

# Tecnologie semantiche per l'analisi della reputazione online delle destinazioni turistiche

Luisa Mich, Mariangela Franch, Nicola Bortoletto<sup>1</sup>, Pamela Negrosanti<sup>1</sup>

Dipartimento di Informatica e Studi Aziendali, Università di Trento

Via Inama 5, 38122 Trento

{luisa.mich, mariangela.franch}@unitn.it

<sup>1</sup>Cogito srl

Via Zeni 8, 38068 Rovereto (Tn)

{nbortoletto, pnegrosanti}@expertsystem.it

*L'analisi della reputazione online per quanti operano nel settore del turismo sta diventando un'attività critica per il successo delle strategie di promozione e commercializzazione. La varietà dei siti web su cui gli utenti possono condividere esperienze ed opinioni e il peso che gli User Generated Content (UGC) pubblicati hanno nelle decisioni dei turisti richiedono strumenti in grado di individuare ed analizzare efficacemente grandi moli di fonti e contenuti. Focalizzando sugli UGC di tipo testuale, tecnologie linguistiche semantiche possono essere applicate per supportare un'analisi integrata della reputazione di una destinazione turistica. Per illustrare l'approccio proposto si fa riferimento a uno studio esplorativo realizzato per il Trentino, una destinazione con un'offerta turistica matura e diversificata. Oltre ad ottenere indicazioni importanti sulla reputazione online, evidenziando un 'sentiment' dei turisti complessivamente positivo, l'approccio semantico ha permesso di segnalare aspetti specifici su cui intervenire per migliorarlo, un input utile per i sistemi di Customer Relationship Management delle destinazioni.*

## 1. Introduzione

Gli sviluppi recenti del Web, indicati con la denominazione Web 2.0 [O'Reilly, 2005], si caratterizzano per un ruolo predominante degli utenti, che contribuiscono in maniera rilevante alla creazione di nuovi contenuti, o UGC (*User Generated Content*). Tali contenuti vengono pubblicati su blog o wiki, condivisi su piattaforme per il *social networking*, sono frequentemente accessibili anche da dispositivi mobili e si diffondono con velocità nella *websfera*. In alcuni casi si tratta di recensioni o valutazioni relative a servizi specifici, come avviene ad esempio sui siti delle OTA (*Online Travel Agency*), per gli alberghi o i ristoranti; in altri casi sono commenti pubblicati nei blog di

un'azienda o in quello personale; micro-messaggi, spesso usati per segnalare eventi o informazioni utili, propagando rapidamente le notizie; preferenze e gusti riguardo a specifici prodotti, sono frequenti anche nei siti 'generalisti' (ad esempio, in FaceBook). In questo contesto, da un lato si registrano percentuali crescenti di iscritti ai siti di *social networking* più famosi (si vedano, ad esempio, le statistiche di Alexa, [www.alexacom.com](http://www.alexacom.com) e Social Networking Watch, [www.socialnetworkingwatch.com](http://www.socialnetworkingwatch.com)) e dall'altro le aziende stanno acquistando maggior consapevolezza dell'impatto della loro reputazione online. Quest'ultimo aspetto trova riscontro nella pubblicazione di articoli sul tema sui quotidiani più autorevoli (a titolo di esempio si possono citare [Pattison, 2009], [Gibson, 2010]), ma anche nell'ampia offerta di servizi di on-line (o web) *reputation analysis* e *monitoring*.

Il giudizio complessivo che emerge online per un prodotto o un servizio viene indicato anche come '*sentiment*' e la sua analisi permette all'azienda di ricavare informazioni utili in termini di *customer relationship management*, sia per comunicare in modo efficace con i clienti (gestione del *complaint* digitale e promozione da parte dei clienti soddisfatti), che per intervenire su prodotti o servizi in modo mirato, e in ultima istanza per le strategie di qualità [Juran e Gryna, 1980], [Juran, 1995]. Per una destinazione turistica, un'analisi costante e mirata dei contenuti pubblicati sul Web che la riguardano può rappresentare un importante vantaggio competitivo, oltre che aumentare la qualità delle relazioni e della comunicazione tra tutti i soggetti coinvolti.

Esistono numerosi strumenti che possono essere usati per l'analisi della reputazione online. Una prima classificazione distingue la ORM (*Online Reputation Management*) rispetto alle citazioni - o menzioni - nei risultati prodotti dai motori di ricerca da quelle pubblicate sugli spazi Web 2.0. Per le prime si possono applicare strumenti di *alert*, quali ad esempio Google Alert o Yahoo Alerts. Sono strumenti che consentono di ricevere una notifica - ad esempio via email con una frequenza definita dall'utente - della pubblicazione sul Web di contenuti relativi alla propria azienda od organizzazione. Si basano su interrogazioni automatiche legate ad un insieme predefinito (e personalizzato) di parole chiave. Il vantaggio di questo tipo di strumenti è che nella versione più semplice sono gratuiti, ma le loro prestazioni sono limitate (con valori bassi sia per la precisione che per il *recall*). Sistemi più sofisticati sono in grado di monitorare le piattaforme di social networking più diffuse, contenuti pubblicati in più lingue, applicando tecnologie di *Information Retrieval* e di analisi linguistica di diversa complessità. Tuttavia, la maggior parte di questi sistemi presenta limiti dovuti prevalentemente ad un approccio sintattico o solo limitatamente semantico alla individuazione delle menzioni e all'analisi del *sentiment*.

In questo lavoro si propone l'applicazione di tecnologie semantiche per realizzare un'analisi degli UGC di tipo testuale che permetta di comprenderne il significato e valutare il giudizio espresso in relazione ad una organizzazione complessa. Per illustrare l'approccio si fa riferimento ad uno studio realizzato nell'ambito del turismo, un settore articolato, non riconducibile ad un singolo dominio [Franch, 2010], in cui l'offerta si caratterizza per il contributo di

numerosi operatori e la domanda, in continua evoluzione, è costituita da segmenti di mercato con esigenze ed aspettative diverse.

Il primo passo prevede la definizione di un modello (*topic model*) che permette di individuare i concetti sui quali si basa la reputazione di una specifica destinazione in base alle strategie promozionali della stessa. A ciascun concetto (*driver*) corrisponde un insieme di indicatori, da ricercare negli UGC rilevanti. L'analisi semantica permette di analizzare più accuratamente anche il sentiment espresso nelle menzioni, evitando i fraintendimenti dovuti all'uso di termini predefiniti in una determinata frase, per cui ad esempio, 'buono' corrisponde a un giudizio positivo, anche in una frase come 'buon quarto d'ora'.

Il resto del lavoro è strutturato come segue. La sezione 2 descrive i sistemi per l'analisi della reputazione online. Nella sezione 3 si illustra l'approccio proposto, evidenziando il ruolo delle tecnologie semantiche. La sezione 4 presenta il caso e infine, la sezione 5 è dedicata alle note conclusive e ai problemi aperti.

## **2. Sistemi per l'analisi della reputazione online**

### **2.1 Il contesto e le sfide**

Per le aziende, la rilevazione di opinioni, giudizi, impressioni, critiche espresse nel Web, può affiancare l'analisi dei dati acquisiti attraverso metodi di ricerca tradizionali per lo studio del grado di soddisfazione del cliente alla base delle attività di *customer care* e in generale dei processi decisionali e strategici. Rispetto agli strumenti tradizionali, che compendono indagini di mercato, interviste telefoniche, questionari allegati a prodotti e servizi, gli UGB offrono la possibilità di conoscere tempestivamente il giudizio incondizionato di un largo pubblico. Tuttavia, spesso il sentiment espresso su fonti aperte risulta difficilmente reperibile o utilizzabile per le aziende: da una parte per la numerosità dei contenuti che richiede un tempo eccessivo per ottenere una visione d'insieme, dall'altra per la difficoltà di identificare e filtrare i contenuti rilevanti per un'effettiva *marketing intelligence*. Da una conoscenza tempestiva e completa del sentiment dei propri clienti scaturisce la capacità di prevedere nuovi bisogni e tendenze e al contempo di apprendere eventuali segnalazioni o problemi che rispettando il normale iter richiederebbero tempistiche molto più lunghe prima di poter essere presi in considerazione. Tutto ciò è tanto più vero quando ci si focalizza sul mondo dei servizi: è qui che le esigenze del cliente diventano prioritarie e le modalità con cui vengono espresse sono assai diverse e personalizzate. Una mera raccolta di istanze, di occorrenze degli UGC (ottenuta ad esempio utilizzando uno strumento di Alert) offre un punto di vista esclusivamente quantitativo che non delinea la reazione del pubblico nei confronti dell'offerta. Ciò che risulta prioritario, non solo in termini di *time effort*, ma soprattutto in termini di efficacia, è un'analisi qualitativa che offra una visione di insieme sul sentiment relativamente all'offerta e al tempo stesso la possibilità di dettagliare tale sentiment, così da poter avere sotto controllo eventuali situazioni critiche.

L'ambito del marketing territoriale e turistico risulta essere uno dei settori più sensibili per sua natura alle esigenze dei clienti. Tuttavia, non è affatto semplice ricevere un feedback strutturato o strutturabile da parte dei turisti, anche perché ogni giudizio per definizione esprime una visione od emozione soggettiva e in quanto tale esprimibile in diversi modi linguistici. Per quanto riguarda il linguaggio, le problematiche per l'analisi automatica degli UGC possono essere classificate in primo luogo rispetto ad ortografia, sintassi e semantica:

- ortografia: esistono svariate forme 'colorite' e regionali; abbreviazioni, che contraddistinguono il linguaggio degli SMS così come quello di forum e dei blog;
- sintassi: un giudizio può essere espresso in una semplice frase, ad es., "mi piace", oppure può essere il risultato di un ragionamento complesso articolato su più frasi;
- semantica: lo studio del significato delle parole mostra la complessità del linguaggio, continua fonte di potenziale ambiguità; due parole scritte nello stesso modo possono riferirsi a due concetti diversi (come calcio inteso come sport o come elemento chimico); due parole scritte in modo diverso possono fare riferimento allo stesso oggetto (come ad esempio, casa e abitazione).

Un sistema semantico, in grado di comprendere il significato dei testi, permette di:

- focalizzarsi sui contenuti dei testi, e quindi su argomenti, concetti, entità salienti, indipendentemente dal modo in cui sono stati espressi e a prescindere dal formato, archivi, ecc.
- garantire un livello più alto di precisione nell'interpretazione del significato delle parole: le tecnologie più comuni per l'elaborazione di un testo scritto nel linguaggio comune raggiungono un grado di precisione che si aggira fra il 45% e il 55%. Alcuni sistemi per applicazioni specifiche o su domini ristretti- arrivano a percentuali più alte, con punte intorno al 70%. Con l'analisi semantica, la precisione può superare il 90% indipendentemente dal contesto e dalla quantità di documenti.

## 2.2 Strumenti

L'esigenza di sviluppare sistemi per supportare l'analisi della reputazione online è stata sottolineata da molti autori. Una descrizione dettagliata ed approfondita delle tecnologie applicabili e degli approcci esistenti si trova in [Pang e Lee, 2008]. Un lavoro antecedente che ha posto le basi per affrontare il tema della reputazione online è [Resnick e al., 2000], in cui gli autori fanno riferimento ai sistemi applicati da alcuni siti di successo. I lavori pubblicati più recentemente affrontano i vari aspetti dello sviluppo e realizzazione di un sistema di *online reputation monitoring* e propongono approcci sempre più raffinati sia per l'individuazione delle menzioni rilevanti che per l'analisi del sentiment. Un quadro dello stato dell'arte si può rilevare dalle pubblicazioni di due recenti workshop, dedicati alle tecnologie linguistiche per ORM e all'analisi del sentiment [Antin e al., 2011], [Balahur e al., 2011]. Tuttavia, nessuno dei sistemi proposti adotta un sistema di *natural language processing* semantico

basato su una base di conoscenza sufficientemente allargata da poter essere applicato a domini diversi, sia pur con adattamenti, per tener conto del linguaggio e delle conoscenze specifiche. Molti si limitano al riconoscimento delle cosiddette *named entity* (nomi geografici, indirizzi, ecc.), o recuperano la semantica dei termini nei commenti collegandoli ad un certo numero di concetti; complessivamente si tratta di approcci che potremo definire *semantically-flavoured*. Fra i sistemi commerciali, un'analisi semantica è effettuata da BlogMeter (<http://www.blogmeter.it>); da citare inoltre il progetto realizzato per il Comune di Milano [Barbesino e al., 2010] che si basa su una piattaforma integrata per l'analisi semantica di più fonti. Per quanto riguarda l'analisi della reputazione delle destinazioni turistiche, uno studio recente ha analizzato l'impatto degli UGC restituiti dai motori di ricerca [Inversini e Cantoni, 2011], evidenziando la varietà delle fonti rilevanti per una destinazione.

### 3. Il nostro approccio

#### 3.1 Il processo di analisi della reputazione online

Il processo di analisi della reputazione online di una destinazione turistica prevede tre fasi principali:

- **individuazione delle fonti web** sulle quali individuare gli UGC collegati alla destinazione tenendo conto dei profili del target dell'offerta turistica e di altri criteri volti ad ottimizzare l'efficacia delle strategie di comunicazione e promozione;
- **definizione e validazione dei concetti** (*driver*) da localizzare all'interno delle opinioni degli utenti Web raccolte;
- **analisi linguistica delle opinioni** per riconoscere e classificare i contenuti rilevanti rispetto agli elementi caratterizzanti l'offerta turistica culturale e valutazione del sentiment.

Per le destinazioni turistiche, i siti di riferimento appartengono a due gruppi (linee guida per l'identificazione delle fonti da monitorare per un hotel sono contenute in [Tourism BC, 2011]):

- social network e altri siti 'generalisti' più usati nelle aree geografiche da cui provengono i turisti (flussi *incoming*), ad esempio Facebook e Twitter perché hanno un largo pubblico (per alcuni mercati sono diversi; ad esempio, per la Cina, alcuni siti sono proibiti ed esistono altri social network più diffusi di quelli citati);
- siti collegati al settore del turismo (denominati anche Travel 2.0), perché frequentati dai turisti per programmare, realizzare e condividere le esperienze di vacanza. Fra questi, Tripadvisor è uno dei più noti, ma ne esistono moltissimi, fra cui si possono citare anche Zoover, Wayn, Triplt.

Il **secondo passo** è il più critico, perché si tratta di focalizzare l'attenzione sulle strategie promozionali della destinazione per stabilire quali sono gli UGC di interesse. Se, ad esempio, si volesse verificare la reputazione online della destinazione rispetto alle attività sportive praticabili per promuovere un'offerta turistica caratterizzata da una vacanza attiva, tipicamente per attirare un segmento di turisti più giovani (un problema tipico delle destinazioni alpine per

le vacanze estive), occorrerà identificare contenuti collegati a tale tema. Si tratta in sostanza di definire un 'topic model'.

A titolo di esempio, la tabella 1 illustra uno dei 'topic' individuati per il modello sviluppato per il caso relativo al Trentino come destinazione turistica culturale. Dalla definizione dei driver di riferimento dipendono le analisi che possono essere effettuate sui dati.

**Tabella 1 - Musei ed istituzioni culturali**

<b>Brand, Dominio</b>	<b>Line: Musei e Istituzioni</b>	<b>Product, Event</b>	<b>Feature part: Microeventi</b>	<b>Qualità</b>
Cultura-turismo in Trentino	MART (Competitor: Museo Madre Napoli, GAM Torino, Museo di Arte Contemporanea Basilea, Museion Bolzano, Guggenheim Bilbao, MOMA NY	Le varie mostre	Micro eventi: conferenze, Testimonial, Vernissage	Costi Accessibilità: Parcheggio, navette, collegamenti vari, diversamente abili, location, edificio, estetica, allestimento, contenuti, flessibilità, orari, servizi accessori:
	Castello Buonconsiglio (Eventuali Competitor)			book, cataloghi, ristorante, varietà dell'offerta: selezione, divertimento, intrattenimento
	Elenco principali musei trentini con eventuali commentatori: Museo guerra Rovereto, Casa d'Arte futurista, ecc.			

Come si può vedere dalle tabella, a ciascuno dei driver individuati per descrivere l'offerta culturale del Trentino è stata associata una tassonomia che permette di rappresentare in modo calibrato gli elementi di interesse: sufficientemente dettagliati, ma non troppo, per evitare, rispettivamente classificazioni generiche o forzate o di frammentare l'analisi a cause di un'alta 'sparsità' dei dati.

Il **terzo passo**, analisi e classificazione degli UGC rispetto a contenuto e sentiment, si basa sull'applicazione di tecnologie semantiche in grado di comprendere quali menzioni sono rilevanti per un determinato driver e quale, se presente, è il giudizio dell'autore rispetto ad esso. Le tabelle (tassonomie) definite per una determinata destinazione e strategia della stessa costituiscono la base per la classificazione delle menzioni, per evidenziare, ad esempio, che uno post su un blog riguarda una specifica esposizione (*Product, Event*) di uno dei musei (*Line: Musei e Istituzioni*) nella destinazione, l'acquisto del catalogo (*Qualità*) in occasione del vernissage (*Feature part: Microeventi*). In generale, un driver, è descritto da una quintupla ordinata costituita da:

<Brand/dominio, Line, Product/Service, Feature/Part, Quality>

Ad ognuna delle informazioni riconducibili a un ramo della tassonomia, viene assegnato un giudizio in senso assoluto normalizzandolo secondo uno schema predefinito. Nel caso descritto nella sezione successiva si è applicato lo schema seguente: 1 = very poor; 2 = poor; 3 = fair; 4 = good; 5 = excellent.

### 3.2 Il sistema per l'analisi semantica dei testi

Rispetto al processo di analisi della reputazione online introdotto nella sezione 3.1, la piattaforma adottata per il progetto, Cogito® (<http://www.expertsystem.net/>) supporta le seguenti attività:

- **reperimento delle informazioni:** agenti software di acquisizione dati, consentono di prelevare e raccogliere tutti i contenuti da analizzare, indipendentemente dalla provenienza e dal formato (*crawling*); in particolare, tali agenti producono post in formato normalizzato aventi la seguente struttura:
  - nome della fonte; localizzatore universale della fonte (URL o equivalente);
  - titolo della discussione (thread); localizzatore universale della discussione (URL o equivalente);
  - localizzazione contestuale della discussione (rispetto alla fonte);
  - titolo del post; localizzazione universale del post (URL o equivalente);
  - localizzazione contestuale del post (rispetto alla discussione);
  - data e ora di pubblicazione; autore; testo; firma dell'autore.
- **individuazione dei driver**, dove per driver si intende una categoria o un insieme di categorie a cui può essere associato un giudizio;
- **analisi linguistica:** sui dati prelevati si svolge l'analisi linguistica completa - morfosintattica, grammaticale, logica, semantica - per estrarre le informazioni necessarie per le successive analisi (un esempio è riportato in Fig.1).
- **interrogazione e reporting:** l'utente finale può visualizzare ed analizzare tutte le informazioni raccolte, attraverso interfacce di interrogazione e monitoraggio. La possibilità di strutturare i dati e renderli facilmente navigabili, permette analisi di vari tipi, sia da parte del top management, che dell'operatore di sportello, senza richiedere competenze tecniche o applicative specifiche.

Per quanto riguarda il sentiment, per interpretare le informazioni individuate e scaricate automaticamente, Cogito applica delle regole che consentono l'identificazione dei giudizi sui driver individuati con la tassonomia. Un esempio è riporta in Fig.2.



Fig.1 – Esempio analisi linguistica di un post

Per un singolo post, viene effettuata l'**estrazione** e la **normalizzazione** dei dati elementari utili, cioè di tutte le espressioni che possono indicare driver e delle espressioni che possono indicare rating; successivamente, viene

effettuata la **composizione** dei giudizi, ossia l'**attribuzione** dei rating ai driver a partire dai dati elementari estratti nella fase precedente.

```
SCOPE SENTENCE IN SEGMENT(AUTHORIAL_TEXT, Comparison)
{
  // non male, non malaccio
  IDENTIFY(FAIR_ADJ)
  {
    @FAIR_ADJ[KEYWORD[("non male", "non malaccio", "non è male", "non è malaccio", "non
è male", "non è malaccio", "nella media", "nella norma", "niente di speciale", "niente di che",
"senza infamia e senza lode", "bellino", "bellina")]
  }
}
```

**Fig.2 – Esempio di regola per la valutazione del sentiment dei post**

La Fig.3 riporta l'analisi complessiva per il seguente post:

*"Il MART è uno dei pochi edifici di Botta che mi piaccia 😊"*

```
BRAND -DOMINIO = MART;  
LINE = MUSEI ED ISTITUZIONI;  
PRODUCT - EVENTI = (GENERAL);  
feature/part - MICROEVENTI = (GENERAL);  
quality = ESTETICA EDIFICIO  
rating: 4
```

**Fig.3 – Analisi dei contenuti e del sentiment di un post**

#### 4. Il caso di studio

L'approccio proposto nella sezione precedente è stato applicato in un progetto per il Trentino. l'obiettivo del progetto è l'analisi in tempo reale e in modo continuo del sentiment che i clienti esprimono liberamente nel web a proposito della qualità dell'offerta turistica del Trentino. Le assunzioni principali hanno riguardato due aspetti fondamentali:

- la scelta di ambiti applicativi significativi sul piano strategico e focalizzati su domini trattabili dal punto di vista computazionale: cultura e sport;
- l'adozione di un sistema di analisi del linguaggio che rappresenta lo stato dell'arte per l'area di *Natural Language Processing*, la piattaforma Cogito.

Nella prima fase del progetto si è focalizzata l'attenzione sulle due aree individuate, approfondendo l'analisi per cinque macro-categorie per ognuna di esse: 5 discipline sportive e 5 eventi, creando una sesta categoria sia per le altre discipline sportive che per le attrazioni culturali non incluse nelle voci esplicitate. I post relativi alle prime 5 categorie sono stati analizzati in dettaglio secondo i topic model sviluppati per ciascuno di essi. A tal fine si è personalizzata la rete semantica (SENSIGRAFO®) al fine di adeguarla al linguaggio da interpretare. Un'analisi preliminare ha evidenziato la presenza di



un numero di post relativi alle aree individuate per l'offerta turistica trentina e di alcuni competitor nell'ordine di qualche decina di migliaia di post, indice della fattibilità ed utilità di un sistema automatico di analisi della reputazione online.

Si sono identificate 5 fonti, fra cui Googlegroups e Googleblogs, che da sole sono risultate contenere il 50% dei post sull'offerta culturale del Trentino (sul totale dei giudizi espressi nel mondo web, estrapolati tramite aggregatori che identificano solo i post che esprimono sentiment sul dominio di riferimento).

Lo studio ha permesso di misurare e valutare il positivo ed utile apporto dell'analisi dei cosiddetti UCG per il turismo trentino. Sulla base delle fonti analizzate, sono emerse per la destinazione alcune indicazioni e tendenze precise. Fra queste, una forte domanda di 'natura', 'tranquillità' e 'tipicità', che talvolta stridono con l'organizzazione di grandi eventi. L'analisi dei post individuati potrebbe contribuire a chiarire come l'offerta turistica trentina possa far convivere queste due caratteristiche della domanda turistica. Altre informazioni sono utili sono state ottenute riguardo alla percezione dell'offerta turistica rispetto a quella dei concorrenti.

## 5. Conclusioni

Punti di forza dell'approccio proposto sono la completa trasparenza del funzionamento del sistema di analisi della reputazione online nei confronti dell'utente finale e l'analisi semantica di testi. Quest'ultima consente di ricavare dagli UGC informazioni utili: per comprendere le esigenze dei turisti, come sono percepiti i *competitors* e verificare se coincidono con quelli attesi; individuare gli aspetti critici su cui intervenire e quelli positivi da valorizzare; identificare 'opinion leader' da contattare e coinvolgere; raccogliere feedback su specifici pacchetti o prodotti dell'offerta turistica. L'individuazione delle *named entity*, e quindi di località, eventi, hotel, ecc. consente di rispondere a domande su quali siano le località o gli eventi citati più frequentemente e se sono quelli attesi; quali siano i siti più suggeriti; se vengono date indicazioni su operatori turistici, ecc.

Sul piano metodologico, queste informazioni confermano l'utilità da parte delle *Destination Management Organization* (DMO) di adottare un efficace sistema di analisi della reputazione online, stabilendo procedure e strumenti per agire a due livelli: operativo, per attività di CRM, rispondendo adeguatamente ai singoli post; strategico, per la pianificazione della promozione e dello sviluppo della destinazione, identificando eventuali trend o minacce per il successo della stessa. L'applicazione di tecnologie semantiche permette a una DMO di svolgere anche un ruolo di coordinamento, importante per il web marketing, laddove venisse utilizzata per segnalare agli operatori sul territorio le opinioni e i commenti che li riguardano (ad esempio, ai singoli alberghi, o alle aziende di promozione turistica d'ambito).

Sviluppi futuri riguardano una valutazione complessiva dei risultati in termini di recall e precision sui tre livelli di applicazione delle tecnologie semantiche: identificazione dei post pertinenti, identificazione dei contenuti nei post e attribuzione del sentiment, laddove i risultati dello studio condotto, benché

preliminari confermano le prestazioni che la tecnologia semantica adottata ha ottenuto negli ambiti applicativi finora affrontati.

**Ringraziamenti:** si ringraziano Nicola Zeni per la collaborazione alla definizione dei topic model e la Trentino Spa.

## Bibliografia

Antin J., Churchill E.F., Chen Bee-Chung, Workshop on online reputation: context, privacy, and reputation management. In Proc. 20th Int. Conf. companion on World wide web - WWW'11, ACM, New York, 2011, 321-322.

Balahur A., Boldrini E., Andres Montoyo A., Patricio Martinez-Barco P. (eds), Proc. 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis - WASSA2011, Association for Computational Linguistics, Portland, Oregon, 2011

Barbesino P., Francalanci C., Petrovich F., Social Media Intelligence Comprendere il "Polso" di un Territorio, Mondo Digitale, AICA, Milano, 45, 2010, 52-63.

Franch M. (ed), Marketing delle destinazioni turistiche. Metodi, approcci e strumenti, McGraw-Hill, Milano, 2010, 277-324.

Gibson M., Repairing Your Damaged Online Reputation: When Is It Time to Call the Experts?, Time Business, April 19, 2011, [www.time.com/time/business/article/0,8599,2065597,00.html#ixzz1a7V5Zlbc](http://www.time.com/time/business/article/0,8599,2065597,00.html#ixzz1a7V5Zlbc)

Inversini A., Cantoni L., Towards online content classification in understanding tourism destinations' information competition and reputation, Int. J. Internet Marketing and Advertising, 6(3): 282-299, 2011.

Juran J., Gryna Jr F.M., Quality planning and analysis, McGraw-Hill, 1980.

Juran J.M., A History of Managing for quality, ASQC Quality Press, Milwaukee, WI, 1995.

Mich L., La web presence delle destinazioni turistiche, in Garibaldi R., Peretta R. (eds), Facebook in Tourism, eBook, FrancoAngeli, Milano, 2011.

O'Reilly, What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software, 2005, <http://oreilly.com/web2/archive/what-is-web-20.html>

Pang B., Lee L., Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1-135, 2008.

Pattison K., Managing an Online Reputation, The New York Times, July 29, 2009, <http://www.nytimes.com/2009/07/30/business/smallbusiness/30reputation.html?adxnnl=1&pagewanted=all&adxnnlx=1318010523-IMI2YTLQVnuQFQPY4Z3BvA>

Resnick P., Zeckhauser R., Friedman E., Kuwabara K., Reputation Systems, Communications of the ACM, 43(12): 45-48 - 2000.

Tourism British Columbia, Online reputation management, 2011, [www.jti.gov.bc.ca/industryprograms/pdfs/OnlineReputationManagementTBEGuide2011\\_Jun15.pdf](http://www.jti.gov.bc.ca/industryprograms/pdfs/OnlineReputationManagementTBEGuide2011_Jun15.pdf)