



Sentiment Analysis and Opinion Mining

Franco Tuveri – tuveri@crs4.it
Manuela Angioni – angioni@crs4.it

NIT – Natural Interaction Technologies:
<http://www.crs4.it/natural-interaction-technologies>

Cagliari, 19 Settembre 2012



Opinion Mining

Opinion Mining: Identificazione e Analisi delle Opinioni nei testi.
Sentiment Analysis: Identificazione e deduzione di stati emotivi nel testo.

Opinion Mining is a new discipline which has recently attracted increased attention within fields such as marketing, personal affective profiling, and financial market prediction. Although often associated with sentiment analysis, which consists in inferring emotional states from text, opinion mining is an independent area related to natural language processing and text mining that deals with the identification of opinions and attitudes in natural language texts.

Erik Cambria et al., 2010
SenticNet: A Publicly Available Semantic Resource for Opinion Mining




Opinioni....

Il recente simposio sulla Sentiment Analysis¹, organizzato da Seth Grimes in New York, si è focalizzato su:

- Lo stato attuale delle ricerche riguardante la Sentiment Analysis e la Text Analytics.
- La fusione di tecnologie e mercato nella analisi delle opinioni e dei comportamenti attraverso i commenti riportati sui social media, news e forum aziendali.
- Diffusione in svariati e nuovi domini di interesse:
 - military intelligence
 - financial markets,
 - strategie di real marketing basate sui contenuti dei social media.

Definizione di deep Marketing Research (DeepMR), "enabled by an ensemble of text analytics, sentiment analysis, behavioral analyses, and psychometric technologies — applied to social and online sources, as well as to traditional surveys — with the potential to revolutionize market research".

¹ <http://sentimentsymposium.com/>



Opinioni....

Le opinioni degli altri possono condizionare le nostre scelte

Diffusione

Passaparola su Web.

- Dal passaparola casuale ai commenti degli utenti descrittivi esperienze, percezioni condivise su blog, forum e siti specializzati per la pubblicazione di review.

Opinioni su scala globale. Non più limitati a:

- Individui o circoli ristretti di amici
- Ambiti di affari circoscritti o localizzati



Applicazioni

Rivolte ad Aziende

Settore Turistico:

- Migliorare il servizio offerto ai clienti, misurare l'analisi dei flussi verso mete turistiche e culturali

Opinion search/retrieval:

- Ricerca di opinioni e pareri in generale

Market intelligence, Business intelligence, Product/Service benchmarking:

- Analisi comparativa di prodotti, servizi, etc.,
- Analisi reputazione aziendale o stato di salute di un brand
- Supporto al lancio di nuovi prodotti
- Identificazione di trend di mercato emergenti
- Verifica efficacia campagne di comunicazione

Voice of the Customer (VOC):

- Analizzare le interazioni e le opinioni dei clienti, via email, note, forum, blog, altri social media



Applicazioni

Rivolte a Persone

Personaggi Pubblici/Politica:

- Monitorare la propria immagine e reputazione (trust) e il proprio gradimento

Consumatori:

sono interessati alle opinioni degli altri quando:

- Devono acquistare un prodotto o un servizio
- Devono trovare opinioni o pareri su argomenti politici

Profilazione utenti:

- Studio delle dinamiche di gruppo
- Individuazione di opinion leader e di gruppi di influenza
- Verifica efficacia campagne di comunicazione



Applicazioni

Rivolte a Istituzioni

Politica:

ascolta la voce dei cittadini

- Sondaggi, Rilevamenti di opinioni

Monitoraggio e analisi dei fenomeni sociali:

- per l'individuazione di situazioni potenzialmente pericolose
- la determinazione dello stato d'animo in generale o relativamente ad un evento


Televisione (comunicazione di massa):

- valutazione della qualità e dell'impatto sociale dei programmi televisivi



Applicazioni

Alcuni esempi



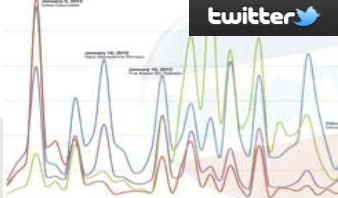
Intenzioni di voto, Sondaggi

crimson hexagon
HOW USE. HOW WHY. WHEN HOW


La Reuters si è rivolta alla Crimson Hexagon, per lo sviluppo di uno strumento di Sentiment Analysis per l'analisi delle intenzioni di voto per le elezioni USA 2010, basato su Twitter.

@mentionmachine

Il Washington Post sviluppa @MentionMachine, un'app che utilizza Twitter (e altri blog) come sistema di analisi real-time degli umori dell'elettorato delle primarie repubblicane del 2012.



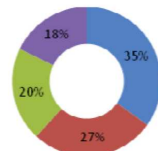
Rank	Candidate	Twitter mentions	Media Mentions
1	Barack Obama	225,752	12,296
2	Mitt Romney	148,336	5,710
3	Ron Paul	66,768	749
4	Newt Gingrich	43,258	1,423



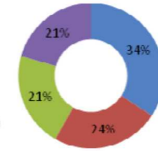
Monitoraggio e analisi dei fenomeni sociali

Nei giorni successivi alla sentenza d'appello per il delitto di Perugia, Expert System ha svolto un'analisi per cogliere le reazioni dei commenti in italiano e in inglese degli utenti di Twitter.

italiano, Amanda




inglese, Amanda




processo
giustizia
criminalità
magistratura

giustizia
imputato
prigione
criminalità



EXPERT SYSTEM SEMANTIC INTELLIGENCE

<http://www.expertsystem.it/>



Feeling breakdown of feelings from people in the last few hours

<http://wefeelfine.org/>



Search for Opinions...

SWOTTI. Search, Rate and Compare.

Most relevant opinions in Internet with semantic rating.

Now follow your brand 30 days for Free

Search opinion for Search

Technology | Mobile Phones | Cameras | MP3 | PDAs | Printers | TVs | Monitors | GPS | Software | Laptops | Automations | Cars | Motorbikes | Destinations | Countries | Cities | Entertainment | Video Games | Consoles | Cinema | Movies | Music | Music Bands | Literature | Books | Personalities | People | Business | Companies

Search⁰¹ Over 20 categories. With the possibility of adding new categories. More information >>

Rate⁰² Over 3 millions opinions. Each day more than 10.000 post-tracked. More information >>

Compare⁰³ From books to electronics. Discover all the variety of Swotti! More information >>

Did you know ... Swotti is possible to seek opinions of movies and music? 

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

The screenshot shows the regulations.gov website. At the top, there is a navigation bar with links for Home, Help, Resources, and Feedback and Questions. Below this is a search bar with a magnifying glass icon and buttons for Search, Browse, and Learn. The main content area is divided into several sections:

- Let Your Voice Be Heard:** A section encouraging users to share their knowledge and make their voice count. It includes a search bar for Rules, Comments, Adjudications or Supporting Documents.
- Are you new to the site?:** A section with links for getting started, such as "How do I find a rule?", "How do I submit a comment?", "How do I find my comment?", and "Do my comments make a difference?".
- Regulations.gov Re-launch:** A section with a "President's Executive Order" and a link to visit the new Facebook page.
- Coming Soon...:** A section with a magnifying glass icon and the text "A Quick & New Way to Search".
- Regulations With Comments Due Soon:** A table listing regulations due soon, categorized by time periods (Today, Next 3 Days, Next 7 Days, Next 15 Days, Next 30 Days, Next 90 Days) and the number of regulations in each category.
- Newly Posted Regulations:** A table listing newly posted regulations, categorized by time periods (Today, Last 3 Days, Last 7 Days, Last 15 Days, Last 30 Days, Last 90 Days) and the number of regulations in each category.

At the bottom of the page, there is a URL: <http://www.regulations.gov/#/home;tab=search>

The slide is titled "Progetto FIRST" and features the CRS4 logo (CRS4 IDEAS BECOME LIFE). The main heading is "Large scale inFormation extraction and Integration infRastructure for SupporTing financial decision making".

The slide includes a line graph showing "Sentiment analysis" over time, with a legend indicating "Positive", "Negative", and "Neutral" sentiment. The graph shows a significant peak in positive sentiment followed by a sharp decline.

The slide lists the following bullet points:

- ABI Lab Conference a Milano: presentato il primo prototipo di modello di supporto alle decisioni basato sugli stati emotivi espressi sul Web.
- Applicato al servizio di micro blogging Twitter o ai dati testuali estratti dai blogs.
- Estrae e mette in relazione gli stati emotivi ai movimenti dei prezzi azionari.
- I risultati indicano una forte relazione positiva tra stato emotivo e volume degli scambi.
- Fornisce all'utente finale spunti importanti sui movimenti dei mercati finanziari e contribuisce al miglioramento del processo decisionale e all'efficienza del mercato.

At the bottom of the slide, there is a URL: <http://project-first.eu/content/press-release-eu-project-first-uses-twitter-co-financial-decision-making>

The slide features the CRS4 logo (CRS4 IDEAS BECOME LIFE) at the top left. The main heading is "Opinion Mining". A large magnifying glass icon is positioned over the text, symbolizing the process of searching for and analyzing opinions.

The slide features the CRS4 logo (CRS4 IDEAS BECOME LIFE) at the top left. The main heading is "Opinion Mining: Overview".

The slide lists the following bullet points:

- Un sistema di Opinion Mining ha come principali obiettivi:
- Analizzare le opinioni riferite ad eventi o a fatti anche in corso di svolgimento, anche quando esse non siano riferite a specifici argomenti.
- Estrarre feature e informazioni significative contenute nelle opinioni, relative a diversi contesti non sempre ben definiti, a partire da molteplici sorgenti di review.
- Esplorare la questione della contestualizzazione delle feature attraverso l'utilizzo di strumenti per la classificazione semantica, per la gestione di reti semantiche e l'utilizzo di risorse linguistiche ad hoc.
- Aggregare e rappresentare i risultati elaborati (Opinion Summarization)

CRS4 Opinion Mining: Overview

Sviluppo di risorse linguistiche

Opinion Summarization

Feature Extraction

Opinion Mining

Development of Linguistic Resource

Sentiment Classification

Opinion Summarization

Text Classification

Text Summarization

Computational Linguistics

Appraisal Theory

Conjuncting Method

PMI Method

WordNet Exploring Method

Gloss Classification Method

PMI Method

Machine Learning Method

NLP Combined Method

Opinion Mining can be roughly divided into three major tasks of development of linguistic resources, sentiment classification, and opinion summarization.

Lee et al., 2008 - Opinion Mining of customer feedback data on the web.

CRS4 Definizione di Opinione

Un'opinione è una quintupla: $(o_j, f_{jk}, so_{ijkl}, h_i, t_i)$

o_j è l'oggetto di riferimento

f_{jk} è una feature dell'oggetto o_j .

so_{ijkl} è il valore dell'opinione espressa dalla persona h_i sulla feature f_{jk} dell'oggetto o_j al tempo t_i . so_{ijkl} può assumere valore pos, neg, o neu, o un rating più granulare.

h_i è la persona che esprime l'opinione (opinion holder).

t_i è il tempo in cui l'opinione viene espressa.

(Liu, a Ch. in NLP handbook)

CRS4 Qualche definizione...

Un oggetto O è definito come un'entità che può essere un prodotto, una persona, un evento, un'organizzazione o un argomento. Un oggetto è associato ad una coppia $O: (T, A)$, dove T indica una gerarchia o una tassonomia di parti, componenti e/o sub-componenti, e A un insieme di attributi di O .

Ciascuna componente può possedere un proprio insieme di sub-componenti e attributi.

(Ding et al., 2008)

Una feature f è definita come una proprietà relativa ad uno specifico oggetto O . Essa è rappresentata dai termini o dalle espressioni, rispetto alle quali le opinioni sono espresse, e da un set di attributi.

CRS4 Putting it all together:

Finding sentiment for aspects

S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, and J. Reynar. 2008. Building a Sentiment Summarizer for Local Service Reviews. WWW Workshop

Reviews

Text Extractor

Sentences & Phrases

Sentiment Classifier

Sentences & Phrases

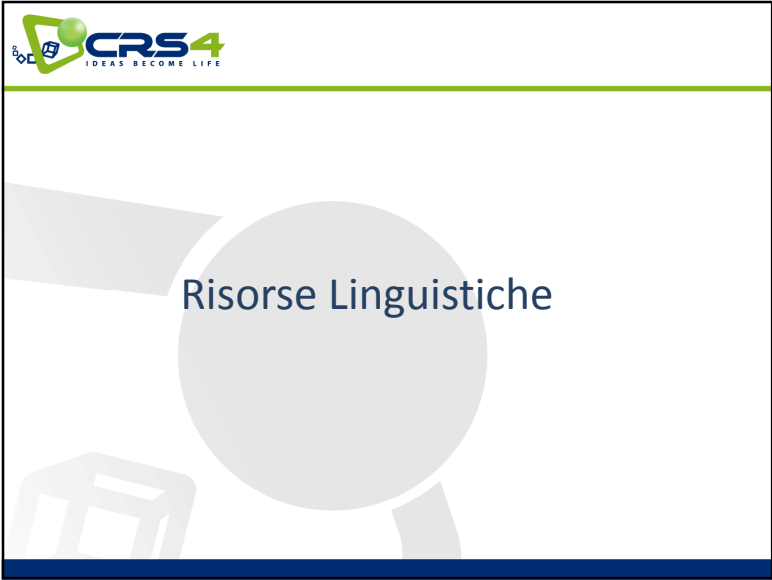
Aspect Extractor

Sentences & Phrases

Aggregator

Final Summary

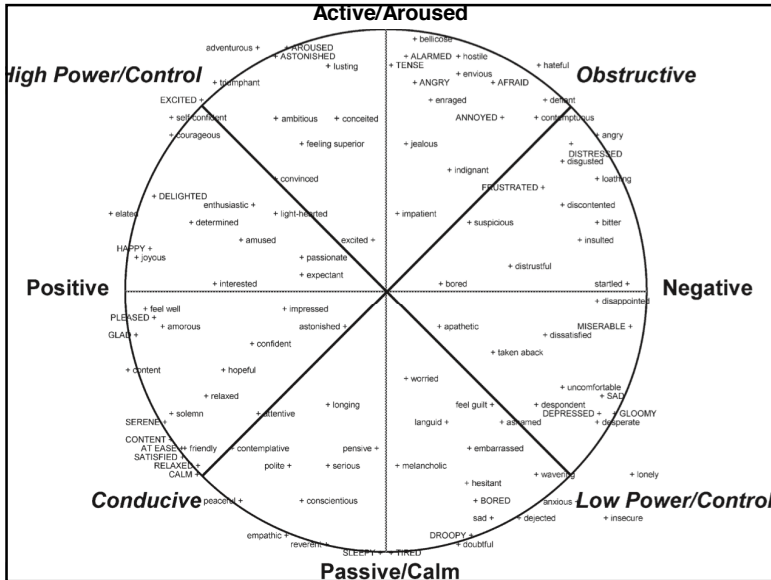
21



CRS4
IDEAS BECOME LIFE

I Differenti Stati Affettivi (secondo Scherer)

- Stati Emotivi (Emotion):** stati d'animo come reazione sincronizzata con un evento rilevante
 - rabbia, tristezza, gioia, paura, vergogna, orgoglio, entusiasmo*
- Stato d'Animo (Mood):** mutazione frequente dello stato d'animo, senza causa specifica, di bassa intensità, di lunga durata
 - allegria, tristezza, irritabilità, apatia, depressione, ottimismo*
- Atteggiamenti Interpersonali (Interpersonal stances):** interazione specifica con un'altra persona
 - amichevole, seduttivo, distante, freddo, caldo, confortante, sprezzante*
- Atteggiamenti (Attitudes):** disposizione duratura, intensa, verso persone e cose
 - simpatico, amorevole, odioso, stimato, desiderabile*
- Personalità (Personality traits):** disposizioni stabili di personalità e tendenze comportamentali tipiche
 - nervoso, ansioso, temerario, cupo, ostile, geloso*



Nel contesto della Human-Computer Interaction (HCI) il modello intende misurare quanto:

- l'utente è contento del servizio
- l'utente è interessato all'informazione fornita
- l'utente è a proprio agio con l'interfaccia
- l'utente è disposto ad utilizzare l'applicazione

E. Cambria et al. - 2010: "SenticNet: A Publicly Available Semantic Resource for Opinion Mining"

Estensioni di WordNet

- **WordNet Domains:**
 - Mapping dei synset di WordNet su un subset di DDC
- **Micro-WNOp:**
 - "Gold Standard" utilizzato per la validazione di SentiWordNet
- **SentiWordNet:**
 - Espande WordNet grazie all'acquisizione semi-automatica di polarità dei termini di WordNet
- **WordNet Affect:**
 - affective-labels o a-labels associate ai synset di WordNet
- **Q-WordNet**
 - risorsa lessicale di significati di WordNet annotati automaticamente con valori di polarità
- **FreeWordNet**
 - Database lessicale di synset arricchiti con un insieme di proprietà relative ad aggettivi ed avverbi

Sentiment Classification

Opinion Retrieval

Creazione di un Corpus di Opinioni:

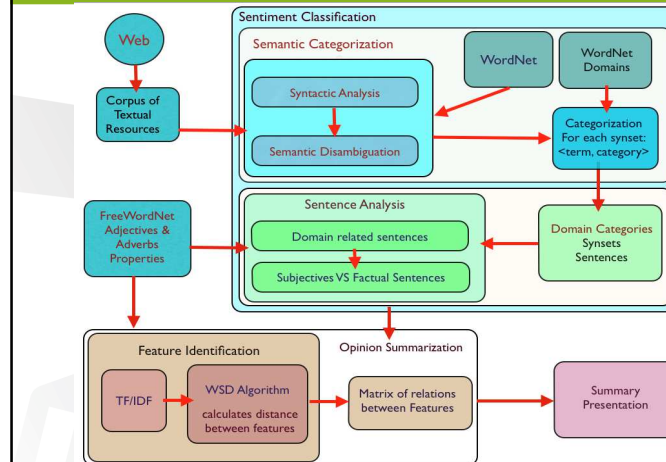
- dalle opinioni di uno specifico Data Provider (es: TripAdvisor, Booking.com, etc)
- dalle opinioni rilevanti per un argomento estratte da diverse fonti (blog, forum, siti di recensioni, giornali e news online, etc)

Analisi Linguistica

Approccio Linguistico

- L'approccio alla Opinion Mining da noi seguito è basato sulla combinazione di avverbi ed aggettivi e sull'uso dei synset di WordNet relativi a ciascun termine.
- Si focalizza sull'analisi delle opinioni attraverso le fasi di: analisi sintattica e semantica delle risorse, di information extraction e di valutazione dell'orientamento semantico.
- Si sviluppa attraverso le fasi di disambiguazione semantica e di classificazione dei testi considerando i diversi significati espressi nelle frasi che compongono il testo.
- L'utilizzo di specifiche risorse linguistiche sviluppate ad hoc, che associano significati di aggettivi ed avverbi a specifiche proprietà rende possibile l'identificazione del contesto di utilizzo dei termini e il loro raggruppamento in specifiche categorie tematiche.

Sentiment Classification



Classificazione Semantica

Il Classificatore

Lavora in diverse fasi del processo di analisi del testo.

Identificazione delle review e delle singole frasi che le compongono

Analisi Sintattica del testo (parser + chunker):


- Identificazione di aggettivi, avverbi, nomi e verbi (POS tagging)
- Identificazione di relazioni tra i termini
- Identificazione delle parole composte
- Correlazione tra aggettivi e avverbi e nomi

Analisi Semantica del testo:

- Disambiguazione del senso più probabile con cui un termine è usato (*WSD, Word Sense Disambiguation*)


Categorizzazione del corpus, delle review e delle frasi:

- Tassonomia di riferimento derivante da WordNet Domains (subset di 160 ca. categorie del DDC)



Qualche esempio di Parser Sintattico

[Montylingua](#)
[LinkGrammar](#)
[TreeTagger](#)




Un esempio di classificazione semantica:

[Classificatore semantico](#)

35



Opinion Summarization



Opinion Summarization

Feature-based o Aspect-based Opinion Summarization

Indica la generazione di sommari di opinioni relativamente a set of aspetti o features.

Aspect/Feature Identification Identify salient topics	<i>battery life, sound quality, ease of use...</i>	La Feature Identification è usata per identificare gli aspetti su cui si focalizzano le opinioni
Sentiment Prediction Determine polarity of text containing topics	<i>battery life is great → +ve long battery life → +ve</i>	La Sentiment Classification o Sentiment Prediction determina la polarità delle opinioni espresse riguardanti le feature individuate
Summary Presentation • Aggregate polarity ratings • Present opinion summaries	<i>Battery Life: ★★★★★ Sound Quality: ★★★★★</i>	La Summary Presentation mostra i risultati ottenuti nei passi precedenti.


Kim et al., 2011: Comprehensive Review of Opinion Summarization



Summary Presentation

Utilizzando i dati della feature identification e della sentiment prediction si possono generare e rappresentare i sommari finali delle opinioni in un formato efficace e facile da capire.

- **Statistical Summary:** usa i dati processati negli step di feature identification e sentiment prediction
- **Filtering:** filtri applicati ai dati processati per la selezione delle informazioni
- **Text Selection:** identifica i periodi rappresentativi per ciascuna feature
- **Aggregated Ratings:** combina la statistical summary e la text selection
- **Summary with a Timeline:** mostra le tendenze delle opinioni legandole ad una timeline




Demo: Opinion Summarization

[Feature Net](#)
[Summary Presentation](#)




Bibliografia

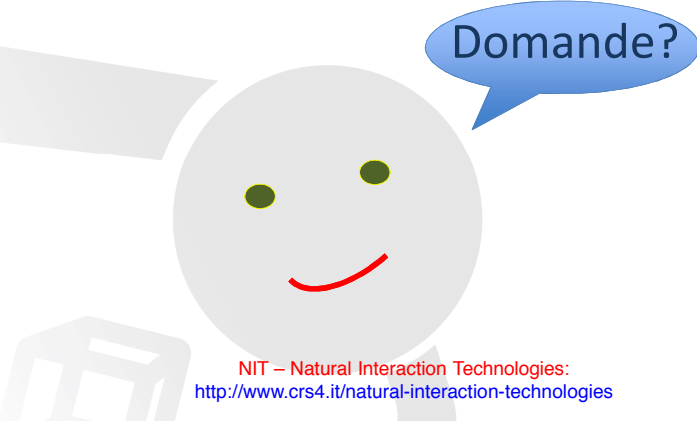
- Tuveri, F., Angioni, M. A Linguistic Approach to Feature Extraction Based on a Lexical Database of the Properties of Adjectives and Adverbs. Global WordNet Conference GWN2012, Matsue, Japan.
- Bing Liu., B. NLP Handbook Chapter: Sentiment Analysis and Subjectivity, 2nd Edition, (Editors) N. Indurkha and F. J. Damerau), 2010.
- Xiaowen Ding, Bing Liu and Philip S. Yu. 2008 A Holistic Lexicon-Based Approach to Opinion Mining. Proceedings of WSDM 2008.
- Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., Jin, C.: Red Opal: product-feature scoring from reviews. ACM Conference on Electronic Commerce 2007: 182-191 (2007)
- Zhai, Z., Liu, B., Xu, H., Jia, P.: Grouping Product Features Using Semi-Supervised Learning with Soft-Constraints. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010), Beijing, China (2010)
- Popescu, A., M., and Etzioni, O.: Extracting Product Features and Opinions from Reviews. Proceedings of the 2005 Conference on Empirical Methods in Natural Language Processing (2005)
- Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010), Beijing, China (2010)
- Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., Subrahmanian, V.,S. 2007. Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. In Proceedings of ICWSM 07 International Conference on Weblogs and Social Media, pp. 203-206.
- Rentoumi, V., Giannakopoulos, G., 2009. Sentiment analysis of figurative language using a word sense disambiguation approach. In International Conference on Recent Advances in Natural Language Processing (RANLP 2009), Borovets, Bulgaria, The Association for Computational Linguistics.



Bibliografia

- Miller, G., A., 1995. WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11
- Leacock, C. and Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In Fellbaum 1998, pp. 265-283
- Lee, D., Jeong, O., Lee, S., 2008. Opinion Mining of customer feedback data on the web. In ICUIMC '08 Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication.
- Wiebe, J., Mihalcea, R., 2006. Word Sense and Subjectivity. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, Sydney, Australia.
- Baccianella, S., Esuli, A., Sebastiani, F., 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In Proceedings of LREC-10, 7th Conference on Language Resources and Evaluation, Valletta, MT, pages 2200-2204.
- Aggeri, R., Garcia-Serrano, A., 2010. Q-WordNet: Extracting polarity from WordNet senses. Seventh Conference on International Language Resources and Evaluation.
- Valitutti, A., Strapparava, C., Stock, O., 2004. Developing affective lexical resources. In Psychology Journal Vol. 2.
- Magnini, B., Strapparava, C., Pezzulo, G., Gliozzo, A., 2002. The Role of Domain Information in Word Sense Disambiguation. Natural Language Engineering, special issue on Word Sense Disambiguation, 8(4), pp. 359-373, Cambridge University Press.
- Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., Gandini, C., 2007. Micro-WNOP: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. In Andrea Sanso', editor, Language resources and linguistic theory: Typology, second language acquisition, English linguistics, pages 200-210. Franco Angeli Editore, Milano, Italy.
- Angioni, M., Demontis, R., Tuveri, F., 2008. A Semantic Approach for Resource Cataloguing and Query Resolution. Communications of SIWN. Special Issue on Distributed Agent-based Retrieval Tools.
- Akkaya, C., Mihalcea, R., Wiebe, J., 2009. Subjectivity Word Sense Disambiguation. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 190-199, Singapore, ACL and AFNLP.





NIT – Natural Interaction Technologies:
<http://www.crs4.it/natural-interaction-technologies>