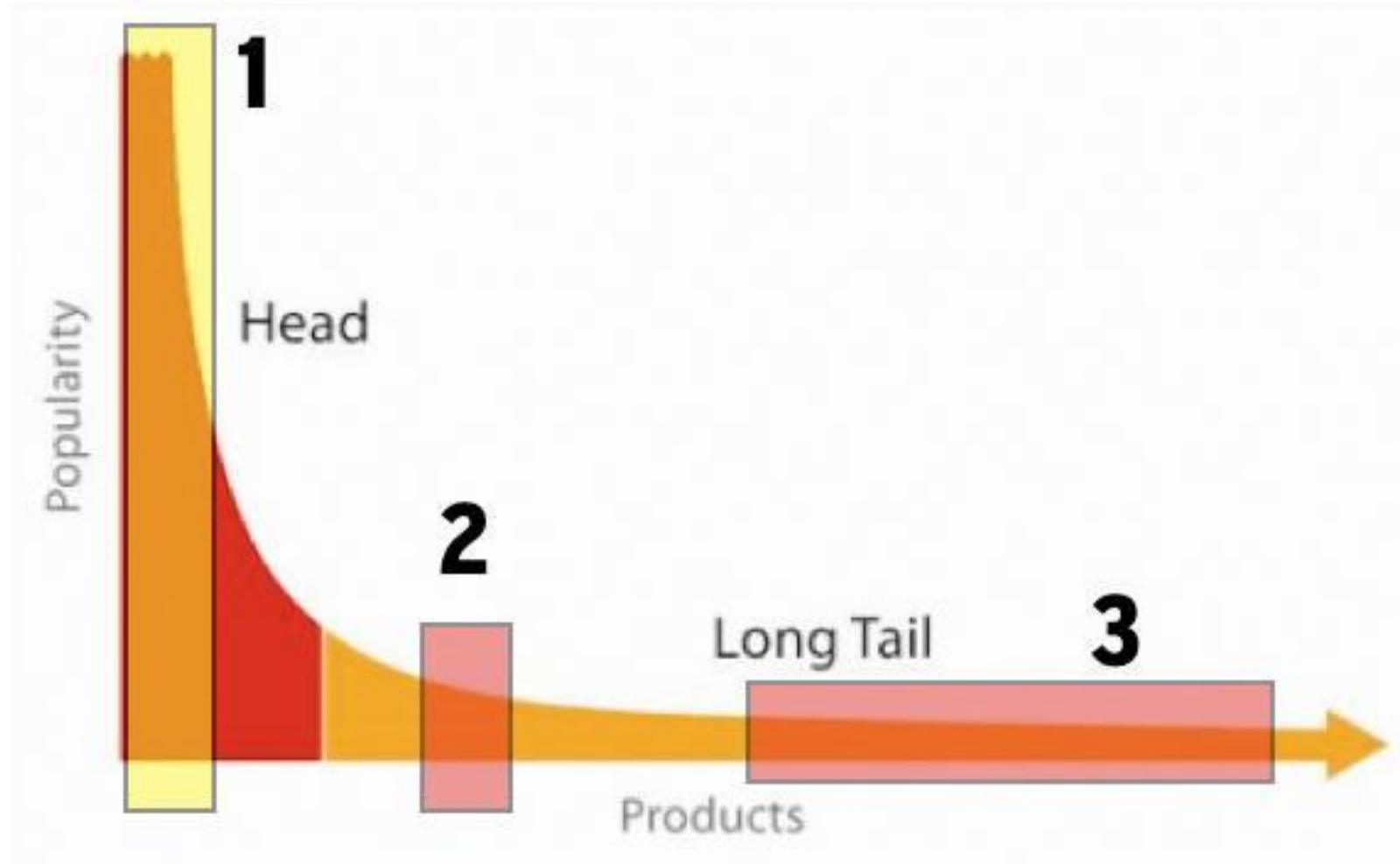
- Very effective learning algorithms available (e.g. Support Vector Machines)
- The ML technology is *portable* while imperative coding is *task (i.e. scenario) specific*
- Very accurate solutions can be obtained
- Gathering training data *much less expensive* than rule coding
- In dynamically evolving scenarios, incremental refinement of the system only consists in re-training



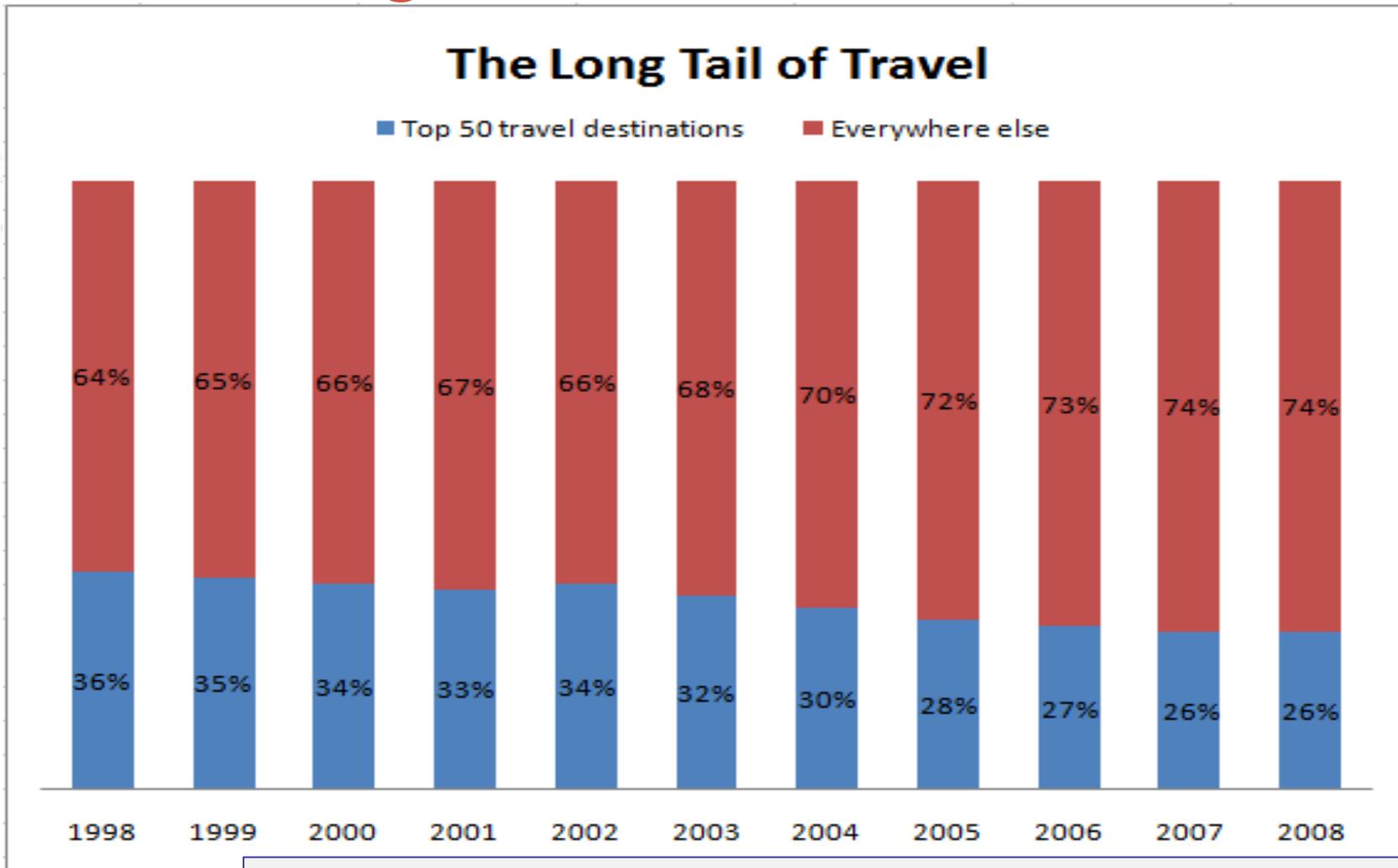
# Data Mining: perspectives and benefits

- Technical advantages
  - Self adaptivity to changing operational conditions (i.e. domain)
  - Better SW management and incremental maintenance
  - More flexibility for special-purpose versioning:
    - No need for re-engineering or independent software developments
    - Just new domain-specific examples are needed
- Cost benefits
  - The data-driven approach has been shown to reduce the development costs up to 80-90% in several NLP tasks
- Market benefits
  - Reduced time-to-market
  - Competitive advantages: the lack of similar products makes the system targeted strongly competitive solutions

# La lunga Coda



# The Long Tail

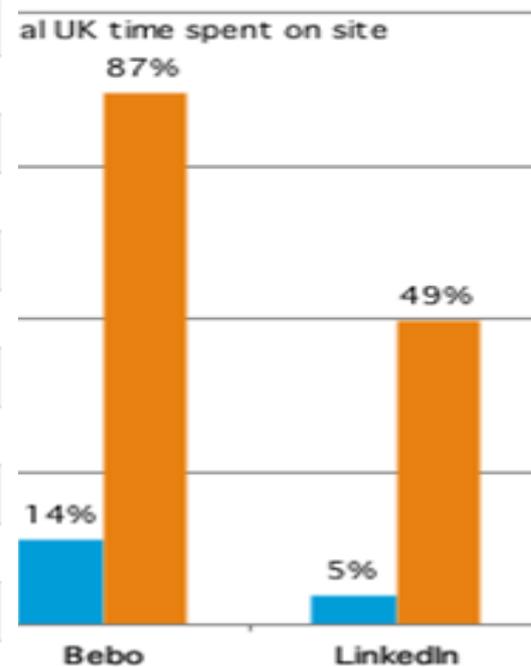


Maren Jinnett over data compiled by the [UK's Civil Aviation Authority](#). (Wired Blog network, Oct 2009)

## Top 20 Websites

The following report shows **websites** for the industry 'All Categories', ranked by **Visits** for the week ending **06/05/2010**.

| Rank | Website           | Visits |
|------|-------------------|--------|
| 1.   | Facebook          | 8.63%  |
| 2.   | Google            | 7.23%  |
| 3.   | Yahoo!            | 3.76%  |
| 4.   | Yahoo! Mail       | 3.68%  |
| 5.   | YouTube           | 2.68%  |
| 6.   | MySpace           | 1.88%  |
| 7.   | msn               | 1.80%  |
| 8.   | Windows Live Mail | 1.63%  |
| 9.   | Yahoo! Search     | 1.30%  |
| 10.  | Bing              | 1.25%  |
| 11.  | Gmail             | 0.86%  |
| 12.  | AOL               | 0.79%  |
| 13.  | eBay              | 0.79%  |
| 14.  | Aol Mail          | 0.57%  |
| 15.  | My Yahoo!         | 0.43%  |
| 16.  | Wikipedia         | 0.42%  |
| 17.  | Yahoo! News       | 0.39%  |
| 18.  | Amazon.com        | 0.35%  |
| 19.  | Pogo              | 0.32%  |
| 20.  | Google Maps       | 0.29%  |



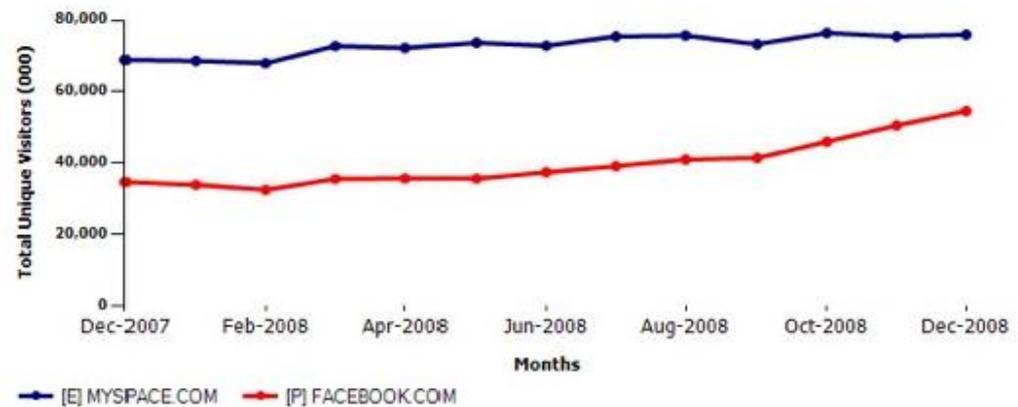
Source: The Nielsen Company/UKOM

# Social Web

## 2008 U.S. Social Network Usage (Comscore)

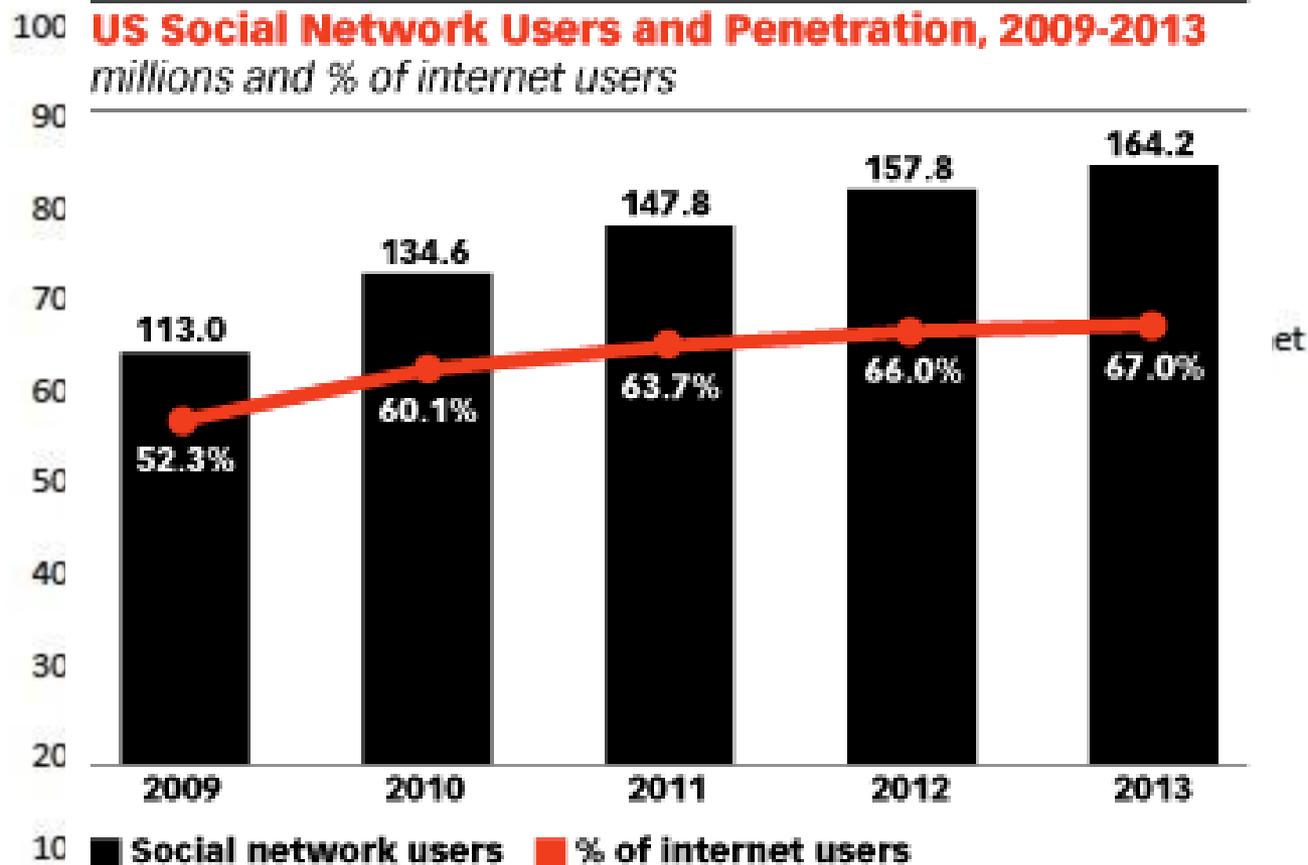
|            | 12/1/2007<br>(millions) | 12/1/2008<br>(millions) | Yearly<br>Growth | Monthly<br>Growth |
|------------|-------------------------|-------------------------|------------------|-------------------|
| MySpace    | 69                      | 76                      | 10%              | 0.8%              |
| Facebook   | 35                      | 55                      | 57%              | 3.8%              |
| Classmates | 10                      | 16.6                    | 66%              | 4.3%              |
| LinkedIn   | 2.9                     | 6.3                     | 117%             | 6.7%              |
| Bebo       | NA                      | 4.9                     |                  |                   |
| Ning       | 0.8                     | 3.9                     | 388%             | 14.1%             |
| Friendster | 1.8                     | 1.7                     | -6%              | -0.5%             |

### MEDIA TREND REPORT



# Social Networking Site Use by Age Group, 2005-2012

% of internet users in each age group who use social networking sites



Note: internet users who use social networks via any device at least once per month

Source: eMarketer, Feb 2011

124527

www.eMarketer.com

# Social Web

## US Internet Users Who Use Social Networks, by Generation, Dec 2008 & May 2010

% of respondents in each group

### Millennials (18-33)



### Gen X (34-45)



### Younger boomers (45-55)



### Older boomers (55-64)



### Silent generation (65-73)



### G.I. generation (74+)



### Total (18+)



■ Dec 2008

■ May 2010

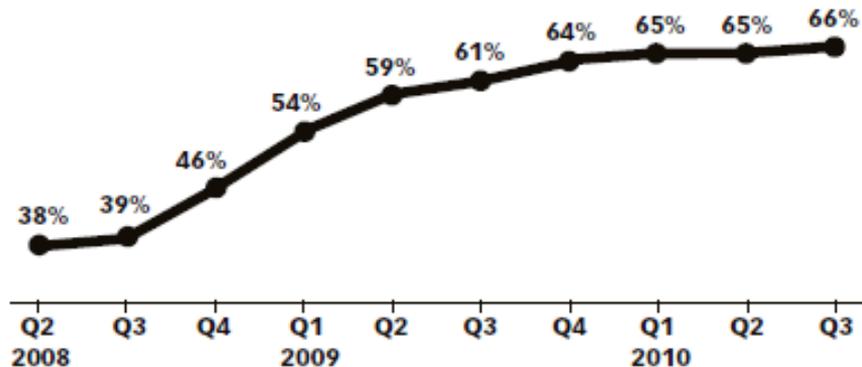
Note: n=2,252

Source: Pew Internet & American Life Project, "Generations 2010," Dec 16, 2010

# Social Web

## US Consumers Who Spend 1+ Hour on Social Network Activities per Week, Q2 2008-Q3 2010

% of respondents



Source: Interpret, "New Media Measure, Q3 2010" as cited by Social Times, Dec 16, 2010

123195

www.eMarketer.com

## Daily Online Activities of US Internet Users, by Age, May 2010

% of each group

### Send or read email



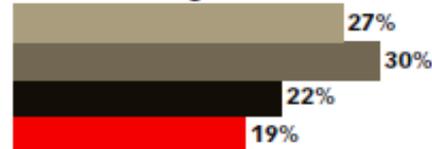
### Social networking



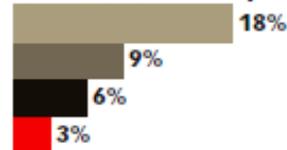
### Get news



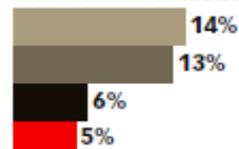
### Online banking



### Twitter or status update service



### Use online classifieds



18-29

30-49

50-64

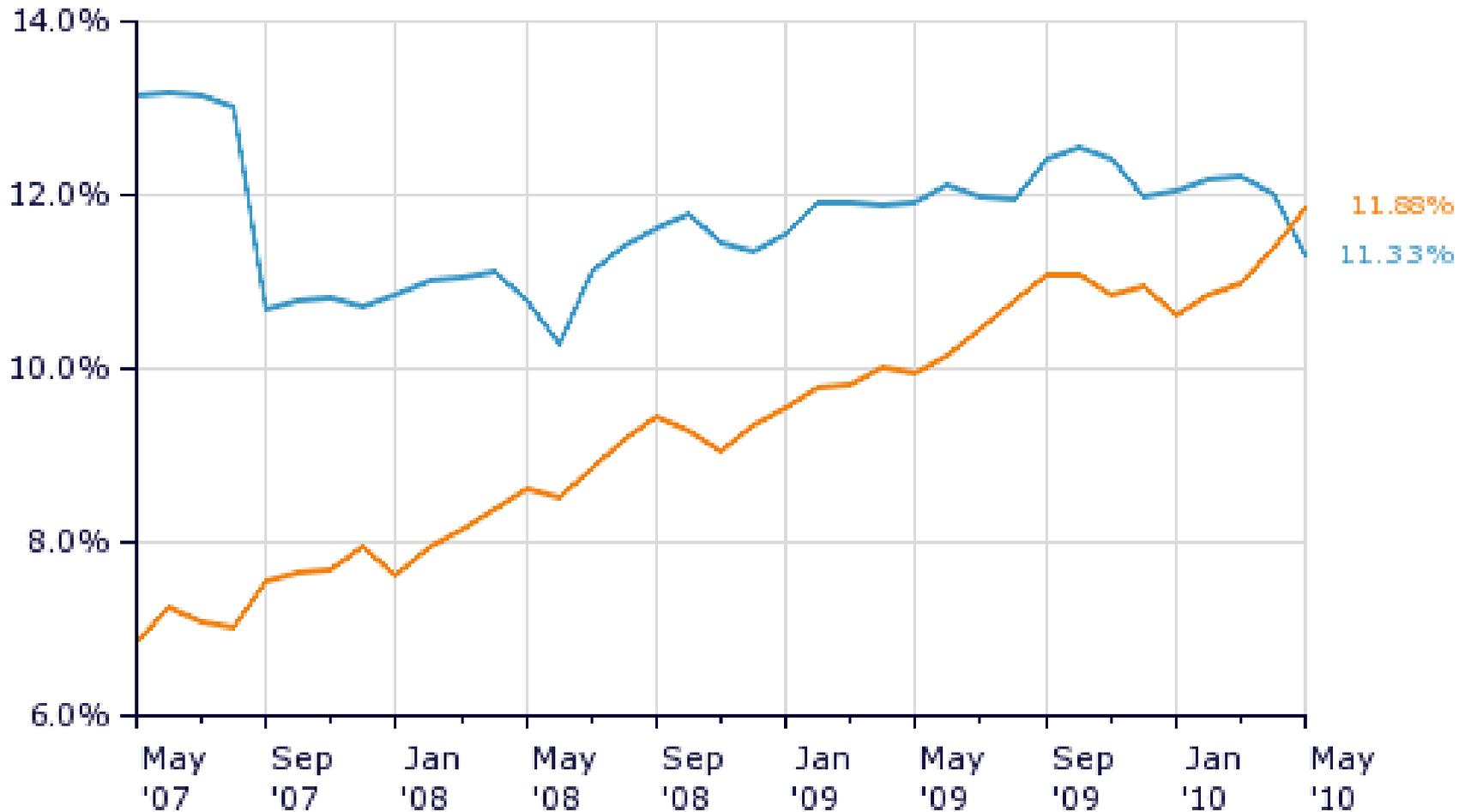
65+

Source: Pew Internet & American Life Project, "Older Adults and Social Media," Aug 27, 2010

119154

www.eMarketer.com

## UK Internet visits to Social Networks and Search Engines

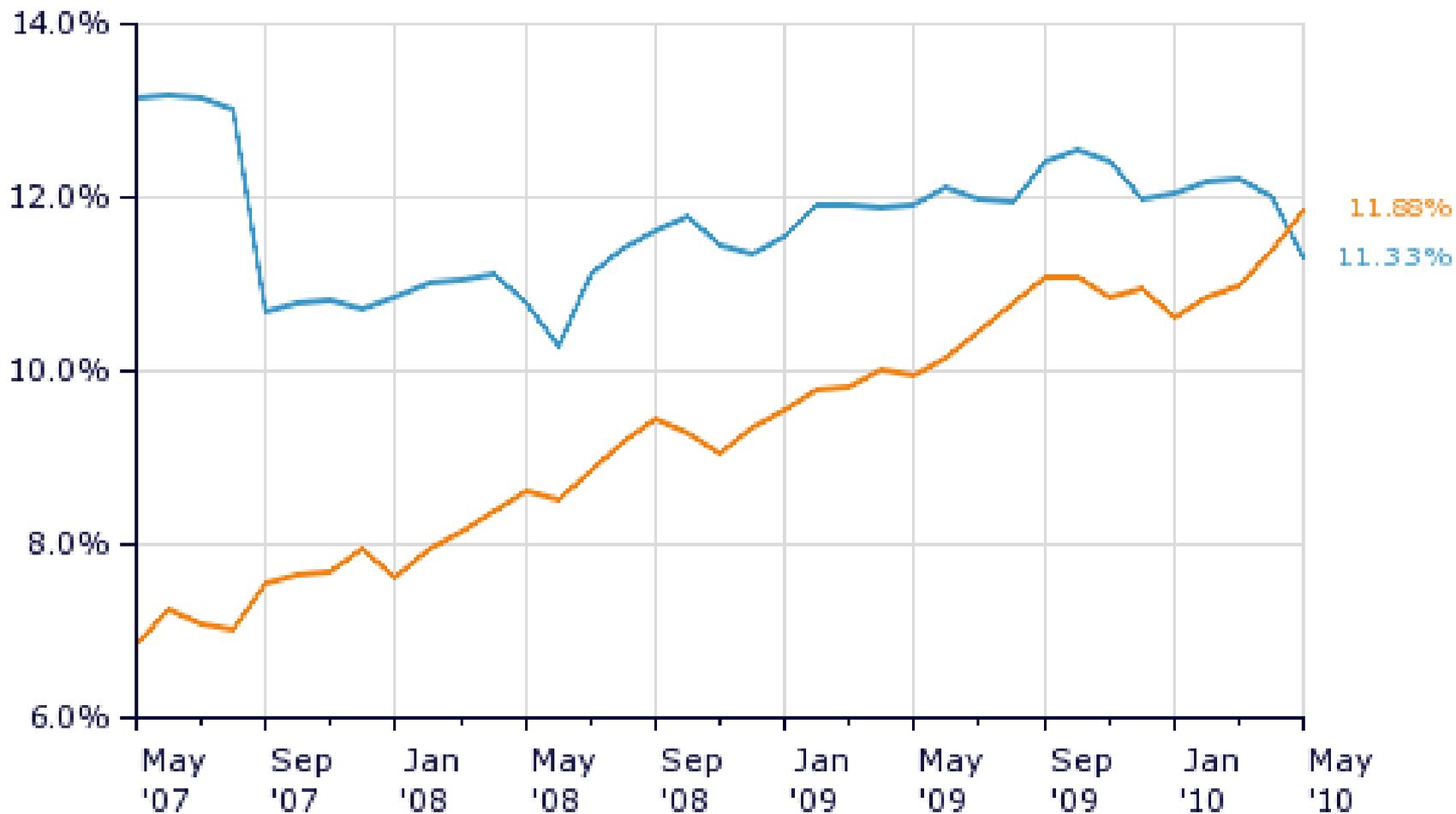


- Computers and Internet - Search Engines
- Computers and Internet - Social Networking and Forums

Monthly market share in 'All Categories', measured by visits, based on UK usage.

Created: 03/06/2010. © Copyright 1996-2010 Hitwise Pty. Ltd. Source: Experian Hitwise UK

## UK Internet visits to Social Networks and Search Engines



- Computers and Internet - Search Engines
- Computers and Internet - Social Networking and Forums

Monthly market share in 'All Categories', measured by visits, based on UK usage.

Created: 03/06/2010. © Copyright 1996-2010 Hitwise Pty. Ltd. Source: Experian Hitwise UK



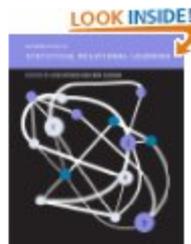
amazon.com

### Recommended for You

Amazon.com has new recommendations for you based on [items](#) you purchased or told us you own.



[Collective Intelligence in Action](#)



[Introduction to Statistical Relational Learning \(Adaptive Computation and Machine Learning\)](#)



[Quantitative Methods In Linguistics](#)

[See More Recommendations](#)

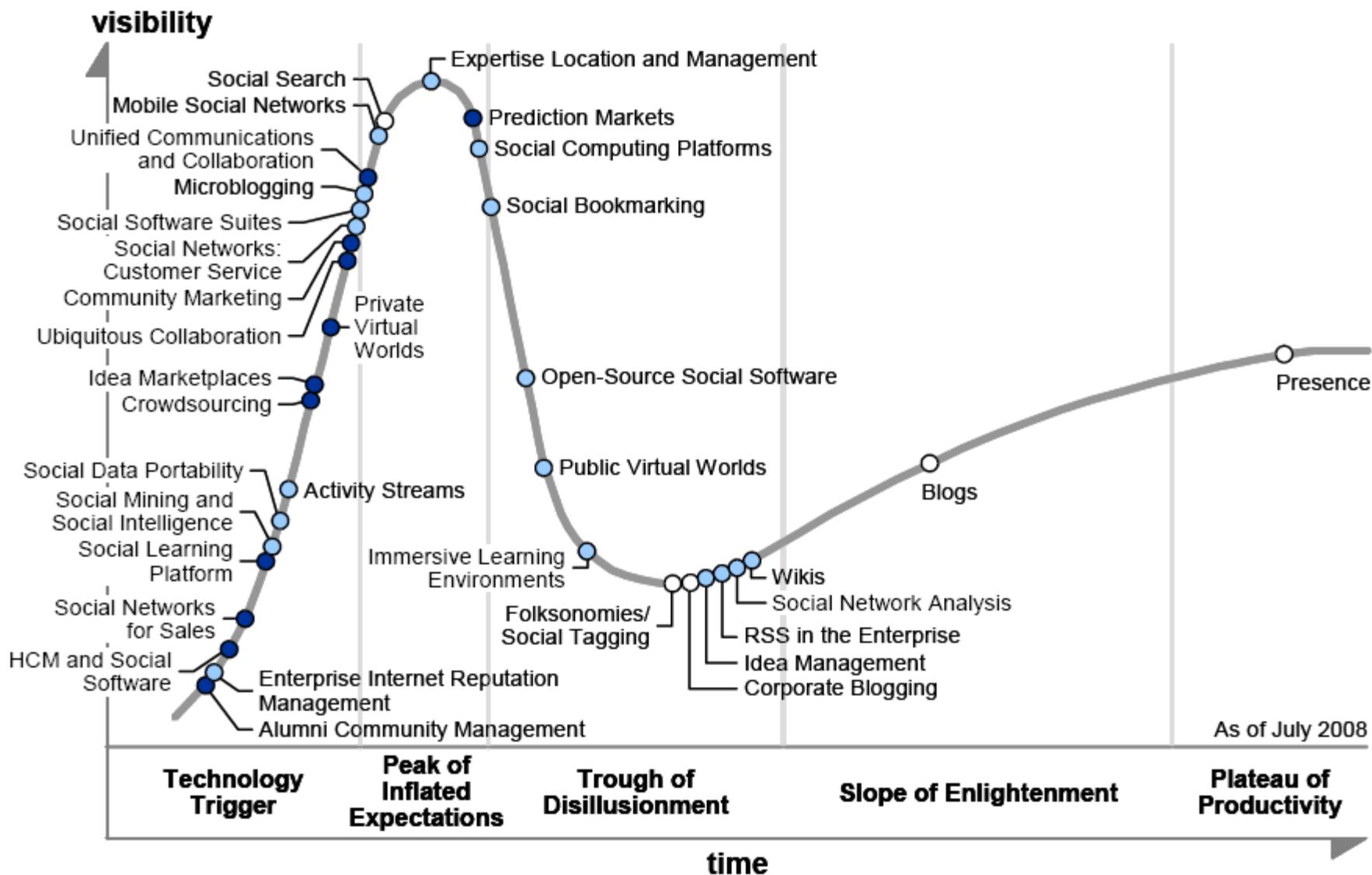


**[Collective Intelligence in Action](#)**  
by Satnam Alag  
Average customer review: ★★★★★

[Add to cart](#)

[Add to Wish List](#)

Figure 1. Hype Cycle for Social Software, 2008



**Years to mainstream adoption:**

○ less than 2 years

● 2 to 5 years

● 5 to 10 years

▲ more than 10 years

⊗ obsolete before plateau

Source: Gartner (July 2008)

# Natural Language Parsing

UK Economy News Headlines - FT.com - Mozilla Firefox

File Modifica Visualizza Cronologia Segnalibri Strumenti Aiuto

http://www.ft.com/world/uk/economy

Più visitati Corso: Basi di dati Gruppi Posta :: Benvenuto a H... ClustrMaps - map of vi... UniversitaCedol Tree Kernels in SVM-lig... Net RicercaAteneo Keysrc Calls EMEROTECA GEMS2010

The image shows a natural language parsing tree for the sentence "Mortgage approvals fell sharply in June, lending yet more weight to the theory...". The tree is rooted at the top with a node labeled "1.00 V\_PP". This node branches into "1.00 V\_Sog" and "1.00 V\_Adv". "1.00 V\_Sog" branches into "Mortgage\_approvals" (type Nom) and "1.00 V\_Adv". "1.00 V\_Adv" branches into "fell" (type VerFin) and "1.00 Adv\_PP". "1.00 Adv\_PP" branches into "sharply" (type Adv) and "in\_June" (type Prep). "in\_June" branches into "in" (type IN) and "'June'" (type NNP). "lending\_yet\_more\_weight" (type Nom) branches into "lending" (type NN) and "yet" (type RB). "to\_the\_theory" (type Prep) branches into "to" (type CongCo) and "more" (type JJR). Each node is expanded to show its constituent words and their grammatical types and morphological features.

Mortgage\_approvals  
type Nom

Mortgage  
type NNP  
morph mas.fem.sing.

approvals  
type NNS  
morph mas.fem.plur.

fell  
type VerFin  
Sentence

sharply  
type Adv

in\_June  
type Prep

in  
type IN  
morph invariante

'June'  
type NNP  
morph mas.fem.sing.plur.

lending\_yet\_more\_weight  
type Nom

lending  
type NN  
morph mas.fem.sing.

yet  
type RB  
morph invariante

more  
type JJR  
morph mas.fem.plur.sing.

to\_the\_theory  
type Prep

to  
type CongCo

Britain's place in the world, and how far it has travelled since 1947 - Jul-29

**Gilts lose lustre for overseas investors**  
Flight from eurozone risk to UK government bonds is moderating - Jul-29

with Alex Barker and Jim Pickard

Mechanical & Electrical Engineering  
Deputy Director of Finance  
London Ambulance Service  
RECRUITERS

http://www.ft.com/westminster

Italiano (Italia)

Today is: 2006-07-06 17:14:55

00:00:02:39

Timeline: [Progress bar]

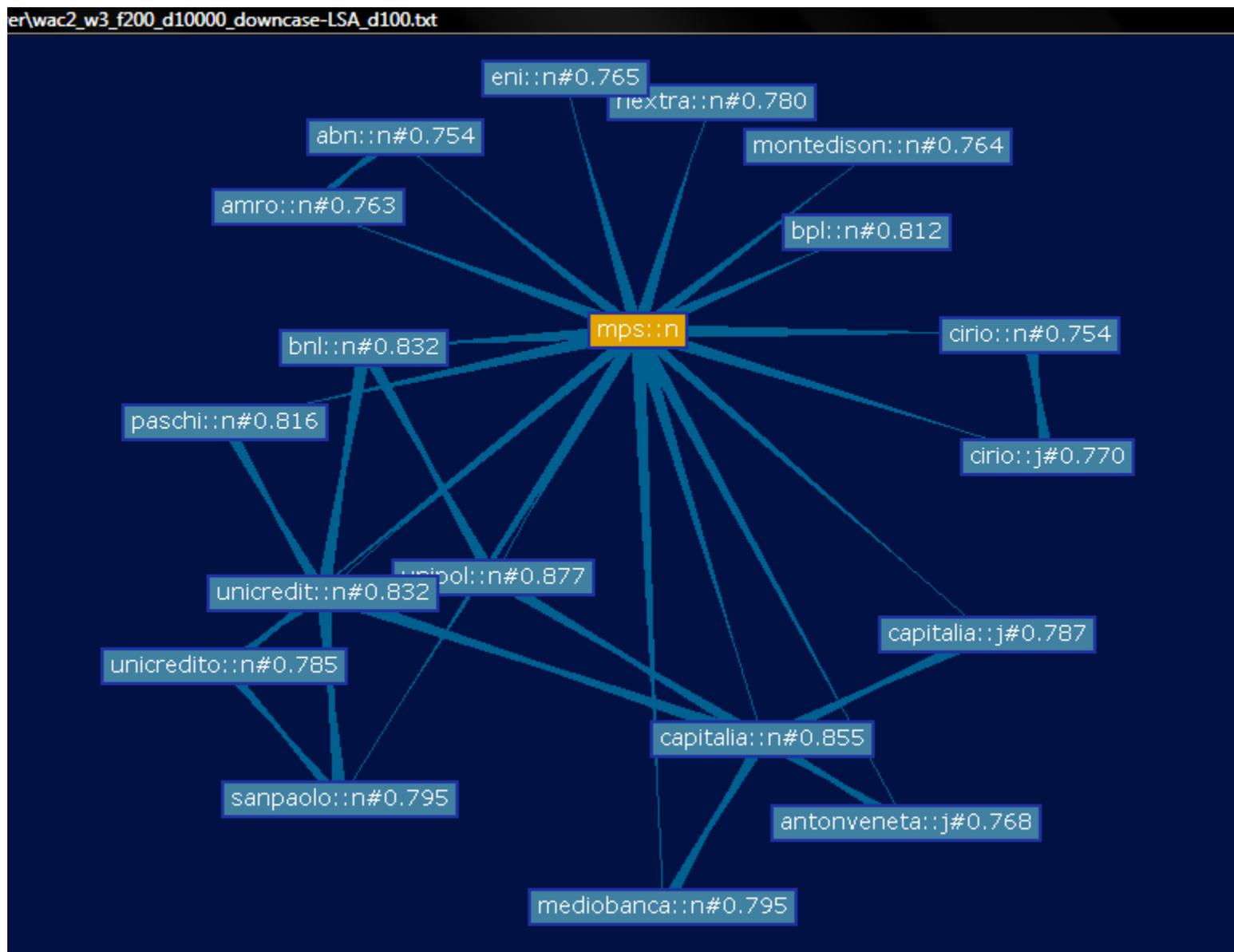
| Info     | Transcription                                                                                                                                                                                                                                                                                                                                                                                                                                | Semantic Analysis | Content Analysis |
|----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|------------------|
| 00:06:36 | chamonix <b>america</b> dove perde forza ma fa sempre paura l' uragano di mallarme <b>italia</b> andiamo nel centro                                                                                                                                                                                                                                                                                                                          |                   |                  |
| 00:06:41 | che in florida riguardasse cento km di costa sull' atlantico si e' formata nel frattempo un' altra tempesta tropicale                                                                                                                                                                                                                                                                                                                        |                   |                  |
| 00:06:52 | ha lasciato una riviera messicana dello jucker puntando verso la florida l' uragano delle corde <b>wilma</b> il dodicesimo ciclone di una stagione eco dell' atmosfera piu' di qualcuno lavatrici su strada a festeggiare lo scampato pericolo mentre dall' altra l' emergenza ha segnato l' inizio dei sa scarseggiano cibo e acqua si e' costretti a fare i conti con la sopravvivenza ad attraversare queste strade inondate sferzata dal |                   |                  |
| 00:07:22 | vento la pioggia per raggiungere i centri della croce rossa vengono distribuiti ieri alla popolazione <b>dino risi</b> ma ha lasciato otto vittime soltanto migliaia di casi devastato la rete ospedaliera abbattuto centrali elettriche che ha causato danni a un milione di persone in florida e' attesa per                                                                                                                               |                   |                  |
| 00:07:44 | e nelle isole di <b>kiss</b> e' gia' iniziata la grande fuga non bastasse sull' atlantico a sud di porto rico si e' formata falla venti di <b>hemingway</b> le                                                                                                                                                                                                                                                                               |                   |                  |
| 00:07:52 | nessuna tempesta tropicale della stagione la buona notizia che dovrebbe essere innocua la brutta notizia che la stagione degli uragani                                                                                                                                                                                                                                                                                                       |                   |                  |
| 00:07:59 | non e' ancora finita nulla fino al trentanove e c' e' stato una sciagura                                                                                                                                                                                                                                                                                                                                                                     |                   |                  |
| 00:08:05 | in nigeria                                                                                                                                                                                                                                                                                                                                                                                                                                   |                   |                  |

- TG1 - 2005-10-23
- Other Classification
- Other Classification
- Other Classification
- Ambiente, Natura e Territorio
- Ambiente, Natura e Territorio**
- Other Classification
- Politica, Partiti, Istituzioni e Sindacati
- Politica, Partiti, Istituzioni e Sindacati
- Other Classification
- Other Classification
- Other Classification
- Other Classification
- Usi e costumi
- Other Classification
- Sanita' e Salute
- Giustizia, Criminalita' e Sicurezza
- Other Classification
- Other Classification
- Giustizia, Criminalita' e Sicurezza
- Other Classification
- Other Classification
- Musica e Spettacolo
- Sport
- Sport

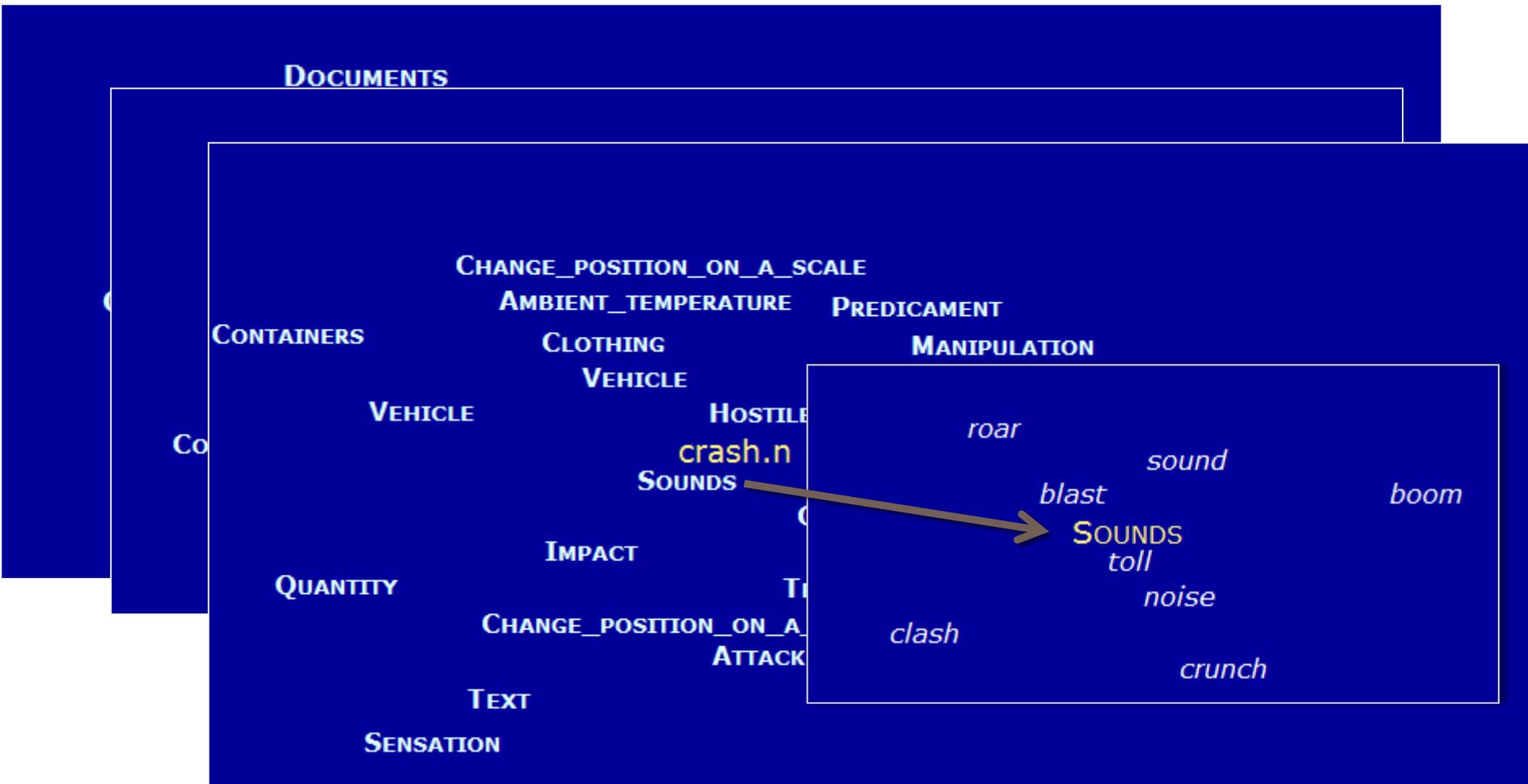
TIMELINE

|             |             |             |             |             |             |
|-------------|-------------|-------------|-------------|-------------|-------------|
| 00:06:36:10 | 00:06:51:07 | 00:06:54:20 | 00:06:56:01 | 00:06:57:10 | 00:07:00:08 |
| 00:07:08:18 | 00:07:12:02 | 00:07:15:23 | 00:07:19:13 | 00:07:22:02 | 00:07:26:06 |
| 00:07:37:08 | 00:07:39:15 | 00:07:45:20 | 00:07:55:06 | 00:07:57:07 | 00:07:59:13 |
| 00:08:05:00 | 00:08:05:00 | 00:08:05:00 | 00:08:05:00 | 00:08:05:00 | 00:08:05:00 |

# Vector Spaces for Lexical Semantics



# Spaces for NL predicates





Home

Preferences

Links

Documents

Contact

Roberto Basili

Search

### All results (99)

- » [alessandro moschitti \(27\)](#)
- » [maria teresa pazienza \(25\)](#)
- » [tor vergata \(12\)](#)
- » [department of computer science systems \(9\)](#)
- » [artificial intelligence and human oriented \(6\)](#)
- » [database of individual seismogenic \(2\)](#)
- » [vigna murata 605 it 00143 \(2\)](#)
- » [fault mapper \(2\)](#)
- » [musical genre a machine learning \(2\)](#)
- » [roberto basili fabio \(2\)](#)

[More clusters](#)

YOU ARE IN "ALESSANDRO MOSCHITTI" CLUSTER WITH 27 DOCUMENTS

#### DBLP: ROBERTO BASILI

Paolo Annesi, **Roberto Basili**: Cross-Lingual Alignment of FrameNet Annotations through Hidden Markov Models. ... **Roberto Basili**, Cristina Giannone, Chiara Del Vescovo, Alessandro ...

<http://dblp.uni-trier.de/db/indices/a-tree/b/Basili:Roberto.html>

#### PUBZONE - ROBERTO BASILI

**Roberto Basili**, Danilo Croce, Cristina Giannone, Diego De Cao. ... **Roberto Basili**, Cristina Giannone, Chiara Del Vescovo, Alessandro Moschitti, Paolo Naggari. Kernel-Based ...

<http://www.pubzone.org/pages/publications/showAuthor.do?userId=79.2000>

#### LEARNING DOMAIN-SPECIFIC FRAMENETS FROM TEXTS

Marco Pennacchiotti, Diego De Cao, Paolo Marocco, **Roberto Basili**, ... Marco Pennacchiotti, Diego De Cao, **Roberto Basili**, Danilo Croce, Michael Roth, Automatic ...

<http://olp.dfki.de/olp3/Basili.pdf>

#### DIEGO DE CAO HOME PAGE

**Roberto Basili**, Danilo Croce, Diego De Cao, and Cristina Giannone. ... **Roberto Basili**, Diego De Cao, Danilo Croce, Bonaventura Coppola, and Alessandro Moschitti. ...

<http://art.uniroma2.it/decao/>



### All results (98)

- » [computer](#) (33)
- » [learning algorithm for machine](#) (11)
- » [machine learning course](#) (6)
- » [machine learning group](#) (6)
- » [introduction to machine learning](#) (3)
- » [machine learning applications](#) (3)
- » [artificial intelligence](#) (7)
- » [applying machine learning](#) (3)
- » [challenges machine learning](#) (3)
- » [machine learning the study](#) (5)

[More clusters](#)

TOP 98 RESULTS OF RETRIEVED FOR THE QUERY MACHINE LEARNING

### MACHINE LEARNING - WIKIPEDIA, THE FREE ENCYCLOPEDIA

**Machine learning** is a scientific discipline that is concerned with ... A major focus of **machine learning** research is to automatically learn to recognize complex ...

[http://en.wikipedia.org/wiki/Machine\\_learning](http://en.wikipedia.org/wiki/Machine_learning)

### MACHINE LEARNING: DEFINITION FROM ANSWERS.COM

**machine learning** ( m??sh?n ?l?rni? ) ( computer science ) The process or technique by which a device modifies its own behavior as the result of its

<http://www.answers.com/topic/machine-learning>

### CMU 15-859 MACHINE LEARNING THEORY

Course description: This course will focus on theoretical aspects of **machine learning**. We will examine questions such as: What kinds of guarantees ...

<http://machinelearning.com/>

### THE INTERNATIONAL MACHINE LEARNING SOCIETY - ABOUT

The International **Machine Learning** Society is a non-profit organisation whose main aim is to foster **machine learning** research and whose main activity ...

<http://www.machinelearning.org/>

### MACHINE LEARNING: WEEKLY STUDY GUIDE

Weekly study guide for the course on **Machine Learning** taught by Vasant Honavar at Iowa

# Semantics and News

Applicazioni Risorse Sistema mar 27 lug, 23.47 dan

Gmail ... x SRL\_EN x Come ... x R Econo... x Googl... x Tanl It... x Frame... x SRL\_EN x Econo... x

file:///home/danilo/Downloads/SRL\_ITA/sorgente/Economia%20-%20Repubblica.it.html

Telefilm in stream... Telefilm in stream... Flash Forward pri... Telefilm in stream... Cronologia Altri Pr



L'ad punta a nuove regole sulla base del modello Pomigliano. L'annuncio, che prevede l'uscita da Federmeccanica, domani al vertice con il governo o giovedì con una lettera a Bombassei. Potrebbe avvenire assieme alla decisione di creare una new company per

Pomigliano di SALVATORE TROPEA

**Cisl-Uil: "L'accordo di categoria non si tocca"** di S. PAROLA

**Sacconi: "Su Fiat partita aperta"**

**Nasce Fabbrica Italia Pomigliano**

## Si dimette il capo di Bp buonuscita un milione di sterline



Oggi l'annuncio: a Tony Hayward subentrerà il direttore esecutivo Robert Dudley. **I costi legati al disastro sono saliti a 32,2 miliardi di dollari**, ma la società li deterrà evitando di versare al fisco Usa 10 miliardi

## Manager Usa, è Ellison di Oracle il più pagato del decennio



Ha guadagnato 1,84 miliardi di dollari. Nella classifica del *Wall Street Journal* sui leader delle società quotate, secondo con 1,14 miliardi il capo di Expedia, terzo Irani di Occidental Petroleum. Solo quarto Steve Jobs

### Il nemico alle porte

#### La Consob e la mano invisibile

Altri articoli

### PICCOLE GRANDI IMPRESE

DI LUCA PAGNI

#### La grande sfida del teleshopping

#### La crisi colpisce anche i porti turistici ma siamo sicuri che sia un male?

Altri articoli

### PERCENTUALMENTE DI ROSARIA AMATO

#### La prova del 9

#### L'export risolve il Pil, ma non le famiglie

Altri articoli

### GLI ESPERTI RISPONDONO

### CASA

A cura di Antonella Donati

#### Compenso extra, quando ne ha diritto l'amministratore

Mia moglie ed il fratello sono proprietari di un appartamento in condominio. Allo stato

Il tuo libro arriva dove hai sempre sognato.

ilmiolibro.it

### 24ORE AGI

**Roma 19:04**  
ACEA: NEL I SEMESTRE UTILE NETTO +52,1% A 20, MLN

**Parigi 18:42**  
AIR FRANCE-KLM: TORNA IN UTILE NEL PRIMO TRIMESTRE

← 3 → Le altre not

### CREDITO ALLE IMPRESE

#### Microimprese: con la crisi aumenta il rischio di credito

IN COLLABORAZIONE CON

# References

- Mitchell, Tom. M. 1997. *Machine Learning*. New York: McGraw-Hill.
- [Kernel machines, neural networks and graphical models](#), P. Frasconi, A. Sperduti, A. Starita, Rivista AI\*IA Numero speciale per i “50 anni di IA”, 2007.
- Very good video lectures by Andrew Ng (Stanford)  
<http://academicearth.org/courses/machine-learning>