CORSO DI
WEB MINING E RETRIEVAL
- INTRODUZIONE AL WM -

Corso di Laurea in Informatica, Ing. Internet, Ing. Informatica, Ing. Gestionale

(a.a. 2015-2016)

Roberto Basili

Overview

- Web Mining & Retrieval: Motivazioni e prospettive
  - Web, User-generated contents, Social Media
  - The role of learning
  - What is Machine Learning?
  - Data-driven algorithms: sources of complexity
- Main Applications
  - Intelligent Web Search
  - User Profiling for Marketing or Brand reputation management
  - Web Recommending
  - Spoken Dialogue Interaction in Robotics or in Web/mobile Interfaces

Do 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 …
  - Communities play a special role in circulating information
  - Search is the mostly collective function used (e.g. # in tweets)
  - Opinion Mining and Sentiment Analysis are important for individuals as well as organisations

WM&R: Motivazioni

- *Cos'e' 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'e' 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?

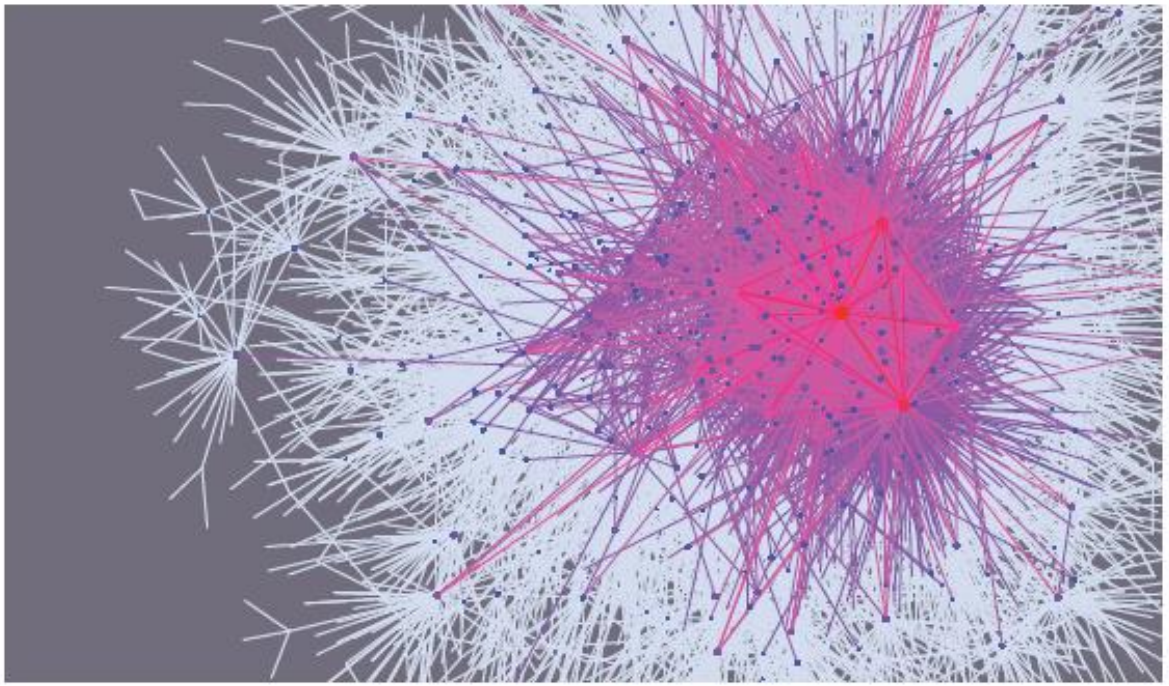
- La taglia delle informazioni in gioco pone il problema della *localizzazione*
- Accedere in modo automatico è possibile solo governando il problema di sapere **dove** si trova una informazione *rilevante*
- La ricerca corrisponde al calcolo di una funzione *aleatoria* di mapping tra requisiti e informazione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

Machine Learning vs IR?

- La eterogeneità delle informazioni produce significativi effetti di incertezza nel processo di ricerca riguardo ad aspetti diversi del processo di IR
  - Incompletezza della informazione:
    - Query brevi come informazione (incomplete) sui fabbisogni informativi
  - Ricchezza di dati, formati e modalità di accesso
    - I contenuti sono sparsi in diverse forme nei dati
  - Requisiti vaghi
    - Spesso molte informazioni sono esplicite solo nel contesto
  - Aspetti soggettivi
    - La rilevanza dipende dallo user e non solo dal contenuto
  - Tempestività ed autorevolezza

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 ΔP 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, …).

ML: una introduzione visuale

- See URL: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/?imm_mid=0d76b4&cmp=em-data-na-na-newsltr_20150826

Algoritmi di Apprendimento

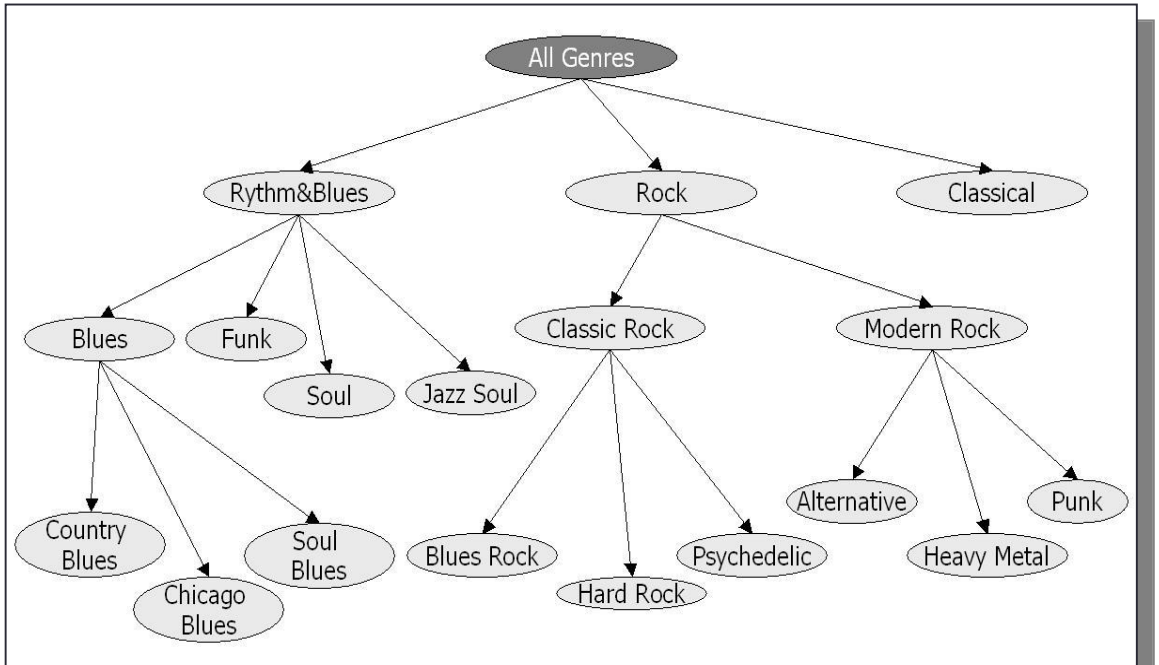
- Funzioni logiche booleane, (ad es., alberi di decisione).
- Funzione di Probabilità, (ad es., classificatore Bayesiano).
- Funzioni di separazione in spazi vettoriali
  - Non lineari: KNN, reti neurali multi-strato,…
  - Lineari, percettroni, Support Vector Machines,…
- Trasformazioni di spazi: embeddings,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

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

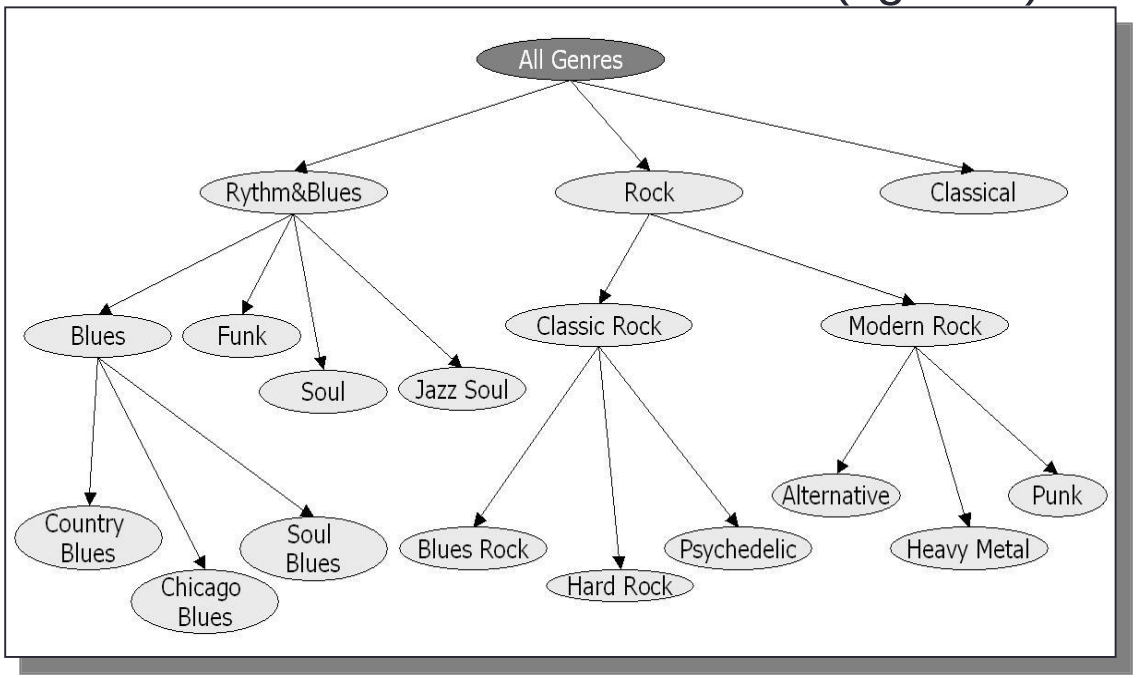
- Esempio:
  - una collezione mp3 può essere organizzata in generi attraverso il raggruppamento di brani simili secondo proprietà audio (*clustering*): tale organizzazione è naturalmente gerarchica
  - Il miglioramento avviene quindi almeno rispetto agli algoritmi di ricerca: la organizzazione gerarchica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



Apprendimento senza supervisione

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)

- Ese
  - u
    a
    a
    g
  - 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).



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

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 Down

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...

+Tu Web Immagini Video Mappe News Gmail Altro -

Google

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...

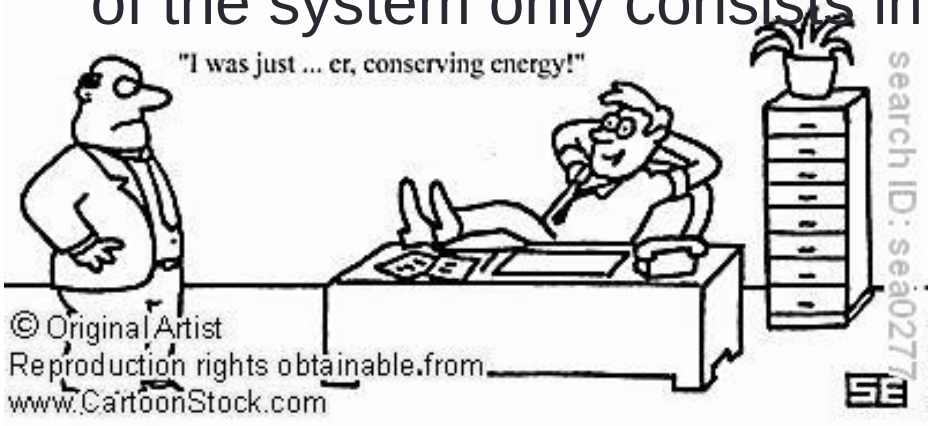
Qualsiasi colore
A colori
Bianco e nero

Qualsiasi tipo
Volti
Foto
Clip art
Disegni

Visual. standard
Mostra dimensioni

Benefits of a data-driven approach

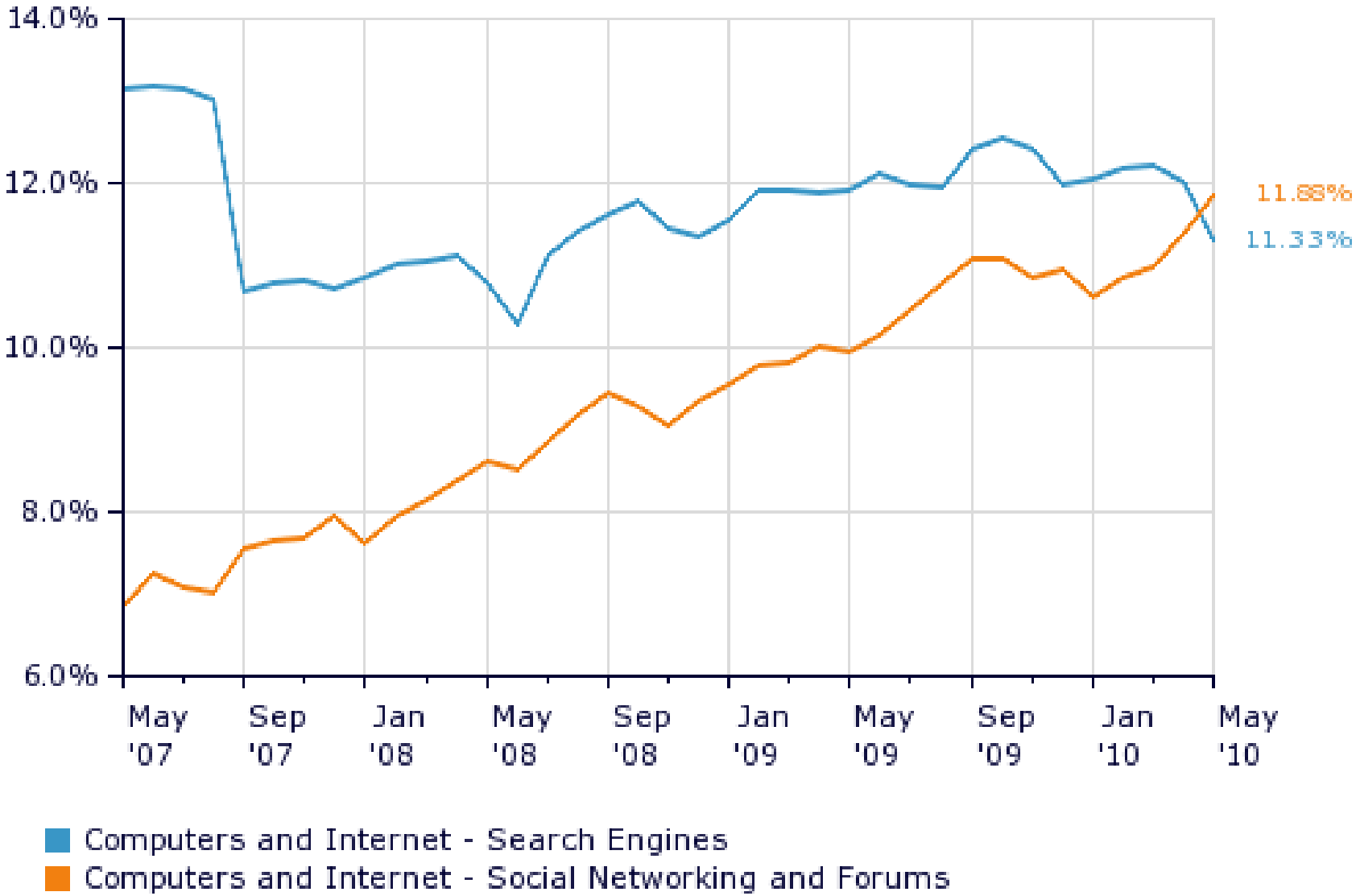
- Very effective learning algorithms available (e.g. Support Vector Machines)
- The ML technology is *portable* while imperative coding is *task (*i.e. *scenario) specific*
- Very accurate solutions can be obtained
- Gathering training data *much less expensive* than rule coding
- In dynamically evolving 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

UK Internet visits to Social Networks and Search Engines



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

Amazon.com recommends "Collective Intelligence in Ac]

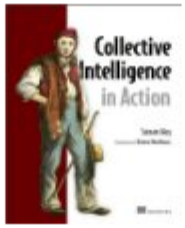
Tools Window Help

Subject: {Disarmed} Amazon.com recommends "Collective Intelligence in

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

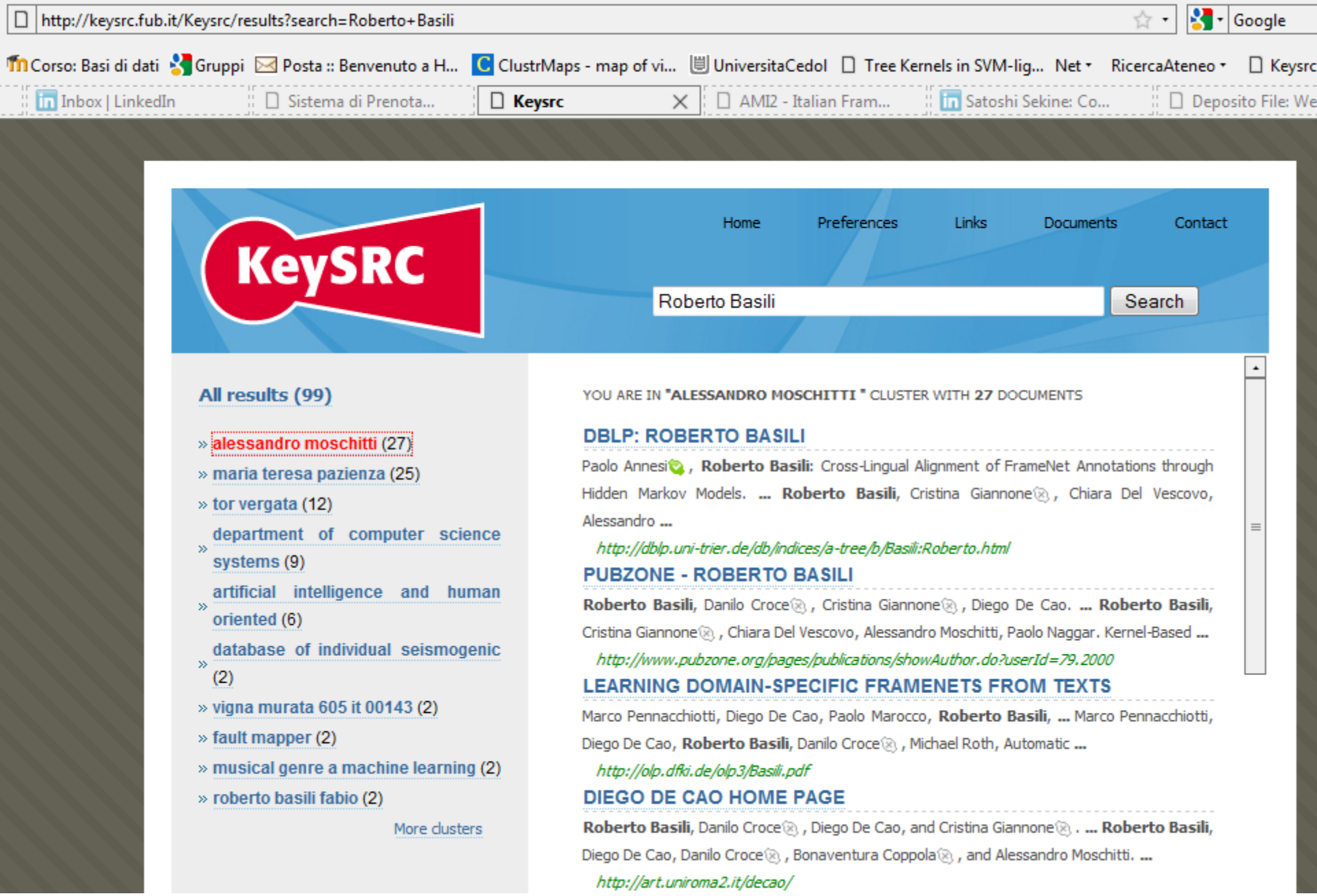
Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)

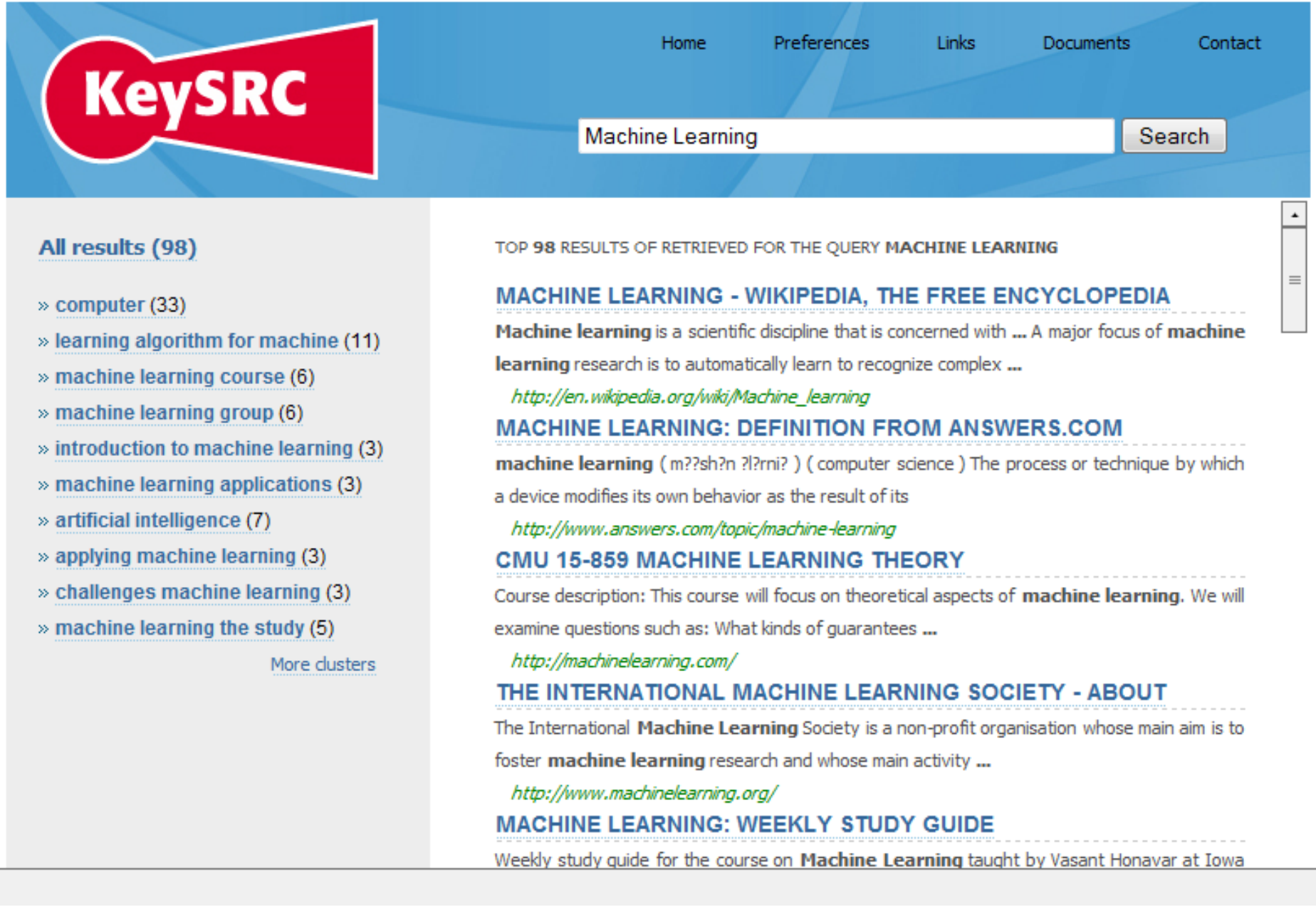
Quantitative Methods In Linguistics

See More Recommendations

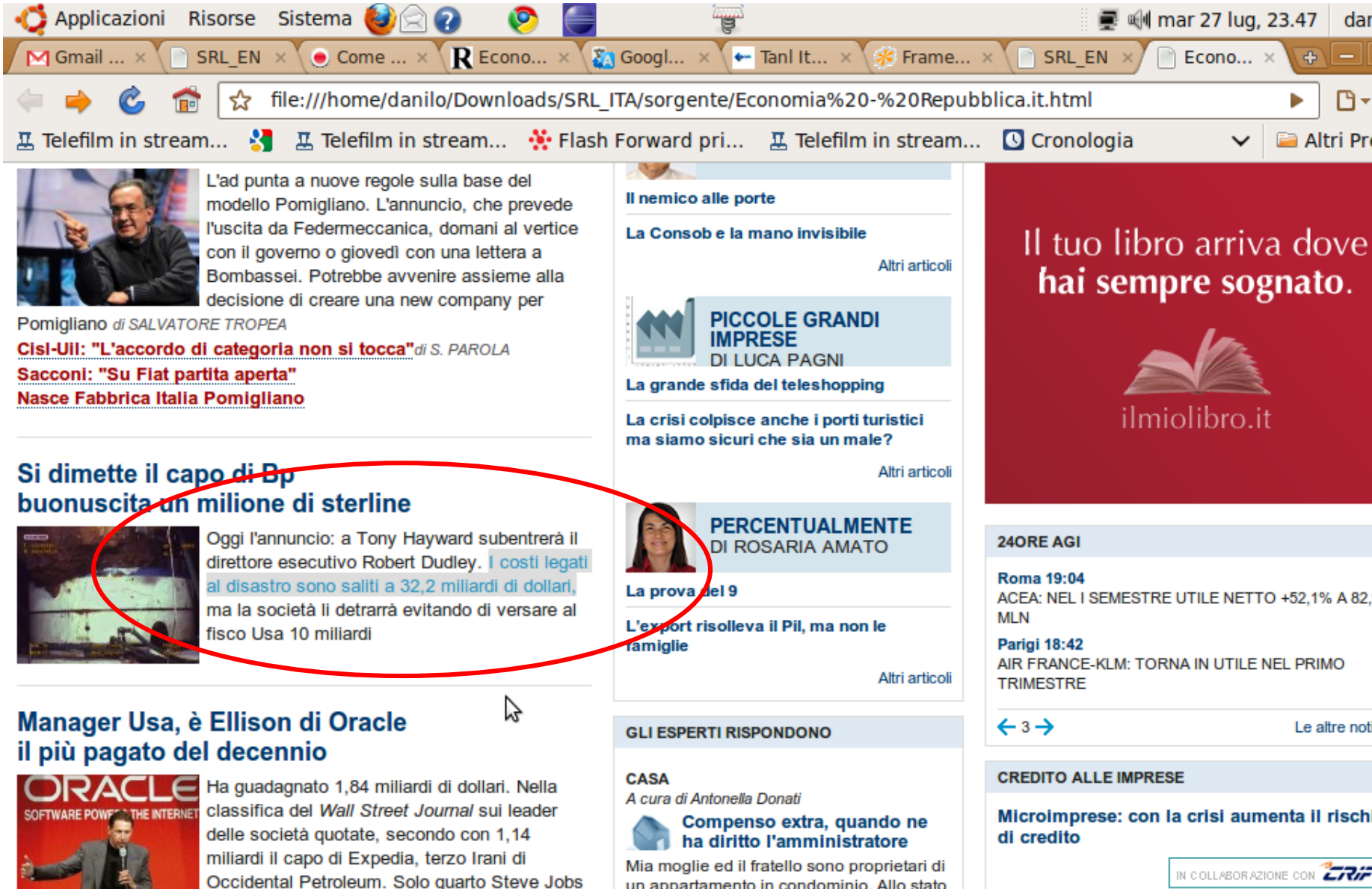
Collective Intelligence in Action
by Satnam Alag
Average customer review:

Add to cart
Add to Wish List


http://keysrc.fub.it/Keysrc/results?search=Roberto+Basili
Google
Corso: Basi di dati
Gruppi
Posta :: Benvenuto a H...
ClustrMaps - map of vi...
UniversitaCedol
Tree Kernels in SVM-lig...
Net
RicercaAteneo
Keysrc
Inbox | LinkedIn
Sistema di Prenota...
Keysrc
AMI2 - Italian Fram...
Satoshi Sekine: Co...
Deposito File: We
KeySRC
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 Naggar. 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/


KeySRC
Home
Preferences
Links
Documents
Contact
Machine Learning
Search
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
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



WEB MINING & RETRIEVAL, a.a. 2015-16

Moduli	Argomenti	Lezioni
Basic ML	Introduzione all'algoritmica di ML. Probabilità e Metriche di similarità. Classificazione mediante algoritmi di base	Introduzione al Corso, ed al WM&R.
		Richiami ai metodi base di ML. Metodi Supervised vs. Unsupervised.
		Metodi probabilistici e generativi. Naive Bayes.
		Esercitazione. Decision Trees. Intro Weka (1)
		Metodi algebrico-lineari. Metriche e similarità semantica.
Clusering and Probabilistic methods	Introduction to clustering algorithms. Generative Models. HMMs	Metodi Unsupervised. Clustering. Metodi Agglomerativi: K-means
		Esercitazione. Metodi avanzati di Clustering. Uso di Weka (2)
		Modelli di Linguaggio. Processi Markoviani.
		Modelli Generativi: HMM. (POS tagging)
From PAC learnability to SVM	PAC learnability. VC-dimension. SVMs. Kernel methods	PAC Learnability. Perceptron
		SVM. Hard Margin.
		Soft margin SVM. La Nozione di Kernel.
		Kernel polinomiali e RBF. Sequence & Tree Kernels.
		Use of Tree and Sequence Kernels in KELP
Advanced ML topics: Neural Networks	Semi-supervised learning: ensemble methods, active learning, EM. On-line learning: Passive-Aggressive. Deep Neural netork architectures.	Improving ML settings: Semisupervised Extensions
		Improving ML settings: *On-line Learning* ***con esercitazione***
		From neural networks to deep learning: perceptrons and MLP
		Deep Learning over MLPs. Recurrent Neural Networks.
		Adopting Convolutional Neural Networks on images, ….
Basic IR topics	Ad hoc IR. Text Classification. IR models for texts and queries. Lexical Resources for IR.	IR tradizionale. Paradigmi e Architettura.
		IR tradizionale. Modelli algebrico-lineari
		Introduzione a Lucene: Perf Eval su Cranfield
		Query Operations/Expansion - Automatic Global Analysis - Thesaurus/SA
Advanced IR topics	Dimensionality Reduction for IR. Distributional Semantic models for IR. Word Embeddings with NNs.	Latent Semantic Analysis. LSA.
		Latent Semantic Kernels & Semantic Kernels. Wordspaces
		Word Embeddings through Neural Networks.
		Word Space Exercise
Web Information Retrieval	Web Applications & IR. Web search and Crawling. Random surfer models: Google PageRank	Introduzione al web and Social Medi Analysis
		Web Search: Rango e Relevance: PageRank. HITS
		Thematic Web Crawling project: Development of an indexed Web dump
		HITS. Web Crawling
Social Media Analytics (*)	IR in Social Media. Community detection. User profiling and Recommending. Sentiment and Emotion Analysis.	Social Media Analysis: Community Detection
		Opinion Mining e Sentiment Analysis: the task
		Introduzione all'OM & SA: Twitter as a case study
		Development of a SA tool
		Social Media Analysis: Recommending
Statistical Natural Language Processing (*)	Standard Natural Langauge Processing. Statistical Parsing. Evaluation of Statistical NLP tools.	Richiami ai metodi di Elaborazione del Linguaggio Naturale: il TAL
		Risorse: Corpora, Lessici, Grammatiche e Basi di Conoscenza per il TAL.
		HMM-based POS tagging.
		Esercitazione: Stanford NLP chain. Evaluation.

WEB MINING & RETRIEVAL, a.a. 2015-16

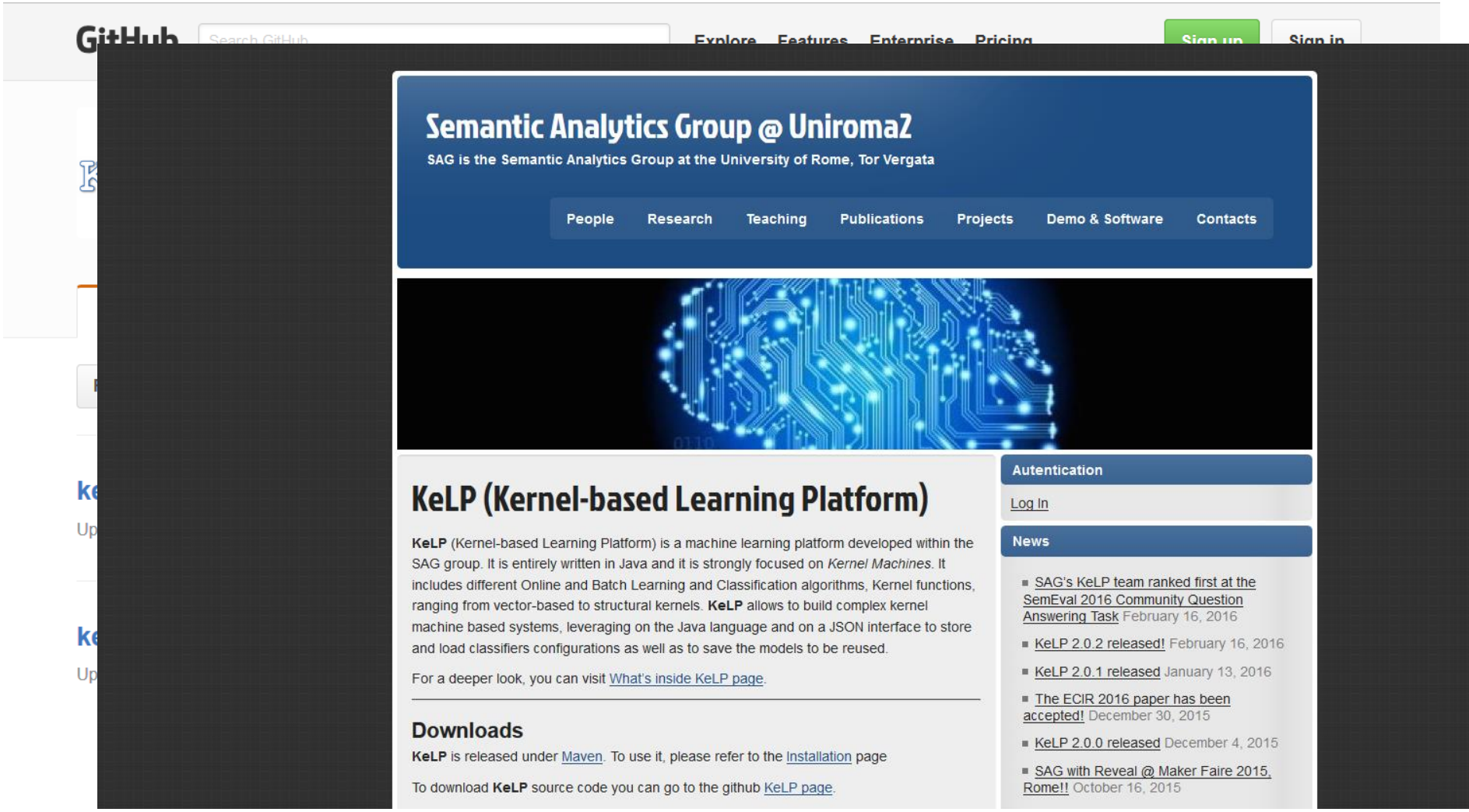
Moduli	Argomenti	Lezioni
Basic ML	Introduzione all'algoritmica di ML. Probabilità e Metriche di similarità. Classificazione mediante algoritmi di base	Introduzione al Corso, ed al WM&R. Richiami ai metodi base di ML. Metodi Supervised vs. U Metodi probabilistici e generativi. Naive Bayes. **Esercitazione. Decision Trees. Intro Weka (1)** Metodi algebrico-lineari. Metriche e similarità semanti
Clusering and Probabilistic methods	Introduction to clustering algorithms. Generative Models. HMMs	Metodi Unsupervised. Clustering. Metodi Agglomerativ **Esercitazione. Metodi avanzati di Clustering. Uso di W** Modelli di Linguaggio. Processi Markoviani. Modelli Generativi: HMM. (POS tagging)
From PAC learnability to SVM	PAC learnability. VC-dimension. SVMs. Kernel methods	PAC Learnability. Perceptron SVM. Hard Margin. Soft margin SVM. La Nozione di Kernel. Kernel polinomiali e RBF. Sequence & Tree Kernels. **Use of Tree and Sequence Kernels in KELP**
Advanced ML topics: Neural Networks	Semi-supervised learning: ensemble methods, active learning, EM. On-line learning: Passive-Aggressive. Deep Neural netork architectures.	Improving ML settings: Semisupervised Extensions Improving ML settings: *On-line Learning* ***con esercitaz*** From neural networks to deep learning: perceptrons a Deep Learning over MLPs. Recurrent Neural Networks. **Adopting Convolutional Neural Networks on images,**
Basic IR topics	Ad hoc IR. Text Classification. IR models for texts and queries. Lexical Resources for IR.	IR tradizionale. Paradigmi e Architettura. IR tradizionale. Modelli algebrico-lineari **Introduzione a Lucene: Perf Eval su Cranfield** Query Operations/Expansion - Automatic Global Analy
	Dimensionality Reduction for IR.	Latent Semantic Analysis. LSA.

Clusering and Probabilistic methods	Introduction to clustering algorithms. Generative Models. HMMs	Metodi Unsupervised: Clustering. Metodi Agglomerativi: K-m **Esercitazione. Metodi avanzati di Clustering. Uso di Weka (2** Modelli di Linguaggio. Processi Markoviani. Modelli Generativi: HMM. (POS tagging)
From PAC learnability to SVM	PAC learnability. VC-dimension. SVMs. Kernel methods	PAC Learnability. Perceptron SVM. Hard Margin. Soft margin SVM. La Nozione di Kernel. Kernel polinomiali e RBF. Sequence & Tree Kernels. **Use of Tree and Sequence Kernels in KELP**
Advanced ML topics: Neural Networks	Semi-supervised learning: ensemble methods, active learning, EM. On-line learning: Passive-Aggressive. Deep Neural netork architectures.	Improving ML settings: Semisupervised Extensions Improving ML settings: *On-line Learning* ***con esercitazione*** From neural networks to deep learning: perceptrons and MLF Deep Learning over MLPs. Recurrent Neural Networks. **Adopting Convolutional Neural Networks on images, ...**
Basic IR topics	Ad hoc IR. Text Classification. IR models for texts and queries. Lexical Resources for IR.	IR tradizionale. Paradigmi e Architettura. IR tradizionale. Modelli algebrico-lineari **Introduzione a Lucene: Perf Eval su Cranfield** Query Operations/Expansion - Automatic Global Analysis - Th
Advanced IR topics	Dimensionality Reduction for IR. Distributional Semantic models for IR. Word Embeddings with NNs.	Latent Semantic Analysis. LSA. Latent Semantic Kernels & Semantic Kernels. Wordspaces Word Embeddings through Neural Networks. **Word Space Exercise**
Web Information Retrieval	Web Applications & IR. Web search and Crawling. Random surfer models: Google PageRank	Introduzione al web and Social Medi Analysis Web Search: Rango e Relevance: PageRank. HITS **Thematic Web Crawling project: Development of an indexe** HITS. Web Crawling
Social Media Analytics (*)	IR in Social Media. Community detection. User profiling and Recommending. Sentiment and Emotion Analysis.	Social Media Analysis: Community Detection Opinion Mining e Sentiment Analysis: the task Introduzione all'OM & SA: Twitter as a case study **Development of a SA tool** Social Media Analysis: Recommending
Statistical Natural Language Processing (*)	Standard Natural Langauge Processing. Statistical Parsing. Evaluation of Statistical NLP tools.	Richiami ai metodi di Elaborazione del Linguaggio Naturale: il Risorse: Corpora, Lessici, Grammatiche e Basi di Conoscenza HMM-based POS tagging. **Esercitazione: Stanford NLP chain. Evaluation.**

Laboratori del Corso

- Nella’mbito dei Laboratori agli studenti saranno resi disponibili:
  - Piattaforme di Machine Learning: Weka, R, KELP
  - Motori di Ricerca: Lucene, Terrier
  - Strumenti di AI per l’elaborazione dei testi:
    - Recursive Neural Networks per l’apprendimento di lessici vettoriali
    - Parser grammaticali di linguaggi naturali (ita,eng)
    - Ambienti di Enterprise Semanic Search su Web

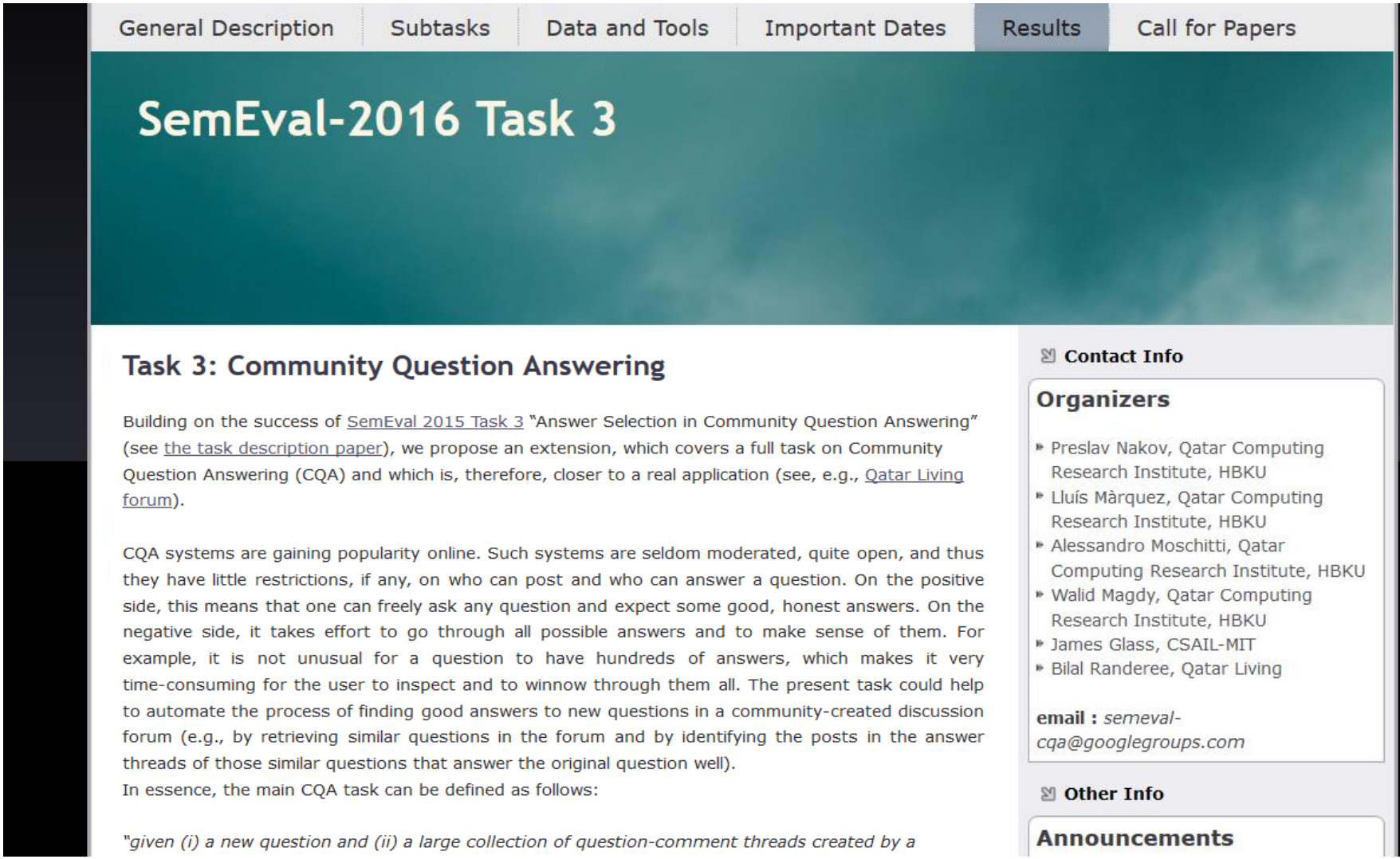
Kelp: Java-based kernel framework



https://github.com/SAG-KeLP

http://sag.art.uniroma2.it/demo-software/kelp/

KELP applications: cQA



KELP applic

General Description

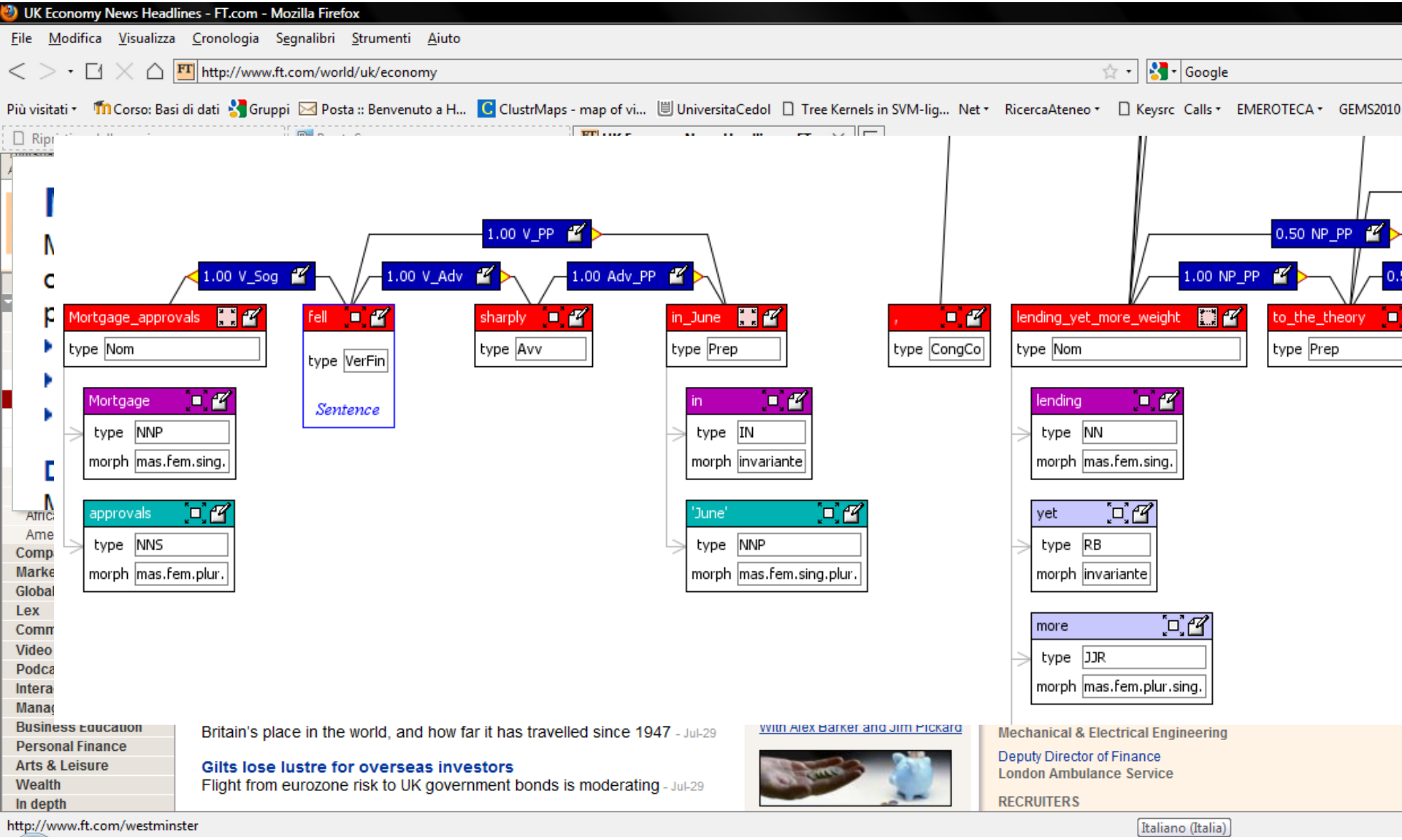
SemEval-20

Team ID	Team Affilia
ConvKN	Qatar Compu
ECNU	East China N
ICL00	Institute of C
ICRC-HIT	Intelligence C
ITNLP-AiKF	Intelligence T
Kelp	University of
MTE-NN	Qatar Compu
overfitting	University of
PMI-cool	Sofia Univers
QAIIIT	IIIT Hyderab
QU-IR	Qatar Univers
RDI_team	RDI Egypt, C
SemanticZ	Sofia Univers
SLS	MIT Comput
SUper_team	Sofia Univers
UH-PRHLT	Pattern Recog
	Universitat P
UniMelb	The Universit
UPC_USMBA	Universitat P

	Submission	MAP	AvgRec	MRR	P	R	F1	Acc
1	**Kelp-primary**	**79.19**$_1$	**88.82**$_1$	**86.42**$_1$	**76.96**$_1$	**55.30**$_8$	**64.36**$_5$	**75.11**$_2$
	ConvKN-contrastive1	78.71	88.98	86.15	77.78	53.72	63.55	74.95
	SUper_team-contrastive1	77.68	88.06	84.76	75.59	55.00	63.68	74.50
2	**ConvKN-primary**	**77.66**$_2$	**88.05**$_3$	**84.93**$_4$	**75.56**$_2$	**58.84**$_6$	**66.16**$_2$	**75.54**$_1$
3	**SemanticZ-primary**	**77.58**$_3$	**88.14**$_2$	**85.21**$_2$	**74.13**$_4$	**53.05**$_{10}$	**61.84**$_8$	**73.39**$_5$
	ConvKN-contrastive2	77.29	87.77	85.03	74.74	59.67	66.36	75.41
4	**ECNU-primary**	**77.28**$_4$	**87.52**$_5$	**84.09**$_6$	**70.46**$_6$	**63.36**$_4$	**66.72**$_1$	**74.31**$_4$
	SemanticZ-contrastive1	77.16	87.73	84.08	75.29	53.20	62.35	73.88
5	**SUper_team-primary**	**77.16**$_5$	**87.98**$_4$	**84.69**$_5$	**74.43**$_3$	**56.73**$_7$	**64.39**$_4$	**74.50**$_3$
	MTE-NN-contrastive2	76.98	86.98	85.50	58.71	70.28	63.97	67.83
	SUper_team-contrastive2	76.97	87.89	84.58	74.31	56.36	64.10	74.34
	MTE-NN-contrastive1	76.86	87.03	84.36	55.84	77.35	64.86	65.93
	SLS-contrastive2	76.71	87.17	84.38	59.45	67.95	63.41	68.13
	SLS-contrastive1	76.46	87.47	83.27	60.09	69.68	64.53	68.87
6	**MTE-NN-primary**	**76.44**$_6$	**86.74**$_7$	**84.97**$_3$	**56.28**$_9$	**76.22**$_1$	**64.75**$_3$	**66.27**$_8$
7	**SLS-primary**	**76.33**$_7$	**87.30**$_6$	**82.99**$_7$	**60.36**$_8$	**67.72**$_3$	**63.83**$_6$	**68.81**$_7$
	ECNU-contrastive2	75.71	86.14	82.53	63.60	66.67	65.10	70.95
	SemanticZ-contrastive2	75.41	86.51	82.52	73.19	50.11	59.49	72.26
	ICRC-HIT-contrastive1	73.34	84.81	79.65	63.43	69.30	66.24	71.28
8	**ITNLP-AiKF-primary**	**71.52**$_8$	**82.67**$_9$	**80.26**$_8$	**73.18**$_5$	**19.71**$_{12}$	**31.06**$_{12}$	**64.43**$_9$
	ECNU-contrastive1	71.34	83.39	78.62	66.95	41.31	51.09	67.86
9	**ICRC-HIT-primary**	**70.90**$_9$	**83.36**$_8$	**77.38**$_{10}$	**62.48**$_7$	**62.53**$_5$	**62.50**$_7$	**69.51**$_6$
10	**PMI-cool-primary**	**68.79**$_{10}$	**79.94**$_{10}$	**80.00**$_9$	**47.81**$_{12}$	**70.58**$_2$	**57.00**$_9$	**56.73**$_{12}$
	UH-PRHLT-contrastive1	67.57	79.50	77.08	54.10	50.11	52.03	62.45
11	**UH-PRHLT-primary**	**67.42**$_{11}$	**79.38**$_{11}$	**76.97**$_{11}$	**55.64**$_{10}$	**46.80**$_{11}$	**50.84**$_{11}$	**63.21**$_{10}$
	UH-PRHLT-contrastive2	67.33	79.34	76.73	54.97	49.13	51.89	62.97
12	**QAIIIT-primary**	**62.24**$_{12}$	**75.41**$_{12}$	**70.58**$_{12}$	**50.28**$_{11}$	**53.50**$_9$	**51.84**$_{10}$	**59.60**$_{11}$
	QAIIIT-contrastive2	61.93	75.22	69.95	49.48	49.96	49.72	58.93
	QAIIIT-contrastive1	61.80	75.12	69.76	49.85	50.94	50.39	59.24
	Baseline 1 (IR)	**59.53**	**72.60**	**67.83**	—	—	—	—
	Baseline 2 (random)	52.80	66.52	58.71	40.56	74.57	52.55	45.26
	Baseline 3 (all 'true')	—	—	—	40.64	100.00	**57.80**	40.64
	Baseline 4 (all 'false')	—	—	—	—	—	—	**59.36**

Table 1: **Subtask A, English (Question-Comment Similarity):** results for all submissions. The first column shows the rank of the primary runs with respect to the official MAP score. The second column contains the team's name and its submission type (primary vs. contrastive). The following columns show the results for the primary, and then for other, unofficial evaluation measures. The subindices show the rank of the

Natural Language Parsing tool: RevNLT



PrestoSpace Publication Platform - Microsoft Internet Explorer

File Modifica Visualizza Preferiti Strumenti ?

Indietro Cerca Preferiti Multimedia

Indirizzo http://localhost:8080/PublicationPlatform/protected/programmes/tg1_2005-10-23T2000/index.html

PrestoSpace

An integra

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

Info | Transcription | Semantic Analysis | Content Analysis | Search

00:06:36	chamonix america dove perde forza ma fa sempre paura l' uragano di mallarme'italia 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 wilma il dodicesimo ciclone di una stagione ecc 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 dino risi 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 kiss e' gia' iniziata la grande fuga non bastasse sull' atlantico a sud di porto rico si e' formata falla venti dihemingway 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

TIME 00:00:02:39

TIMELINE

- 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


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

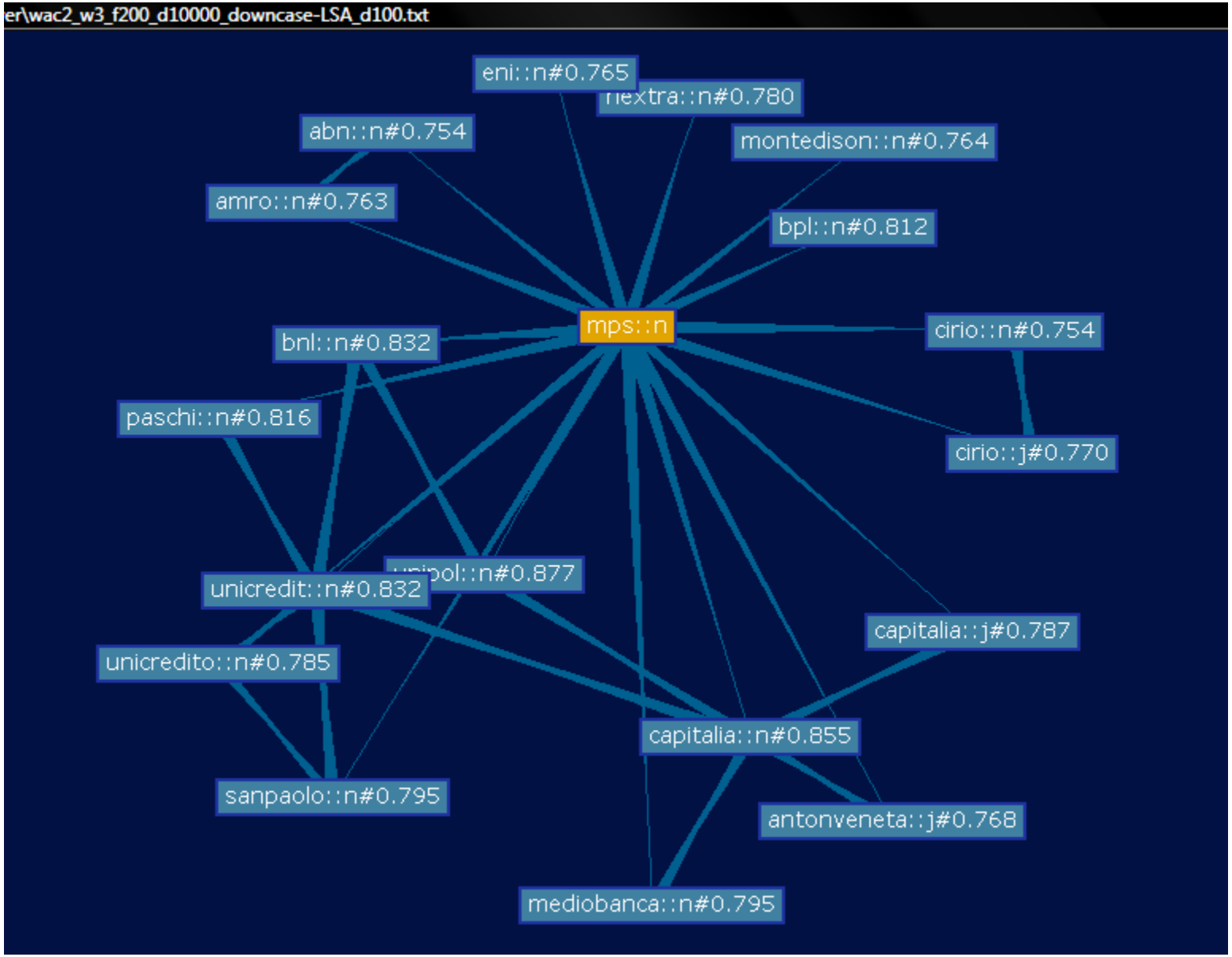

00:07:45:20


00:07:55:06

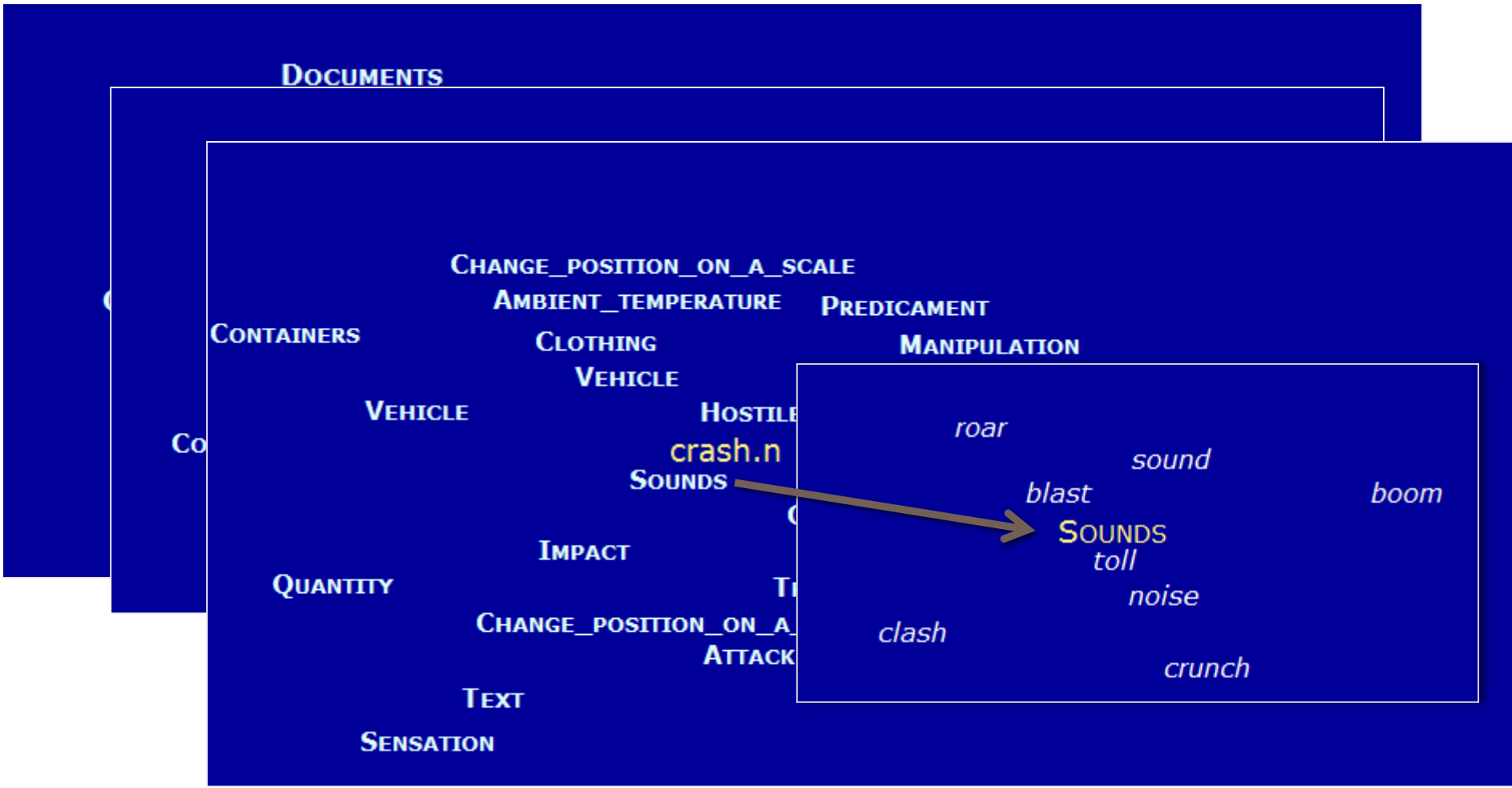

00:07:57:07


00:07:59:13

Vector Spaces for Lexical Semantics



Spaces for NL predicates

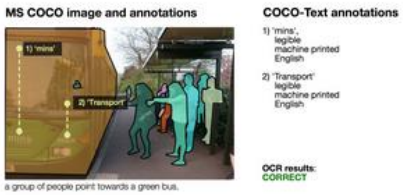


Sperimentazioni attive su http://mscoco.org/

Microsoft Common Objects in Context

cocodataset@outlook.com

Home People Explore Dataset External



COCO-Text

COCO-Text is for both text detection and recognition. The dataset annotates scene text with transcriptions along with attributes such as legibility, printed or handwritten text.

FM-IQA

The Freestyle Multilingual Image Question Answering (FM-IQA) dataset contains over 120,000 images and 250,000 freestyle Chinese question-answer pairs and their English translations.

VQA

VQA is a new dataset containing open-ended questions about images. These questions require an understanding of vision, language and commonsense knowledge to answer.

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**