

**Politecnico di Milano**  
**Scuola di Ingegneria Industriale e**  
**dell'Informazione**



**Corso di Laurea in Ingegneria Informatica**  
**Dipartimento di Elettronica, Informazione e**  
**Bioingegneria**

**SVILUPPO ED INTEGRAZIONE DI UN**  
**MOTORE SEMANTICO PER IL**  
**RILEVAMENTO E LA**  
**GEOLOCALIZZAZIONE DI EMERGENZE,**  
**NELL'AMBITO DEL PROGETTO TORCIA**

**Relatore: Prof. Chiara Francalanci**  
**Correlatore: Ing. Alessandro Poli**

**Tesi di Laurea di: Federico Negri      matr. 779318**  
**Luca Muccignato      matr. 783274**

**Anno Accademico 2013-2014**



*Desideriamo innanzitutto ringraziare la Prof.ssa Chiara Francalanci per la sua costante disponibilità, i preziosi insegnamenti e la sua simpatia che ha reso piacevoli le numerose ore dedicate alla nostra tesi.*

*Ringraziamo, inoltre, il Dr. Ing. Alessandro Poli per averci seguito in maniera continua durante il presente lavoro, fornendoci costantemente suggerimenti e consigli tecnici utili.*

*Intendiamo poi ringraziare anche le imprese afferenti al progetto TORCIA, per averci dato la possibilità di partecipare attivamente alle riunioni di brainstorming e aver condiviso con noi informazioni e dati utili per la realizzazione di questa tesi.*

*Un ringraziamento particolare va alle nostre famiglie che ci hanno appoggiato e sostenuto durante tutti questi anni di studio, sempre presenti nei momenti di bisogno.*

*Milano, 23 Marzo 2014*

*Federico e Luca*



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Contributi . . . . .	2
1.2	Organizzazione . . . . .	4
<b>2</b>	<b>Stato dell'arte</b>	<b>5</b>
2.1	La gestione delle emergenze . . . . .	5
2.1.1	Ciclo di vita delle emergenze . . . . .	6
2.1.2	I piani d'emergenza . . . . .	7
2.1.3	Classificazione delle emergenze . . . . .	8
2.2	I social network e il loro ruolo nella gestione delle emergenze . .	11
2.2.1	I social network . . . . .	11
2.2.1.1	Twitter . . . . .	13
2.2.2	Il Crowdsourcing . . . . .	15
2.2.2.1	Il Crowdsourcing e le fasi dell'emergenza . . . .	16
2.2.3	I social network come possibile aiuto nella gestione delle emergenze . . . . .	17
2.2.3.1	I Social Media e l'Emergency Management secondo l'OECD . . . . .	17
2.2.3.2	I Social Network al Festival del Volontariato 2014	19
2.2.4	Casi reali di utilizzo dei social network durante le emergenze . . . . .	20
2.2.4.1	Una nuova strategia di comunicazione: il caso NORAD NORTHCOM . . . . .	20
2.2.4.2	Uragano Gustav in Louisiana (Agosto 2008) . .	21
2.2.4.3	Eruzione del vulcano islandese Eyjafjöll (Aprile 2010) . . . . .	21

## INDICE

---

2.2.4.4	Giappone: terremoto, tsunami, e disastro nucleare (Marzo 2011) . . . . .	22
2.2.4.5	Terremoto dell'Emilia (Maggio 2012) . . . . .	23
2.2.4.6	Uragano Sandy . . . . .	23
2.2.4.7	Alluvione in Sardegna (Novembre 2013) . . . . .	25
2.2.4.8	American Red Cross Digital Operations Center	26
2.2.4.9	Emergency Situation Awareness - Automated Web Text Mining (ESA-AWTM) . . . . .	27
2.2.4.10	Una piattaforma di crowdsourcing: Ushaidi . . . . .	28
2.2.4.11	Emergenza24 . . . . .	30
2.3	Analisi conclusiva . . . . .	31
<b>3</b>	<b>Il Progetto TORCIA</b>	<b>33</b>
3.1	Motivazioni e obiettivi . . . . .	34
3.2	Il contesto specifico di analisi del progetto . . . . .	36
3.2.1	Scelta della sorgente . . . . .	37
3.2.2	I dati in input . . . . .	38
3.2.3	Scelta del contesto . . . . .	39
3.3	I partner del progetto . . . . .	40
3.3.1	Alcatel-Lucent . . . . .	41
3.3.2	ACT Solutions . . . . .	42
3.3.3	Beta 80 Group . . . . .	43
3.3.4	Vidienne Services . . . . .	43
3.3.5	Dipartimento di Elettronica, Informazione e Bioingegneria del Politecnico di Milano . . . . .	45
3.3.6	Fondazione Politecnico di Milano . . . . .	46
3.4	Le attività core del sistema . . . . .	47
3.5	Implementazione del progetto TORCIA: piano di attività . . . . .	49
3.5.1	WP1 – Requisiti del sistema, architettura e analisi degli scenari . . . . .	49
3.5.2	WP2 – Progettazione e realizzazione del sistema raccolta dati . . . . .	50
3.5.3	WP3 – Progettazione del motore di analisi . . . . .	50

3.5.4	WP4 – Progettazione e realizzazione del motore di calcolo per l’indirizzamento ed il supporto alle decisioni .....	51
3.5.5	WP5 – Progettazione e realizzazione del cruscotto di monitoraggio e framework di distribuzione delle informazioni con supporto cartografico .....	51
3.5.6	WP6 – Applicativo mobile web 2.0 .....	52
3.5.7	WP7 – Progettazione infrastruttura di rete .....	52
3.5.8	WP8 – Integrazione, dimostrazione e validazione .....	52
3.5.9	WP9 – Disseminazione .....	53
3.6	L’architettura core del sistema .....	54
3.6.1	Input del sistema .....	54
3.6.2	L’analisi semantica .....	55
3.6.3	Output del sistema .....	55
3.6.4	Destinatari del sistema .....	56
3.7	Complessità di progetto e impieghi futuri .....	56
3.8	Il nostro lavoro di tesi .....	59
<b>4</b>	<b>Il Motore Semantico</b> .....	<b>61</b>
4.1	Funzionamento Generale .....	62
4.2	Architettura del Motore Semantico .....	64
4.2.1	Il Web Service .....	65
4.2.2	Database .....	65
4.2.3	Motore Semantico .....	68
4.2.3.1	Classificatore .....	68
4.2.3.2	Entity Discovery .....	69
4.3	Web Service .....	69
4.3.1	Cos’è un Web Service .....	69
4.3.2	SOAP e REST .....	72
4.3.3	Architettura di PUSH TWEET e PUSH LUOGHI .....	75
4.3.4	Funzionalità del servizio PUSH TWEET .....	76
4.3.5	Sviluppo .....	80
4.3.6	Funzionalità del servizio PUSH LUOGHI .....	81
4.4	Motore Semantico: Classificatore .....	82
4.4.1	Algoritmo di Carcaci .....	83

## INDICE

---

4.4.2	Modifiche alle Classi . . . . .	84
4.4.3	Risultati Ottenuti . . . . .	86
4.4.4	Riprogettazione del Tool di Classificazione . . . . .	89
4.4.5	Considerazioni sul Tool di Classificazione . . . . .	89
4.5	Motore Semantico: Entity Discovery . . . . .	90
4.5.1	Descrizione generale . . . . .	90
4.5.2	Metodologia di sviluppo . . . . .	92
4.5.3	Descrizione dettagliata dell’algoritmo . . . . .	93
4.5.3.1	Schema generale per il rilevamento di un emer- genza . . . . .	93
4.5.3.2	Schema generale per la gestione della segnala- zione . . . . .	97
4.5.3.3	Schema generale per il rilevamento di un’allerta	103
4.5.3.4	Soglie . . . . .	106
4.5.3.5	Gestione della Localizzazione . . . . .	108
4.5.3.6	Utilizzo di Blacklist . . . . .	114
4.5.3.7	File di Configurazione . . . . .	115
4.6	Tool di Monitoraggio . . . . .	117
4.6.1	Architettura di Torcia Monitor . . . . .	118
4.6.2	Monitoraggio del Motore Semantico . . . . .	118
4.6.3	Monitoraggio del Servizio PUSH TWEET . . . . .	120
4.6.4	Statistiche . . . . .	121
4.6.4.1	Tweet . . . . .	122
4.6.4.2	Segnalazioni . . . . .	122
4.6.4.3	Possibili sviluppi per Torcia Monitor . . . . .	123
4.7	Possibili nuove funzionalità . . . . .	124
<b>5</b>	<b>Verifiche Empiriche e Miglioramenti</b>	<b>127</b>
5.1	Analisi mensile . . . . .	127
5.1.1	Gennaio 2013 . . . . .	128
5.1.2	Febbraio 2013 . . . . .	131
5.1.3	Marzo 2013 . . . . .	133
5.1.4	Aprile 2013 . . . . .	136
5.1.5	Maggio 2013 . . . . .	138
5.1.6	Giugno 2013 . . . . .	140



## INDICE

---

5.1.7	Luglio 2013 . . . . .	142
5.1.8	Agosto 2013 . . . . .	144
5.1.9	Settembre 2013 . . . . .	146
5.1.10	Ottobre 2013 . . . . .	147
5.1.11	Novembre 2013 . . . . .	149
5.1.12	Dicembre 2013 . . . . .	151
5.2	Analisi di Allerte e Segnalazioni . . . . .	153
5.2.1	Nubifragio Catania - 21 Febbraio . . . . .	154
5.2.2	Alluvione Vicenza - 15 Maggio . . . . .	155
5.2.3	Nubifragio Roma - 7 Luglio . . . . .	156
5.2.4	Nubifragio Roma - 21 Luglio . . . . .	157
5.2.5	Nubifragio Roma - 27 agosto . . . . .	158
5.2.6	Allagamenti Toscana - 21 Ottobre . . . . .	159
5.2.7	Alluvione Sardegna - 18 Novembre . . . . .	160
5.3	Riflessioni post-testing . . . . .	161
<b>6</b>	<b>Conclusioni e sviluppi futuri</b>	<b>165</b>
<b>A</b>	<b>Categorie di classificazione</b>	<b>167</b>
A.1	Categorie di partenza . . . . .	167
A.2	Categorie migliorate . . . . .	169
	<b>Elenco delle figure</b>	<b>173</b>
	<b>Elenco delle tabelle</b>	<b>175</b>
	<b>Bibliografia</b>	<b>177</b>



# Capitolo 1

## Introduzione

Col presente lavoro di tesi abbiamo realizzato uno strumento in grado di far leva sull'informazione da social media per aiutare le autorità competenti nell'identificazione preventiva di stati di emergenza e affiancare autorità e cittadini nelle fasi di gestione della crisi stessa, durante le quali è necessario poter inviare e ricevere informazioni e coordinarsi nel più breve tempo possibile. L'innovatività del nostro sistema risiede nel tipo di sorgente adottata per reperire i dati alla base delle analisi e dell'intero progetto, che non si limita ai tradizionali call center, ma integra i messaggi postati da utenti comuni su piattaforme di tipo social, Twitter in particolare.

L'affermazione dei social network e la contemporanea diffusione di internet e dei dispositivi mobili hanno modificato radicalmente il modo di fare informazione. Infatti i mass media vengono pian piano affiancati dai contributi degli utenti, secondo il concetto di Web 2.0. Diversi studi del settore hanno evidenziato come l'“informazione sociale” sia più immediata e fruibile rispetto ai classici canali di informazione. In particolar modo, l'utilizzo dei social media si rileva preponderante durante le situazioni di crisi, permettendo agli utenti di scambiarsi informazioni relative all'emergenza in atto, o di richiedere l'intervento delle autorità, soprattutto quando gli altri mezzi di comunicazione si rivelano inefficaci.

Uno dei principali contributi del lavoro di tesi è la messa a punto di un algoritmo semantico di allerta basato su informazione da social media e la sua verifica empirica su un campione di oltre un anno di dati in lingua

italiana riguardanti le emergenze. Il parlato sociale assume un'importanza fondamentale e si presenta quindi la necessità di avere a disposizione uno strumento che possa gestire i meccanismi di crowdsourcing, ovvero il modello secondo il quale le conoscenze di tutti i cittadini vengono sfruttate a fini comuni per la segnalazione e la risoluzione di problemi. Tale strumento può essere utilizzato dalle istituzioni per avere la possibilità di agire tempestivamente a sostegno del cittadino, sia in fase di allerta che in quella di emergenza. La rete e il suo popolo assumono, quindi, un doppio ruolo: da un lato, possono diventare contributori attivi, segnalando le situazioni critiche in atto sul territorio e aiutando perciò le autorità nell'analisi dei rischi e/o delle conseguenze connesse ad una emergenza; dall'altro, rimangono destinatari dell'informazione istituzionale, avendo però a disposizione un mezzo adeguato per la ricezione tempestiva di tale informazione.

La criticità delle emergenze in atto porta alla diffusione di altissimi volumi di informazione da parte degli utenti, che pertanto non possono essere trascurati. Per questo motivo è necessario raccogliere e gestire queste informazioni in maniera istantanea e automatica, portando ad enormi vantaggi sia per il cittadino che per le istituzioni. Con il nostro strumento, l'analisi delle conversazioni è effettuata in maniera sistematica, al fine di costituire un punto di partenza ben definito per la definizione di una metodologia operativa efficiente per la gestione delle emergenze in ambito social.

### 1.1 Contributi

Il presente lavoro di tesi si propone di creare una metodologia che, a partire da un'analisi dettagliata del parlato sociale, offra un sistema integrato per la gestione di eventi critici all'interno del contesto delle emergenze.

Il progetto si articola nei seguenti punti:

- Analisi del parlato con lo scopo di verificare la reale importanza dell'informazione sociale e i relativi volumi: in questo modo, è stato possibile identificare le caratteristiche delle informazioni, al fine di raffinare i passi successivi della metodologia.

## 1.1 Contributi

---

- Customizzazione del Tool di Classificazione, al fine di adattarlo all'ambito di studio da noi definito.
- Riprogettazione, sviluppo e testing di un sistema di Entity Discovery, attraverso il quale è possibile analizzare sintatticamente il parlato sociale e rilevare automaticamente la presenza di eventi critici in atto mediante deduzioni semantiche.
- Creazione di strumenti di supporto, quali ad esempio il Database i Web Service o il Tool di Monitoraggio, atti ad interfacciare il nostro Motore Semantico con le altre parti del sistema TORCIA in cui si inserisce il lavoro di tesi.

### 1.2 Organizzazione

Il presente elaborato di tesi è così organizzato:

- Nel capitolo 1 definiamo gli obiettivi e lo scopo della tesi.
- Il capitolo 2 riguarda l'analisi dello stato dell'arte, indicando esplicitamente il background informativo relativo alla gestione delle emergenze e all'utilizzo dei social network all'interno di tale contesto.
- Nel capitolo 3 viene presentato ed analizzato il progetto TORCIA, in cui si inserisce il nostro lavoro di tesi.
- Il capitolo 4 fornisce la descrizione completa del Motore Semantico e di tutte le parti che lo compongono, con particolare riferimento ai Web Service, al Tool di Classificazione e al Tool di Entity Discovery.
- Nel capitolo 5 presentiamo i testing effettuati sul nostro sistema, chiarendo il significato dei risultati ottenuti e le migliorie che ci hanno permesso di apportare.
- Infine, nel capitolo 6, sono riassunti tutti i risultati da noi raggiunti attraverso il nostro lavoro di tesi e definiti i possibili sviluppi futuri e migliorie del sistema ideato.

# Capitolo 2

## Stato dell'arte

Questo capitolo definisce il background informativo della gestione delle emergenze e dell'utilizzo dei social network all'interno di tale contesto.

La sezione 2.1 definisce cosa si intende per gestione delle emergenze, descrivendo nel dettaglio quali attività la caratterizzano.

La sezione 2.2 si focalizza, invece, sulla descrizione dei social network e del loro uso nella gestione delle emergenze.

Infine nella sezione 2.3 faremo un'analisi dello stato dell'arte, indicando quali sono i limiti e le soluzioni che proponiamo nel nostro lavoro di tesi.

### 2.1 La gestione delle emergenze

La gestione dell'emergenza può essere vista come un lungo processo di collaborazione tra enti e soggetti secondo le diverse funzioni e peculiarità. Nella sua rappresentazione classica, essa risulta in un processo ciclico formato da diverse fasi, come mostrato nella sezione 2.1.1. L'elemento che ritroviamo in tutte le fasi del processo è il piano d'emergenza (vedi sezione 2.1.2) che viene steso nella fase precedente all'emergenza vera e propria e che verrà utilizzato, invece, in fase operativa. Per una sua stesura corretta è necessario classificare i diversi tipi di eventi che possono generare l'emergenza, in modo da concentrarsi sulle loro caratteristiche e creare un piano operativo il più possibile efficiente. Le modalità di classificazione degli eventi sono descritti nella sezione 2.1.3.

### 2.1.1 Ciclo di vita delle emergenze

Il processo ciclico di gestione delle emergenze è formato da diverse fasi, nelle quali si indicano le azioni da intraprendere al fine di prepararsi all'evento, ridurre il più possibile gli effetti e riportare il più velocemente possibile la situazione ad un livello di normalità. Le quattro fasi principali che si possono individuare descrivono un processo ciclico e continuativo, scandito da eventi che determinano, in modo più o meno netto, il passaggio alla fase successiva, come rappresentato in Figura 2.1.



Figura 2.1: Le fasi di gestione dell'emergenza.

#### **Preparedness**

L'obiettivo principale di questa fase è quello di creare un adeguato livello di prontezza nel rispondere alle emergenze. Tale preparazione deve realizzarsi a diversi livelli organizzativi e decisionali (governo, regioni, comunità) attraverso un'approfondita analisi dei rischi, al fine di valutare la possibilità che un evento si verifichi, seguita da una fase di diffusione della conoscenza e di creazione di scenari di rischio e piani di emergenza condivisi.

#### **Response**

La fase di response è la fase operativa il cui inizio è determinato dal verificarsi dell'evento calamitoso. In questa fase si mettono in pratica le procedure e i processi descritti nei piani di emergenza a partire dall'attivazione delle risorse disponibili e dalla ricerca di feriti e superstiti, alla creazione di rifugi, luoghi di ricovero e centri operativi di controllo e coordinamento. L'esito positivo della



## 2.1 La gestione delle emergenze

---

stessa è strettamente correlato con la gestione real-time di tutti gli aspetti fisici e naturali dell'evento che differiscono da quelli previsti in sede di definizione del Piano di Emergenza. Tale fase è anche un'occasione di verifica dove si può valutare tutto il lavoro svolto in quella precedente e di conseguenza migliorarne i punti di maggior criticità.

### Recovery

L'obiettivo di tale fase è il ritorno alla normalità nel più breve tempo possibile e si attiva non appena si è chiusa l'emergenza in senso stretto. È il momento chiamato comunemente "*ricostruzione*" nel quale si gettano le basi per evitare il ripetersi dell'emergenza o limitarne, in futuro, gli effetti dannosi su persone e cose.

### Mitigation

È la fase di chiusura del processo ed è di lungo periodo; si ricercano contromisure a tutti i livelli per evitare il ripetersi dell'evento calamitoso o per ridurre le conseguenze, in accordo e continuazione con la fase di recovery.

La fase di mitigation deve sfociare in una fase di preparedness, iniziando così di nuovo il ciclo con l'obiettivo di evitare il rischio o, nel caso in cui ciò non sia possibile, migliorare la risposta e ridurre gli effetti negativi.

### 2.1.2 I piani d'emergenza

Il piano di emergenza è un documento che rappresenta il supporto operativo ed è utilizzato dalle autorità per la gestione delle emergenze. È il risultato di uno studio approfondito sulle vulnerabilità del territorio e contiene al suo interno la definizione di alcuni probabili scenari di rischio. Assume la forma di un progetto che, a partire da eventi o situazioni contingenti, descrive lo stato e il workflow delle attività da svolgere per garantire un celere intervento e un conseguente ritorno alle normali condizioni.

Durante il suo ciclo di vita devono essere svolte in maniera continuativa tre attività:

- l'aggiornamento periodico dello stesso, in quanto è evidente che lo scenario di rischio risente della dinamicità dell'assetto del territorio;

- l'attuazione di esercitazioni al fine di verificare la sua validità;
- l'informazione alla popolazione per far sì che essa sia cosciente del rischio e addestrata all'evenienza.

Il modello del piano di emergenza descritto è il “*Metodo Augustus*”, creato dal geologo Elvezio Galanti in base alle disposizioni contenute nella legge 255/92 [1].

### 2.1.3 Classificazione delle emergenze

Eeguire una classificazione delle emergenze, ovvero indicarne i parametri fondamentali come origine, evoluzione ed entità coinvolte, è estremamente utile al fine di definire buone strategie per il contenimento e la risoluzione delle stesse. Infatti non tutti gli eventi critici che si verificano sul territorio richiedono piani di intervento predefiniti: alcuni possono essere incidenti banali, risolvibili senza nessuna preparazione preventiva, mentre in altri possono verificarsi casi di criticità più elevata per cui richiederanno piani operativi mirati e più elaborati.

Per tal motivo sono stati definiti diversi criteri di classificazione [14]:

**Secondo il Grado di Impatto:** Fornisce la misura quantitativa e qualitativa della potenza intrinseca di un certo evento critico. Viene descritto da tre dimensioni di analisi ovvero:

- *livello di gravità;*
- *estensione geografica;*
- *durata temporale.*

La combinazione dei valori che possono assumere queste tre dimensioni produce una suddivisione in tre classi principali, classificando il grado di impatto come *lieve*, *moderato* o *alto*.

**Secondo la Scala:** Una variante della classificazione precedente nella quale ci si concentra maggiormente sulla dimensione sociale all'interno della quale un dato evento impatta. Le tre dimensioni di analisi qui considerate, sono:

- *magnitudine*, cioè l'ampiezza dei soggetti ai quali l'evento ha imposto una modifica sostanziale delle abitudini di vita;

## 2.1 La gestione delle emergenze

---

- *scope*, cioè i confini sociali e geografici coinvolti;
- *duration*, cioè la durata degli effetti sociali;

Grazie a queste dimensioni, possiamo classificare un evento secondo il parametro di scala in *emergenza*, *disastro* o *catastrofe*.

**Secondo il Tipo:** Indica che cosa viene colpito e la suddivisione avviene tra *disastri comunitari*, dove viene colpito un ampio insieme di risorse fisiche e umane, e *disastri settoriali*, dove viene colpito un segmento specifico della comunità

**Secondo l'Anticipabilità:** Gli effetti prodotti da un evento sono molto diversi a seconda di quanto un evento risulti più o meno atteso, ovvero in base a quanto le organizzazioni risultino adeguatamente preparate riguardo ad esso. Le dimensioni di analisi in base a cui si valuta l'anticipabilità sono:

- *predicibilità*, cioè quanto l'evento è predicibile;
- *influenzabilità*, cioè quanto realistiche sono le strategie di riduzione dei danni dell'evento.

Combinando queste dimensioni possiamo definire un evento come *convenzionale* (facilmente influenzabile e predicibile), *inaspettato* (facilmente influenzabile ma difficilmente predicibile), *intrattabile* (difficilmente influenzabile ma facilmente predicibile) o *fondamentale* (difficilmente influenzabile e predicibile).

Una classificazione più generale è data dalla suddivisione di eventi critici in due tipologie[14]:

- *eventi di origine naturale*;
- *eventi di origine umana/tecnologica*.

### Eventi di origine naturale

Gli eventi di origine naturale sono quelli causati da fenomeni naturali. Sono tantissimi e di diversa natura e spesso sono strettamente connessi tra di loro poiché possono generarsi sia come una conseguenza dell'altro, sia contemporaneamente. I principali sono:

## Stato dell'arte

---

- *Inondazioni e alluvioni*: si tratta di un allagamento in tempi brevi (da ore a giorni) di un'area ben definita e abitualmente asciutta da parte di una massa d'acqua. Quando l'inondazione è provocata dalla tracimazione di corsi d'acqua ingrossati per piogge elevate si parla anche di alluvione;
- *Terremoti*: sono vibrazioni o oscillazioni improvvise della crosta terrestre provocate dallo spostamento improvviso di una massa rocciosa nel sottosuolo. I terremoti sono gli eventi naturali più potenti sulla terra e possono causare gravi distruzioni, alte perdite di vite umane e altri eventi come inondazioni, frane, incendi o fuoriuscite di materiali pericolosi. Se il sisma avviene sotto la superficie oceanica si possono generare maremoti o tsunami;
- *Uragano e Tornado*: si tratta di un ciclone tropicale, cioè un sistema tempestoso caratterizzato da un largo centro di bassa pressione e da numerosi temporali che producono forti venti e pesanti piogge. Invece una tromba d'aria o tornado è un violento vortice d'aria che si genera alla base di un cumulonembo e giunge a toccare il suolo;
- *Eruzioni vulcaniche*: consiste in una spaccatura della superficie terrestre in corrispondenza del quale si ha la fuoriuscita di magma tramite un'eruzione. Essa può causare distruzione e danni ingenti;
- *Frane*: si indicano tutti i fenomeni di movimento o caduta di materiale roccioso a causa della rottura di un equilibrio statico preesistente. Possono causare danni e perdite di vite umane e, se avvengono in prossimità di corsi d'acqua, possono provocare esondazioni.

### Eventi di origine umana/tecnologica

Questi tipi di eventi sono dovuti all'innovazione tecnologica e allo sviluppo umano. Possono essere causati dall'uomo volontariamente o involontariamente ma spesso hanno intrinsecamente una concausa naturale. I principali sono:

- *Dighe*: sono sbarramenti volti a modificare il corso dei fiumi o al contenimento di bacini idrici. I problemi con esse possono essere causati da fatto che sono mal progettate o che non sia stata fatta la manutenzione necessaria, e per tali motivi possono cedere, provocando ingenti danni.

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

Inoltre le forti piogge e le inondazioni possono originare uno straripamento oltre la diga o addirittura, nel caso in cui l'acqua assuma una forza dirompente, distruggerla;

- *Caduta di edifici*: può essere determinata da una cattiva progettazione o da eventi naturali come terremoti, frane, inondazioni;
- *Epidemie*: sono malattie infettive e sorgono in relazione a specifiche condizioni igienico-sanitarie.

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

In questa sezione tratteremo dei social network e di come possono essere utilizzati nella gestione delle emergenze.

Inizieremo dando una loro descrizione generale, spiegando, in particolare, quali caratteristiche li accomunano e le teorie che sono state sviluppate su di essi. Ci soffermeremo, poi, su un particolare social network, Twitter, spiegando nel dettaglio il suo funzionamento e il suo “*successo*” in Italia e nel mondo.

Successivamente introdurremo il concetto di *crowdsourcing*, concetto altamente legato ai social network, approfondendo particolarmente il suo utilizzo durante le quattro fasi della gestione delle emergenze.

Concluderemo la trattazione descrivendo una serie di casi emblematici, dove l'utilizzo dei social network ha avuto un ruolo predominante nella gestione delle emergenze.

### 2.2.1 I social network

I *social network* (in italiano *reti sociali*) sono definiti come strutture di individui connessi tra di loro mediante legami sociali. Tali legami possono essere di natura diversa, per esempio lavorativa, di amicizia, di interesse o familiare.

Lo studio dei social network è iniziato nel lontano 1900 grazie al filosofo e sociologo Simmel che ha fondato una prima teoria per l'interpretazione di fenomeni sociali [16]. Nel 1964 lo psicosociologo Moreno propose la prima

formale rappresentazione dei social network come combinazione di archi e nodi [13], rappresentazione che ripresero e sfruttarono, successivamente, Haray e Cartwright per la creazione di *sociogrammi* derivanti dalla teoria dei grafi [6]. Mediante l'introduzione di tale rappresentazione iniziarono ad essere teorizzati i complessi modelli sociali che stanno alla base dei social network.

Da qui si svilupparono due scuole di pensiero:

- *l'approccio sociocentrico*, focalizzato sull'identificazione di sottogruppi di persone appartenenti alla stessa rete sociale e sullo studio delle relazioni che avvengono tra di essi;
- *l'approccio egocentrico*, focalizzato invece sullo studio dell'intera comunità;

A quest'ultima scuola di pensiero apparteneva lo psicologo Milgram, che enfatizzò l'importanza dei social network come strumento di condivisione di informazioni e conoscenza. Inoltre Milgram introdusse il concetto di "*sei gradi di separazione*" che aveva l'obiettivo di dimostrare la "*teoria del piccolo mondo*" e cioè che, nonostante ogni elemento tenda ad avere relazioni prevalentemente con solo pochi altri, esso può comunque essere "*vicino*" a qualsiasi altro elemento della rete mediante pochi intermediari [12].

La diffusione del web e del termine *social network* ha fatto sì che si creassero delle ambiguità di significato. Infatti la rete sociale vera e propria è nata come rete sociale fisica mentre quelle web-based sono considerate forme di comunicazione in rete: tali piattaforme web permettono di organizzare le relazioni umane in una mappa informatica.

Secondo la definizione data dagli studiosi Boyd e Ellison [3] si possono definire i *social network sites* come servizi Web che permettono di:

- costruire un profilo pubblico o semi-pubblico all'interno di un sistema vincolato;
- articolare una lista di altri utenti con cui avere una connessione;
- vedere la lista dei propri contatti e ciò che fanno all'interno del sistema.

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

I *social network sites* che a livello mondiale e italiano sono più diffusi e hanno il maggior numero di utenti sono Facebook, Twitter e LinkedIn [10]. I servizi essenziali che questi social network mettono a disposizione sono:

- *creazione di un profilo utente*: in esso sono contenuti una serie di dati, alcuni inseriti dall'utente mentre altri creati automaticamente in base all'utilizzo della piattaforma. In base al tipo di social network site il profilo può essere reso pubblico, semi-pubblico o privati agli utenti della rete;
- *creazione di gruppi*: consente una condivisione mirata di conoscenza e contenuti;
- *ricerca contatti*: è possibile ricercare altri utenti dalle liste di contatti degli amici o da una lista pubblica.

Lo sviluppo sempre più massiccio degli smartphone ha modificato ulteriormente il panorama dei social network. Infatti una parte del traffico sui social network è generato attraverso dispositivi mobili: secondo uno studio condotto da BI Intelligence nel 2013, in media, la percentuale del traffico mobile rispetto a quello totale è pari al 40%, percentuale che sale al 50% per Facebook e al 70% per Twitter [9].

Più in generale, i *social network sites* rientrano nel gruppo più ampio dei *social media*, l'insieme di tutte le tecnologie e pratiche online usate dalle persone per condividere contenuti testuali, immagini, audio e video.

### 2.2.1.1 Twitter

Twitter è uno dei *social network sites* più diffusi a livello mondiale. E' un servizio online gratuito di *social networking* e *microblogging* e dà la possibilità agli utenti di pubblicare e leggere messaggi testuali di lunghezza massima di 140 caratteri, chiamati "*tweet*". Ogni utente ha un profilo personale aggiornabile mediante messaggi di stato che, una volta inseriti, divengono immediatamente disponibili su di esso e negli aggiornamenti di stato di tutti coloro che sono *follower* dell'utente. Infatti la relazione che intercorre tra gli utenti, cioè la relazione di *follower/following*, è di tipo monodirezionale e in particolare si ha che:

## Stato dell'arte

---

- un utente può vedere gli aggiornamenti di stato e i profili personali di tutti i *following* (utenti che egli “*segue*”). Potrà inoltre commentare uno specifico tweet, pubblicato da uno di loro, tramite la funzionalità di *mention*(@) o condividere lo stesso mediante la funzionalità di *retweet*(RT).
- il profilo personale e tutti gli aggiornamenti di stato dell'utente sono, invece, visibili solo dai suoi *follower* cioè da tutti gli utenti che hanno deciso di “*seguirlo*”.

Nonostante non esista in Twitter il concetto di “*conversazione*”, è possibile ugualmente raggruppare i tweet mediante l'utilizzo di un *hashtag*(#). Gli hashtag definiscono l'argomento della discussione e aiutano nella ricerca di tutti i tweet recenti ad esso riferiti. Tale ricerca può essere fatta anche mediante *keyword* e restituisce un insieme di messaggi pubblici contenenti tale keyword scritti da diversi utenti, indipendentemente dal fatto che chi ha eseguito la ricerca sia loro *follower*. Ciò dimostra il carattere giornalistico di Twitter e di come sia diventato una piattaforma di riferimento per la diffusione di news.

Twitter è costruito interamente su architettura Open Source e mette a disposizione oltre al *social network site* anche una *social network mobile app*, scaricabile gratuitamente dai più popolari store, e numerose API, che permettono a varie applicazioni di utilizzare al loro interno i suoi servizi <sup>1</sup>.

Twitter rappresenta un ottimo strumento per la diffusione delle informazioni, mediante la struttura basata su *follower*, che stimola la condivisione e genera il fenomeno sociale della *influence*: una minoranza di utenti, chiamati *influencers*, eccellono nel convincimento di tutti gli altri. Tale meccanismo non dipende, essenzialmente, dal numero di *follower* di un utente, ma dal valore intrinseco del messaggio che provoca i *retweet* e dalla stima percepita verso l'autore che stimola la funzione di *mention* (risposta/ citazione di un tweet). Questo porta alla teoria fondamentale chiamata “*million follower fallacy*”: non è il numero di *follower* a garantire l'influenzabilità di un utente, ma piuttosto la pertinenza delle cose che egli scrive [11].

---

<sup>1</sup><http://dev.twitter.com>



## 2.2 I social network e il loro ruolo nella gestione delle emergenze

### Twitter in Italia

A livello mondiale Twitter è il social network con la più veloce crescita di utenti attivi. Secondo una statistica<sup>2</sup> condotta nel Dicembre 2013, a livello mondiale dal 2010 ad oggi il numero di utenti Twitter è aumentato da 172 milioni a più di 500 milioni. Da 55 milioni di tweet al giorno si è passati a circa 500 milioni, con una media di 570 mila tweet al secondo nell'ultimo anno.

In Italia, nel 2010, erano presenti circa 1.3 milioni di utenti registrati, 350 mila dei quali effettivamente attivi. Ad oggi siamo arrivati a più di 9.5 milioni, di cui ben 3.5 attivi. Di questi ultimi, circa 2.2 milioni, il 64% quindi, è solito connettersi via mobile.

In uno studio meno recente, invece, si è osservata la natura dei tweet pubblicati: durante i primi sei mesi del 2012 sono stati lanciati in rete in Italia 60 milioni di tweet, di cui il 26% contiene un link, il 19% è un retweet, il 26% ha un hashtag e solo il 5% è geolocalizzato. Facendo, inoltre, una più approfondita analisi è stato mostrato come Twitter venga utilizzato soprattutto per scopi seri come la testimonianza e la condivisione di fatti di cronaca: un esempio significativo è dato dai tweet riferiti alla bomba di Brindisi e al terremoto che ha colpito l'Emilia Romagna. A dimostrazione di ciò sono significative le parole dette dal direttore de "Il Post", un quotidiano online italiano: "*Noi al Post sappiamo le cose che accadono nel mondo prima da Twitter e poi dalle agenzie. E ciò non riguarda solo il tempismo ma anche l'affidabilità*" [2].

### 2.2.2 Il Crowdsourcing

Secondo la definizione di Estellés e González [8] :

*"Il crowdsourcing è una tipologia di attività online partecipativa nella quale una persona, istituzione, organizzazione, non a scopo di lucro o azienda propone ad un gruppo di individui, mediante un annuncio aperto e flessibile, la realizzazione libera e volontaria di un compito specifico. La realizzazione di tale compito, di complessità e modularità variabile, e nella quale il gruppo di riferimento deve partecipare apportando lavoro, denaro, conoscenze e/o esperienza, implica sempre un beneficio per ambo le parti. L'utente otterrà, in*

---

<sup>2</sup><http://wearesocial.it/blog/2013/12/latest-3-years-twitter-34-milioni-di-utenti-attivi-italia/>

## Stato dell'arte

---

*cambio della sua partecipazione, il soddisfacimento di una concreta necessità, sia essa economica, di riconoscimento sociale, di autostima, o di sviluppo di capacità personali; il crowdsourcer, d'altro canto, otterrà e utilizzerà a proprio beneficio il contributo offerto dall'utente, la cui forma dipenderà dal tipo di attività realizzata".*

Quindi il crowdsourcing può essere visto come un modello di produzione e risoluzione di problemi dove la soluzione viene richiesta ad un gruppo non definito di persone, la “*crowd*”, che normalmente si riuniscono in comunità online. Spesso gli utenti che hanno contribuito a trovare la soluzione vengono ricompensati in denaro o mediante premi e riconoscimenti. Grazie allo sviluppo di Internet e del Web, il crowdsourcing oggi viene mutuato solo attraverso questi canali.

### 2.2.2.1 Il Crowdsourcing e le fasi dell'emergenza

L'informazione sociale garantita attraverso il crowdsourcing può essere utilissima nella gestione delle emergenze. Queste informazioni possono essere di supporto in tutto il ciclo delle emergenze e possono essere di natura diversa. Vediamo, quindi, ogni fase da che tipo di informazione può essere caratterizzata.

#### **Preparedness**

In questa fase i cittadini possono fornire informazioni relative allo stato di rischio del territorio e identificando le infrastrutture critiche. Le segnalazioni normalmente raggiungono le autorità sotto forma di lamentela ed è necessario che esse vengano sfruttate al fine di elaborare un migliore piano di emergenza e in particolare di effettuare una migliore analisi dei rischi. Da parte delle autorità è molto utile in questa fase l'attivazione del *crowdfeeding*, cioè l'attivazione di una serie di strumenti per istruire i cittadini su cosa fare in caso di emergenza.

#### **Response**

Essendo la fase più critica le informazioni che giungono dai cittadini sono molto significative e devono essere utilizzate a supporto delle decisioni strategiche nel

## **2.2 I social network e il loro ruolo nella gestione delle emergenze**

momento di risoluzione dell'emergenza. Queste informazioni possono contenere una descrizione della situazione generale del territorio, dati relativi allo stato dei cittadini colpiti oppure richieste di intervento.

### **Recovery**

In questa fase i cittadini possono informare le autorità relativamente a quali siano gli interventi necessari per la ricostruzione post emergenza. Inoltre il crowdsourcing può essere utilizzato dalle autorità per chiedere l'aiuto di esperti esterni per le problematiche di ricostruzione.

### **Mitigation**

Durante questa lunga fase, il crowdsourcing può essere utile per raccogliere lamentele e problematiche sociali derivanti dall'evento, dati che devono essere sfruttati nella successiva fase di preparazione.

### **2.2.3 I social network come possibile aiuto nella gestione delle emergenze**

L'utilizzo dei social nel campo dell'analisi dei rischi è, come già accennato, ancora in fase sperimentale, agli albori. Di seguito presentiamo alcune riflessioni ed iniziative su tale tema, in ambito sia internazionale che italiano, per far capire quanto il miglioramento delle tecniche, in questo senso, sia desiderabile.

#### **2.2.3.1 I Social Media e l'Emergency Management secondo l'OECD**

L'Organization for Economic Co-operation and Development (OECD) è un'organizzazione internazionale di studi economici per i paesi membri dell'UE, paesi sviluppati aventi in comune un sistema di governo di tipo democratico ed un'economia di mercato. L'organizzazione svolge prevalentemente un ruolo di assemblea consultiva che consente un'occasione di confronto delle esperienze politiche, per la risoluzione dei problemi comuni, l'identificazione di pratiche commerciali ed il coordinamento delle politiche locali ed internazionali dei paesi membri.

Un nuovo paper dell'OECD del mese di Gennaio 2014 raccoglie riflessioni sull'uso che si può fare dei social media nella comunicazione per "*Risk and Crisis Management*". Molto interessante la prima frase delle conclusioni: "*l'uso dei social media per questi scopi è ancora agli albori*".

Il report "*The Use of Social Media in Risk and Crisis Communication*" contiene i punti salienti di un workshop organizzato congiuntamente dall'OECD e dall'International Risk Governance Council. Approfondisce l'utilizzo che si potrebbe fare dei social media per le comunicazioni riguardanti la gestione dei rischi e delle condizioni di emergenza. Il concetto che viene affermato in modo perentorio è la necessità di far uso dei social media nella comunicazione per la gestione del rischio e dell'emergenza. Anche il silenzio parla e la scelta di non comunicare attraverso facebook o twitter si è rivelata un punto di debolezza per la credibilità e l'efficacia delle istituzioni. Questo poiché la pervasività dei social media e la "*connessione permanente*" dei singoli alla moltitudine e alle istituzioni è motivo di rassicurazione, di creazione e consolidamento della fiducia reciproca e contribuisce a determinare la reputazione dell'istituzione (come anche del blogger o dell'esperto), qualità che, per la gestione della comunicazione su temi critici, è parte integrante del messaggio. È la chiave di lettura della componente emotiva della comunicazione in emergenza, i social media possono dare qualcosa in più. Possono contribuire alla movimentazione dei volontari, a veicolare la comunicazione di early warning (principio fondante dell'emergency management completamente ignorato in Italia), ma anche attivare l'early detection capillare e, di conseguenza, trasformare il ruolo passivo dei cittadini in comportamento attivo. Per fare questo salto di qualità è necessario superare il modello top-down e articolare adeguatamente la comunicazione, sia rispetto all'architettura istituzionale che rispetto agli elementi sociali (dall'amministrazione centrale a quella locale, dalla persona alle associazioni, dal comune cittadino all'esperto).

Naturalmente i social media non sono tutti uguali e non tutti rispondono alle esigenze necessarie. I social network come twitter e facebook sono strumenti di comunicazione realtime e non sarebbero in grado, per esempio, di mappare da soli le informazioni geografiche della crisi. Questa è la finalità di strumenti classificati come volunteer technology communities (come Ushahidi o Sahana). Così come dagli sharing media (Youtube, Flickr, Wiki, blog) non

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

bisogna aspettarsi molto di più della condivisione e diffusione della conoscenza. Però tutti sono caratterizzati da una qualità essenziale: l'informazione che gestiscono è qualificata in termini sia geografici che temporali.

Quest'uso dei social media richiede una legislazione rinnovata per la giusta tutela della privacy e delle informazioni personali. L'utilizzo estensivo delle informazioni create nei social media per via del loro uso nella comunicazione in emergenza può essere preziosissimo. Questi dati dovrebbero essere e rimanere "open", un requisito che sarebbe opportuno non sottovalutare se si vuole evitare di amplificare altre forme di rischio già presenti nella società dell'informazione. Effettivamente non c'è altro modo per concludere: l'uso dei social media per questi scopi è ancora agli albori.

### 2.2.3.2 I Social Network al Festival del Volontariato 2014

Una delle attività promosse dal Festival del volontariato 2014<sup>3</sup>, in programma a Lucca dal 10 al 13 aprile, sarà quella di raccogliere esperienze e buone pratiche di utilizzo dei social media per una comunicazione del rischio e in situazioni di emergenza di protezione civile. E' infatti on-line un apposito modulo da compilare<sup>4</sup>, realizzato dal CNV (Centro nazionale per il Volontariato) in collaborazione con il Dipartimento nazionale della Protezione civile, attraverso il quale non solo le organizzazioni di volontariato e di protezione civile, ma anche tutte le numerose componenti di questo variegato sistema (a partire dalle amministrazioni locali e altri enti pubblici) possono raccontare le proprie esperienze nello sviluppo e nell'utilizzo delle varie applicazioni e dei social media.

Obiettivo di questa iniziativa è raccogliere storie ed informazioni che permettano di proseguire nel percorso di confronto sull'utilizzo dei social (e non solo) come strumento di rapporto diretto con i cittadini per i temi di protezione civile e sul loro impatto nelle dinamiche di comunicazione nell'ambito di una possibile gestione dell'emergenza. Un percorso la cui prima tappa è stata la giornata di studio "La protezione civile e i social media: comunicare il rischio e il rischio di comunicare"<sup>5</sup>, organizzata presso il Dipartimento della protezione civile lo scorso novembre, che ha stimolato un vivace dibattito con la partecipa-

<sup>3</sup><http://www.festivalvolontariato.it/>

<sup>4</sup><http://www.festivalvolontariato.it/socialprociv/>

<sup>5</sup><http://www.ilgiornaledellaprotezionecivile.it/index.html?pg=1&idart=10793&idcat=1>

zione delle varie realtà del sistema, oltre a studiosi, giornalisti ed esperti della comunicazione 2.0. L'iniziativa prende il via dal continuo confronto che si sta sviluppando, anche all'interno del sistema di Protezione civile, sull'evoluzione che i social network stanno avendo come nuovi mass-media e sulla velocità nello scambio di comunicazioni polidirezionali che essi offrono. L'obiettivo è mappare e verificare l'esistenza di procedure che possano garantire, pur in un contesto caratterizzato da una continua evoluzione, l'attendibilità e la validazione delle informazioni messe a disposizione dei cittadini, soprattutto in situazioni di crisi.

### **2.2.4 Casi reali di utilizzo dei social network durante le emergenze**

In questa sezione si mostreranno una serie di esempi significativi che possano far capire come è possibile l'utilizzo dei social network all'interno di scenari reali di emergenza, con lo scopo di fornire una testimonianza concreta di come essi abbiano avuto un ruolo importante se non addirittura fondamentale nella loro gestione e risoluzione.

Spiegheremo come le reti sociali possano rappresentare una nuova strategia di comunicazione e come possano offrire una più efficiente distribuzione delle informazioni rispetto ai sistemi tradizionali che si affidano al mezzo stampa.

Analizzeremo eventi passati nei quali i social network hanno avuto un ruolo predominante nella diffusione di informazioni da parte di autorità e hanno permesso alle persone in difficoltà di richiedere assistenza o rimanere semplicemente informate sul quadro dell'emergenza.

Infine descriveremo alcuni sistemi adottati da enti e organizzatori, estere ed italiane, per estrarre informazioni sociali dai messaggi presenti nei social network per costruire dati aggregati utili per la prevenzione e la risoluzione di eventi disastrosi.

#### **2.2.4.1 Una nuova strategia di comunicazione: il caso NORAD NORTHCOM**

NORAD (North American Aerospace Defense Command) e NORTHCOM (U.S. Northern Command) sono le principali agenzie militari americane per la difesa e adottano strategie intelligenti per lo sfruttamento dei social media.

## **2.2 I social network e il loro ruolo nella gestione delle emergenze**

Nel piano di comunicazione proposto nel 2010, ai canali tradizionali (agenzie di stampa), sono stati associati account sui principali social network (Twitter, Facebook, Youtube e Flickr) e un blog gestito dal comandante per informazioni importanti e tempestive.

Il meccanismo di diffusione delle informazioni sfrutta le ricondivisioni assumendo una modalità esponenziale: il comando invia i messaggi ad una rete di dipendenti interni, i quali devono ricondividere tale messaggio con la propria rete privata di contatti. Ovviamente anche quest'ultimi eseguiranno lo stesso tipo di operazione e di conseguenza l'informazione tenderà ad espandersi in maniera esponenziale e raggiungendo un'alta percentuale di cittadini.

E' possibile pensare che le informazioni più importanti, cioè quelle relative a situazioni di emergenza, siano quelle ad essere spontaneamente più ricondivise e ad avere maggior "*successo sociale*".

### **2.2.4.2 Uragano Gustav in Louisiana (Agosto 2008)**

L'uragano Gustav fu il settimo ciclone tropicale e il secondo uragano maggiore nella stagione degli uragani atlantici del 2008. Memori della tragedia avvenuta anni prima, il governo decise di far evacuare l'intero stato per evitare vittime. Al fine di coordinare meglio gli spostamenti, il Dipartimento di Sicurezza Americano chiese aiuto al social network MySpace, allora molto popolare, richiedendo lo sviluppo di un *widget*, un'applicazione predefinita, che aveva l'obiettivo di aiutare nella ricerca delle persone scomparse con un sistema di geolocalizzazione, donare fondi e permettere di seguire l'avanzamento in tempo reale dell'uragano verso la costa. L'utilizzo di tale widget fu di molto aiuto nella fase di evacuazione e permise una riduzione del traffico telefonico d'emergenza.

I social network e soprattutto Facebook, risultarono utili anche per gli Stati chiamati ad accogliere gli sfollati della Louisiana per contattare i volontari al fine di rispondere in modo celere alla chiamata di emergenza.

### **2.2.4.3 Eruzione del vulcano islandese Eyjafjöll (Aprile 2010)**

Il 20 Marzo 2010 iniziò una serie di eventi di natura vulcanica sul vulcano Eyjafjöll in Islanda, che sfociò nell'eruzione del 14 Aprile 2010. Essa provocò

seri problemi per il traffico aereo fino al 23 Aprile, con conseguenti chiusure degli aeroporti del nord Europa.

Si rese necessaria una strategia di comunicazione nei confronti dei turisti e dei lavoratori bloccati a terra. I social network divennero il mezzo principale per tale comunicazione: le principali compagnie aeree si unirono alle conversazioni dei maggiori social network per raggiungere più facilmente i propri clienti e fornire un servizio di assistenza funzionale. In particolare su Twitter, alcuni hashtag si diffusero molto velocemente raggiungendo oltre 55000 mention. Anche la piattaforma Flickr contribuì alla diffusione delle informazioni tramite immagini scattate nei giorni dell'emergenza.

I danni registrati dalle compagnie aeree nei primi sette giorni dell'emergenza ammontarono a 1.7 miliardi di dollari, cifra che sarebbe potuta essere maggiore in assenza di una strategia di comunicazione adeguata. Ciò testimonia l'importanza dei social network nella diffusione delle informazioni utili.

### **2.2.4.4 Giappone: terremoto, tsunami, e disastro nucleare (Marzo 2011)**

Nel Marzo del 2011, un sisma di magnitudo 9 e con epicentro in mare a 30 km di profondità, colpì la costa settentrionale del Giappone. Si generò in seguito uno tsunami con onde di oltre 10 metri di altezza e con una velocità stimata di 750 km/h.

Il sisma provocò anche lo spegnimento di 11 centrali nucleari tramite un sistema di emergenza automatico. Alcuni reattori della centrale di Fukushima comunque esplosero, a causa del cattivo funzionamento dei meccanismi di raffreddamento dovuto al danneggiamento procurato dal sisma. Come conseguenza di tali eventi, venne imposta l'evacuazione per più di 110 mila persone intorno a Fukushima, oltre ad un alto numero di persone che dovettero lasciare le proprie case in seguito ai danni subiti per il sisma ed il maremoto.

La comunicazione attraverso le reti telefoniche fu resa praticamente inaccessibile dalle migliaia di richieste. La rete Internet, però, risultò funzionante, dando la possibilità alle persone di comunicare sui social network ed usufruire delle informazioni di supporto diffuse dagli account delle principali città ai propri cittadini.



## **2.2 I social network e il loro ruolo nella gestione delle emergenze**

La rete venne sfruttata anche per la raccolta di fondi e il lancio di iniziative benefiche a favore degli sfollati. Inoltre il web developer Yusuke Wada creò un database per raccogliere informazioni riguardo agli scomparsi provenienti da Twitter e reclutò, sempre tramite i social network, 200 volontari per aiutarlo nella gestione del sistema.

### **2.2.4.5 Terremoto dell'Emilia (Maggio 2012)**

Un primo esempio locale si ha il 20 Maggio 2012 quando un forte terremoto di magnitudo 5.9 con epicentro a Finale Emilia (MO), colpì l'Emilia Romagna provocando 7 morti, circa 50 feriti, 5000 sfollati e ingenti danni stimati a 13 miliardi di euro.

Per l'evento, venne istituito un sito internet coordinato da persone comuni, raccogliendo oltre 610.000 euro di donazioni e divenendo il canale ufficiale di contatto tra istituzioni e cittadini. Alcuni canali informativi non ufficiali, però, aprirono la strada alla diffusione di informazioni piuttosto generiche e a volte anche false. Per questo il sindaco di San Felice sul Panaro richiese la certificazione dei post da parte delle istituzioni al fine di non creare falsi allarmismi.

I social network risultarono utili per l'invio di richieste di aiuto e di fondi. Inoltre vennero creati alcuni gruppo all'interno del social network Facebook con lo scopo di raccogliere riferimenti di persone o famiglie bisognose di un posto letto.

### **2.2.4.6 Uragano Sandy**

L'uragano Sandy è stato il decimo uragano atlantico del 2012 e ha colpito molti stati americani tra cui la Giamaica, Cuba, le Bahamas e la costa orientale degli Stati Uniti. Secondo le più recenti stime, l'uragano Sandy è stato classificato come il secondo più costoso uragano atlantico dopo Katrina, a causa dei numerosi danni, interruzioni di servizio e perdite umane che ha provocato. Sandy si è sviluppato da un'onda tropicale nel Mar dei Caraibi il 22 ottobre 2012 e ha creato danni e distruzione in tutta la costa orientale dell'America fino al 29 ottobre, giorno dell'arrivo a New York dell'uragano.

## Stato dell'arte

---

Proprio il giorno del 29 ottobre, mentre a New York le linee telefoniche erano in tilt, Manhattan si trovava al buio e molti siti internet crashavano, il solo strumento che non ha mai smesso di rilevarsi utile è stato il social network Twitter: difatti è stato utilizzato come vera e propria linea di soccorso - o 911- e ha permesso la divulgazione di informazioni istituzionali, foto in tempo reale dei luoghi allagati, rassicurazioni ma soprattutto richieste di aiuto.

Già nei giorni precedenti, molte agenzie governative e organi ufficiali, come il governatore di New York Andrew Cuomo, l'Ente federale per la gestione delle emergenze e NotifyNYC, account ufficiale per la gestione delle emergenze della città di New York, avevano twittato aggiornamenti e ordini di evacuazione.

Il giorno dell'emergenza l'account dei Vigili del Fuoco e quello della Croce Rossa Americana sono stati letteralmente presi d'assalto, con richieste di qualunque tipo ma di cui la maggior parte erano richieste di aiuto. Grazie a questi dati la Croce Rossa Americana ha tracciato una mappa delle emergenze nella metropoli, individuando le aree in cui era necessario concentrare i soccorsi.

Per l'occasione, Twitter ha messo a disposizione ai newyorkesi strumenti che tradizionalmente sono riservati a inserzionisti pubblicitari: per la prima volta, l'azienda ha creato una pagina evento con hashtag "#Sandy" in modo da raggruppare tutti i tweet relativi all'uragano e permettere ai suoi utenti di conoscere in tempo reale gli ultimi aggiornamenti sul da farsi. E' stata data, inoltre, la possibilità alle istituzioni di pubblicare tweet promossi, cioè tweet solitamente riservati ad inserzionisti e usati per raggiungere un pubblico più vasto possibile, e una serie di servizi basati sul codice postale dell'utente, per raggiungere un pubblico residente in una determinata area.

Twitter ha fatto sì che venisse utilizzato anche Instagram, il sito di photo sharing che ha raccolto migliaia di scatti riguardanti i disastri causati da Sandy. Purtroppo questo ha fatto anche sì che venissero postati scatti clamorosamente falsi, come immagini di squali che nuotavano nei cortili del New Jersey o foto di un sub all'interno della metropolitana allagata. Ma nonostante il fatto che sul social network non fosse stato impostato nessun tipo di censura o filtro, gli account più prolifici di fake sono stati immediatamente individuati e ogni cinguettio fasullo è stato in breve tempo smascherato. Insomma, Twitter, come

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

ha riferito il giorno dopo la tempesta un esperto di social media su BuzzFeed, è una grande “*fact-processing machine*”, una macchina che processa velocemente informazioni e che tanto velocemente si libera anche di quelle sbagliate [15].

### 2.2.4.7 Alluvione in Sardegna (Novembre 2013)

Nel mese di Novembre 2013 la tempesta Cleopatra ha investito la Sardegna con una mole d’acqua mai vista prima e questa terra, un tempo abituata alla siccità più che ai diluvi, non era pronta. Non lo erano le famiglie sorprese dalla furia dell’acqua nei seminterrati di Olbia, non lo erano gli sfollati di Terralba che hanno dormito in chiese e palestre, non lo erano gli abitanti di Bitti e Arzachena dove l’acqua ha devastato case e locali commerciali.

Il Capo della Protezione Civile Gabrielle disse che in Sardegna aveva trovato più buona volontà che organizzazione. Lo stato d’emergenza era stato dichiarato per tempo, ma la macchina statale spesso è troppo lenta nel mettersi in moto. E allora gli abitanti della Sardegna pensarono di darsi da fare per aiutare al più presto chi era rimasto senza casa e non aveva più un posto dove ripararsi. Con questo scopo nacque la pagina Facebook “*Apriamo le case ai nostri concittadini*”, in cui chi aveva un alloggio libero poteva dare la propria disponibilità per ospitare chi aveva bisogno.

“*A causa dell’odierna alluvione che ha colpito la nostra isola, in particolare Olbia, invito tutti i miei contatti ad aprire la loro casa per ospitare chi ha difficoltà ad arrivare alla propria abitazione. Aiutiamoci tra di noi!*”, fu l’invito iniziale. Seguirono annunci coi numeri di telefono di chi offriva ospitalità in un B&B, una villetta al mare o semplicemente nelle camere libere della propria abitazione. “*A chiunque avesse parenti, amici, conoscenti, le cui case son state sommerse dall’acqua metto a disposizione il mio b&b . senza problemi contattatemi*”, scrisse Jobee subito subissato da “*Mi piace*” e da commenti di chi lo incoraggiava o lo ringraziava per l’offerta. Eleonora scrisse che “*La nostra casa è aperta sia a pranzo che a cena...un pasto caldo e tanto affetto è quello che possiamo offrirvi*”, e ci furono donne che si offrirono di badare ai bimbi delle famiglie colpite dalla tempesta.

### 2.2.4.8 American Red Cross Digital Operations Center

La Croce Rossa Americana creò, nel 2012, il primo centro operativo digitale per la gestione delle emergenze mediante il monitoraggio costante e intelligente dei social network. All'interno del centro, i monitor forniscono costantemente dati aggregati ricavati dalle informazioni provenienti dai principali social network, utilizzati per prevenire e gestire al meglio, in ogni fase, le emergenze. Al centro operativo è sempre presente una persona fisica, anche se i dati sono disponibili attraverso il Web e i volontari, in caso di emergenza, hanno accesso ad un *front-end* telematico mediante il quale possono informare gli altri sulle situazioni di pericolo o ricevere istruzioni su come muoversi condividendo la propria identità e posizione. In questo modo è possibile richiamare i volontari anche fuori dall'orario di servizio ufficiale e far sì che il sistema di segnalazione risulti tempestivo.

I dati registrati non costituiscono, di per sè, motivo per agire, ma rappresentano semplicemente un supporto informativo per le decisioni strategiche. In caso di disastri di grandi dimensioni, viene usato il *Disaster Operations Center*, un centro operativo che permette, oltre alla funzionalità di aggregazione intelligente dei dati, l'invio di informazioni *push* a tutti i cittadini che hanno l'*app* della Croce Rossa per smartphone o sono iscritti ai suoi account nei diversi social network. Il *Disaster Operations Center* permette, inoltre, di inviare report continui alle centrali operative e comunica direttamente con gli operatori mediante specifiche console.

I dati aggregati vengono forniti agli operatori attraverso diversi formati di visualizzazione:

- *heat map*: illustra il volume delle conversazioni sulla mappa, permettendo l'identificazione dei punti caldi;
- *community*: illustra i profili delle persone che parlano di un certo topic e quali sono gli utenti con più follower;
- *universe*: illustra i volumi di conversazione attuali riguardanti topic specifici, permettendo di effettuare sentiment analysis su di essi;

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

- *conversation dashboard*: illustra i volumi di traffico lungo un arco temporale, evidenziando quando dei specifici topic smettono di generare traffico;

I dati così visualizzati, permettono di costruire *topic profiles*, cioè profili di conversazione relativi ad uno specifico avvenimento e sono molto importanti, in quanto rappresentano un insieme di informazioni mirate.

Esiste inoltre un applicazione mobile della Croce Rossa Americana, sviluppata su piattaforma iOS, utilizzabile come supporto di carattere puramente informativo per il cittadino. In essa sono raccolti tutti i punti di ricovero (gli *shelters*) che possono essere raggiunti e ai quali è possibile chiedere accoglienza in caso di eventi critici. Tali punti di sono visualizzati all'interno di una mappa, con relative informazioni specifiche sul ricovero.

Inoltre è anche possibile visualizzare una serie di news, attraverso un collegamento diretto al sito della Croce Rossa, con l'obiettivo di fornire al cittadino informazioni dettagliate e precise.

### **2.2.4.9 Emergency Situation Awareness - Automated Web Text Mining (ESA-AWTM)**

ESA-AWTM è un sistema di *data mining*, creato per dare un supporto decisionale ai centri di coordinamento crisi presenti in ogni centro urbano dell'Australia. I dati vengono prelevati dal social network Twitter e dopo esser stati collezionati, revisionati, aggregati ed organizzati, vengono analizzati al fine di trovare dati specifici riguardo ad eventi che possono impattare sulla sicurezza dei cittadini.

Il problema principale nella costruzione del sistema è convertire un grandissimo flusso di messaggi in informazioni utili, il tutto in maniera veloce, pratica ed efficiente. Le funzioni, quindi, implementate includono:

- *ricerca di incidenti*: il sistema esamina le parole che sono contenute nei tweet e usa dati storici per costruire un modello statistico per la ripetizione di termini;

## Stato dell'arte

---

- *sintesi e indicizzazione dei messaggi*: creazione mediante motori esterni di cluster di tweet che parlano di un medesimo argomento;
- *classificazione e la revisione di tweet importanti*: sviluppo di un sistema statistico che permetta la classificazione di tweet in base all'interesse che viene determinato da fattori quali la ricondivisione o la presenza di keyword ad alto significato semantico;
- *identificazione, tracciamento e gestione dei problemi*: algoritmo che si occupa di unire, dividere o cancellare cluster con contenuto simile lungo medesimi archi temporali;
- *analisi forense*: allarmi memorizzati in un database storico per analisi future.

L'output di ESA-AWTM permette la creazione di una mappa al fine di posizionare tutti i tweet importanti e indicare quali sono le zone di possibili incidenti. Vengono indicati anche altri dati statistici come il volume dei tweet catturati.

Tramite questo sistema, durante il periodo di prova durato nove mesi, sono state meglio gestite dodici importanti situazioni di emergenza in diverse città australiane [4].

### 2.2.4.10 Una piattaforma di crowdsourcing: Ushaidi

Ushahidi Inc. è una azienda no-profit con sede in Kenya che sviluppa software open source gratuito per la raccolta, visualizzazione e geocalizzazione interattiva di informazioni. L'azienda mira a sfruttare il paradigma del crowdsourcing rivolto all'attivismo sociale e alla responsabilità pubblica. Offre, quindi, la possibilità ai cittadini di inviare informazioni utili mediante smartphone o rete internet al fine di creare archivi temporali e geospaziali di eventi di interesse pubblico.

Il concetto alla base del sito creato da Ushaidi<sup>6</sup> è quello di rendere visibile agli utenti mappe di dominio pubblico, con possibilità di segnalazioni da parte di chiunque, al fine di raccogliere dati mirati in riferimento ad uno specifico evento, scelto da colui che genera la mappa. I dati raccolti possono

---

<sup>6</sup><http://www.ushahidi.com>

## 2.2 I social network e il loro ruolo nella gestione delle emergenze

essere utilizzati sia dai cittadini, per identificare le zone di pericolosità, sia dalle istituzioni per meglio risolvere le emergenze e a supporto delle decisioni critiche.

Negli anni il software è stato utilizzato in varie situazioni che necessitavano attenzione, nei più disparati settori e in varie parti del Mondo. Alcuni esempi sono il terremoto di Haiti nel 2010, durante le proteste democratiche del Medio Oriente nel 2011 o le elezioni parlamentari in Armenia nel 2012. Nel corso degli anni il programma è stato più volte aggiornato e modificato, anche se l'ultima versione risale ad Aprile 2012. Ushaidi ha ricevuto veri premi, l'ultimo dei quali nel 2013, quando gli è stato conferito il MacArthur Award per la sua effettiva efficienza e la sua innovatività nell'aiuto per la risoluzione di problemi di interesse globale.

### **Utilizzo di Ushaidi durante il terremoto di Haiti (Gennaio 2010)**

Durante la situazione di emergenza creata in seguito al gravissimo terremoto di Haiti del 2010, venne utilizzata la piattaforma di crowdsourcing Ushaidi, mediante la partecipazione di numerosi volontari e organizzazioni umanitarie al Ushaidi Haiti Project (UHP).

UHP permise la diffusione e raccolta di informazioni relative ai luoghi colpiti, facilitando gli spostamenti attraverso la generazione di mappe stradali sulle quali furono identificati i percorsi sicuri e le zone praticabili.

La piattaforma venne inoltre utilizzata dal Dipartimento Statale Americano degli Analisti e dall'esercito per conoscere nello specifico la situazione locale e organizzare al meglio le operazioni di soccorso. L'importanza del suo utilizzo nasce dal fatto che è l'unico e il più completo aggregatore di informazioni riguardanti le aree colpite, in un territorio dove i supporti informatici efficienti sono piuttosto carenti. In più numerose interviste a responsabili ufficiali hanno testimoniato il reale apporto di Ushaidi nel salvataggio di numerose vite umane.

Purtroppo non mancarono i problemi di utilizzo della piattaforma: in alcune zone la connessione internet era totalmente assente o troppo lenta ed inoltre l'aggregazione dei dati fornita da Ushaidi non corrispondeva a quella richiesta dalle autorità per prendere concrete decisioni e pertanto numerose informazioni non poterono essere sfruttate per tempo.

### 2.2.4.11 Emergenza24

Si tratta di un esperimento di Social Network for Emergency Management, una sorta di Social Network per le emergenze realizzato da un gruppo di persone provenienti da diversi settori, quali ad esempio ingegneria, geologia, ma anche protezione civile e GIS<sup>7</sup>. Non si tratta di un'associazione vera e propria ed è totalmente no profit e aperta a tutti. Chi collabora sono tutti volontari con l'obiettivo di cercare di dare una mano alle autorità responsabili durante le emergenze.

Il funzionamento è semplice: agli utenti che sono iscritti al progetto, e quindi al corrente del regolamento del sito, è richiesto, in caso di emergenza, di fornire informazioni dettagliate utilizzando su Twitter dei messaggi contenenti l'hashtag #Emergenza24 e una struttura fissa in modo da permettere di filtrare facilmente questo tipo di messaggi e raccogliere informazioni il più strutturate e precise possibili.

Nel sito, in caso di emergenza, viene aperta una pagina apposita in cui sono specificati eventuali numeri verdi e per le ricerche dei dispersi, come anche informazioni sulla situazione ricavate, appunto, dall'analisi dei messaggi filtrati dal sistema.

L'idea di base è molto buona poiché prevede l'integrazione con un Social Network (Twitter) e la partecipazione attiva dei cittadini che possono ricevere e, soprattutto, inviare segnalazioni. Sotto questi aspetti, il nostro progetto TORCIA, di cui parleremo approfonditamente nei capitoli a seguire, vi somiglia molto. Ma le funzionalità del sito di Emergenza24 sono molto più limitate di quelle del nostro sistema, in quanto i tweet che riescono a raccogliere ed elaborare sono solo quelli contenenti l'hashtag sopra citato e strutturati secondo dei precisi canoni. In altre parole non hanno alcun modo di raccogliere ed analizzare i tweet di utenti che non sono iscritti al loro progetto e, quindi, non hanno idea di come funzionino le modalità di segnalazione dell'emergenza. Infine non sono presenti nel sito alcune funzionalità che l'app mobile TORCIA si propone di realizzare, come fornire le mappe delle zone in cui è in corso l'emergenza, con eventuali percorsi di fuga e luoghi di raccolta.

---

<sup>7</sup><http://www.emergenza24.org/>



## **2.3 Analisi conclusiva**

La gestione delle emergenze è un processo lungo e complesso, che comporta l'interazione tra enti e soggetti che hanno funzioni e peculiarità differenti tra loro.

Se ci concentriamo soprattutto sulla fase di *response*, possiamo notare come il piano di emergenza, spesso, non sia un sufficiente supporto decisionale per gli enti e le organizzazioni che devono agire. Nel momento stesso dell'emergenza, possono occorrere una serie di situazioni che non sono state previste in fase di stesura del piano, ma che devono essere gestite in maniera real-time. A tal fine è necessario avere una base di informazioni ricca e, soprattutto, affidabile. Questa base può essere creata sfruttando le conoscenze di tutti quei cittadini che sono in primo piano testimoni delle emergenze. Difatti, la fase di *response* è quella in cui il crowdsourcing dà l'apporto più significativo, proprio perchè, per definizione, è la fase più critica.

Ci sono stati una serie di casi emblematici dove l'utilizzo dei social network ha avuto un ruolo fondamentale nella gestione delle emergenze. Tra tali casi quello più significativo è quello relativo all'American Red Cross Digital Operations Center. Questo centro operativo sfrutta l'informazione sociale proveniente dai social network come supporto informativo per le decisioni strategiche, basandosi sulla teoria da noi esposta precedentemente. Inoltre mediante la *app*, creata appositamente per i dispositivi mobili, gli operatori possono inviare informazioni push a tutti i cittadini al fine di mantenere aggiornata la popolazione sullo stato dell'emergenza.

In questo ultimo particolare però, risiede il limite di tale centro operativo: non esiste un sistema integrato che permetta una diretta interazione tra cittadini e autorità; i flussi informativi presenti sono tutti di tipo unidirezionale. Questo tipo di interazione è di vitale importanza e la dimostrazione l'abbiamo vista nel caso emblematico dell'uragano Sandy. Durante questo tragico evento, il social network Twitter è stato un punto di incontro sia per i cittadini, che si scambiavano informazioni su ciò che stava succedendo, sia per le autorità che hanno potuto avere informazioni in tempo reale sulle situazioni di pericolo. Inoltre, essendo basato su crowdsourcing, Twitter è riuscito a identificare autonomamente le informazioni che erano errate, facendo sì che il supporto

informativo creato per le autorità fosse completamente affidabile.

Altro limite che è possibile identificare dall'analisi dei casi emblematici precedentemente esposti è la totale mancanza di un servizio di alerting che riduca sensibilmente i tempi di reazione. La presenza di un sistema di questo tipo garantirebbe la segnalazione immediata di un particolare evento critico alle autorità, con conseguente rapidità nella consultazione dei contenuti e nel loro utilizzo in ambito decisionale.

Il progetto TORCIA, in cui si inserisce il presente lavoro di tesi, parte proprio dall'analisi dei dati relativi all'informazione sociale e dai limiti esposti precedentemente. Ci si propone quindi di costruire una metodologia e un'architettura che sia in grado di offrire ai diversi attori un sistema integrato capace, da un lato, di creare una base di informazioni affidabile a supporto decisionale e un meccanismo di alerting per le autorità; dall'altro, di fornire sia ai cittadini che alle autorità stesse un mezzo per la loro diretta interazione in tutte le fasi del ciclo di emergenza ma soprattutto in fase di *Response*. Il nostro sistema si baserà interamente su crowdsourcing, cercando di mettere in atto tutte le dinamiche che lo caratterizzano e prendendo spunto da piattaforme di successo come l'American Red Cross Digital Operations Center, l'Emergency Situation Awareness - Automated Web Text Mining e Ushaidi.

# Capitolo 3

## Il Progetto TORCIA

Il progetto TORCIA - PiaTtafORma di gestione CollaboratIvA delle emergenze - è un progetto finanziato da Regione Lombardia e ha come obiettivo quello di utilizzare tecnologie del Web 2.0 a supporto dei processi di pianificazione e di controllo nella gestione delle emergenze, garantendo la resilienza nelle infrastrutture di trasporto. Il team del progetto mira a definire un sistema integrato che possa utilizzare l'informazione sociale proveniente dai cittadini, informazione che viene organizzata e opportunamente interpretata, per poi essere fornita a supporto di enti e organizzazioni che si occupano di gestione delle emergenze.

L'aspetto innovativo del progetto sta nell'utilizzo dei social network in modo completo, cioè sia per acquisire e interpretare informazioni in fase di emergenza, sia per fornirle alle autorità competenti. In tal modo gli utenti del Web 2.0 diventano protagonisti nel fornire, mediante i social network, informazioni georeferenziate e dati utili per la gestione delle emergenze. Il progetto intende sfruttare al meglio queste informazioni creando un'infrastruttura che possa raccoglierle, interpretarle e opportunamente condividerle. Tale infrastruttura di elaborazione è basata su *piattaforme cloud* geograficamente distribuite e collegate tra di loro mediante rete ottica ad elevata capacità.

Questo capitolo presenta, a grandi linee, il Progetto TORCIA:

La sezione 3.1 espone quali sono le motivazioni e gli obiettivi per i quali è stato pensato il sistema;

La sezione 3.2 descrive il contesto in cui si cala il progetto e le scelte effettuate in fase di analisi;

## **Il Progetto TORCIA**

---

La sezione 3.3 effettua una panoramica sugli attori principali coinvolti nel progetto;

La sezione 3.4 presenta le attività principali che il sistema si prefigge di compiere;

La sezione 3.5 riassume il Workflow operativo che è stato stabilito e seguito nella realizzazione del progetto;

La sezione 3.6 dà una visione globale di quella che è l'architettura core del sistema;

La sezione 3.7 espone quali sono state le principali problematiche affrontate durante lo sviluppo del sistema e quali potranno essere i suoi impieghi futuri;

Infine la sezione 3.8 è un'introduzione al successivo capitolo 4, in cui parleremo del Motore Semantico.

### **3.1 Motivazioni e obiettivi**

La resilienza delle infrastrutture di trasporto è un aspetto critico da affrontare durante gli eventi di crisi. Infatti, l'accadimento di un fenomeno di tipo catastrofico che porta ad un black-out di specifiche reti stradali, interagendo con il sistema potrebbe generare tutta una serie di conseguenze negative sulle cose, le attività e l'ambiente del sistema socioeconomico e territoriale della regione circostante. Lavorare sulla resilienza significa, nel lungo termine, costruire infrastrutture di back-up in funzione del rischio, ma nel brevissimo periodo, all'atto dell'accadimento dell'evento catastrofico, può significare garantire la circolazione e lo scorrimento delle arterie considerate più strategiche nell'area considerata.

In ambito di Protezione Civile per consentire questo risultato si opera sui Piani di Emergenza, all'interno dei quali sono indicate le operazioni da fare, come ad esempio modifica del senso di circolazione e lo sgombero di determinate arterie da parte della polizia locale per consentirne l'utilizzo ai soli mezzi di soccorso. L'attuazione del piano tuttavia deve essere tempestiva e, per questo, occorre una comprensione immediata di quello che sta accadendo e l'elaborazione immediata delle comunicazioni da effettuare. L'obiettivo è di capire il prima possibile che si rientra nelle condizioni che devono attivare il Piano di Emergenza e comunicare con tempestività alla cittadinanza le azioni intraprese, finalizzate a mantenere le condizioni di sicurezza nonché quelle di

### 3.1 Motivazioni e obiettivi

---

percorribilità dell'infrastruttura stradale. Questi sono elementi chiave per una corretta gestione dell'emergenza.

Il progetto sfrutta le tecnologie attualmente disponibili per:

- facilitare le operazioni di raccolta dati sul territorio
- comprendere cosa si sta verificando
- dare una comunicazione tempestiva di quanto sta accadendo e delle iniziative da compiere

Questo è reso possibile dalla combinazione congiunta degli strumenti mobile e dei social network. Moltissime persone sono attualmente “always-on” e forniscono commenti in tempo reale su quanto sta accadendo e sulle situazioni che si verificano. Queste persone possono, inoltre, essere localizzate, dal momento che i dispositivi che utilizzano per condividere i propri pensieri sono spesso dotati di GPS. Uno dei primi obiettivi del progetto è quello di raccogliere, razionalizzare ed interpretare le informazioni che la cittadinanza condivide in modo autonomo, rendendole disponibili e fruibili quindi alle autorità preposte alla gestione dell'emergenza.

Obiettivo successivo è quello di poter elaborare questi dati, per lo più georeferenziati, per poter avere informazioni previsionali relativamente alla situazione dell'occupazione dell'infrastrutture di trasporti, elaborando delle indicazioni che devono essere intraprese sia dalle autorità che dalla cittadinanza. Per concludere si vogliono anche utilizzare le nuove tecnologie per fornire la possibilità di una comunicazione efficace, il più immediata possibile, sia tra i cittadini, ma anche tra autorità e cittadini stessi.

L'obiettivo del progetto TORCIA è quello di supportare la gestione delle emergenze tramite la raccolta di informazioni sul territorio e sulle infrastrutture di trasporto integrata dai contenuti pubblicati dagli stessi cittadini su siti web 2.0. Il sistema TORCIA struttura queste informazioni per renderle disponibile ai diversi attori coinvolti nella gestione dell'intervento, inclusi gli individui stessi colpiti dall'emergenza. L'infrastruttura tecnologica di TORCIA sarà progettata in modo da stimolare la produzione di informazione ed erogare servizi informativi secondo un approccio di *Crowdsourcing*.

## **Il Progetto TORCIA**

---

Il Crowdsourcing rappresenta un meccanismo di coordinamento orientato allo sfruttamento delle conoscenze e delle capacità di un numero elevato di individui, secondo procedure collaborative che stimolano e traggono vantaggio dalla naturale volontà individuale di contribuire alla soluzione di situazioni problematiche.

Da un punto di vista tecnologico il grado di innovazione più rilevante è l'utilizzo dei social network in modo pieno e che sarà utilizzato sia per interpretare l'informazione sia per fornirla. Twitter ed altri sistemi Web 2.0 sono ormai sempre più utilizzati per la raccolta e scambio di informazioni. Basti ricordare, per esempio, gli eventi della recente alluvione in Sardegna (17/11/2013), in cui per comprendere quello che stava accadendo e dove intervenire maggiormente, le autorità consultavano i post di Twitter. O ancora la rivoluzione in Egitto del Febbraio 2011, durante la quale la popolazione, per organizzarsi, ha comunicato esclusivamente tramite social network, Facebook in particolare.

Raccogliere le informazioni in tempo reale dai social media, eliminare il rumore, organizzarle e strutturarle mettendole al servizio delle istituzioni è una sfida tecnologica importante e che porterà benefici rilevanti. Un aspetto fondamentale, inoltre, è la resilienza del sistema progettato. Se questo sistema, al verificarsi di un evento catastrofico non fosse più operativo, creerebbe ancora più danni del non essere presente. Per questo motivo l'infrastruttura di elaborazione a supporto del sistema è da considerarsi un elemento chiave. Per far fronte a queste evenienze si è pensato ad una infrastruttura di elaborazione basata su piattaforme cloud geograficamente distribuite e collegate tra loro da una rete ottica resiliente ad alta capacità.

### **3.2 Il contesto specifico di analisi del progetto**

Al fine di eseguire un'analisi delle informazioni fornite dagli utenti, soprattutto quelle provenienti dai social network, il team del progetto ha deciso di stabilire da quali sorgenti di input attingere le informazioni.

Verrà quindi spiegato, nella sezione 3.2.1, quale social network si è stabilito di prendere come sorgente e quali motivazioni hanno portato a tale scelta.

## 3.2 Il contesto specifico di analisi del progetto

---

Successivamente, visti gli obiettivi del progetto, è stato deciso di considerare solo gli eventi che possono provocare emergenze in ambito cittadino e, tra di essi, è stata scelta un'unica tipologia di evento, descritta nella sezione 3.2.3, eseguendo la restrizione seguendo criteri logici basati sulle finalità di TORCIA.

### 3.2.1 Scelta della sorgente

I social network hanno una struttura di base comune, ma presentano, al tempo stesso, molti aspetti che li differenziano notevolmente gli uni dagli altri: per esempio le funzionalità implementate, i tipi di informazioni che veicolano, gli aspetti tecnologici e le impostazioni di *privacy*. Le considerazioni su questi aspetti hanno fatto sì che si profilasse un profilo ben preciso della sorgente ricercata:

1. deve essere utilizzata da un gran numero di persone che hanno un'elevata conoscenza del mezzo che stanno utilizzando;
2. deve possedere ridotte impostazioni di *privacy* per consentire l'acquisizione del maggior numero di messaggi possibile;
3. deve possedere un'interfaccia mediante la quale eseguire l'acquisizione;
4. deve permettere l'inserimento e la lettura di messaggi istantanea, senza essere moderata a priori;
5. deve poter veicolare un'informazione piuttosto concisa;
6. deve permettere la geocalizzazione del messaggio;
7. deve indicare precisamente la data e l'ora in cui il messaggio viene inserito;
8. deve inserire un meccanismo di ricondivisione per una maggior diffusione delle informazioni;

In base a queste considerazioni, la scelta è ricaduta sul social network *Twitter*.

Date le sue caratteristiche primarie, riesce a soddisfare tutti i punti precedentemente indicati. L'unica eccezione è per il punto 6: il *geotagging* non è una funzionalità obbligatoria per tutti i messaggi ma è comunque

impostabile opzionalmente dagli utenti.

In aggiunta ai Tweet provenienti da questa sorgente poi, saranno tenuti in considerazione i messaggi strutturati provenienti dagli utenti che utilizzeranno l'applicativo mobile, che sarà realizzato da Beta80. Si parla di messaggi strutturati in quanto sarà l'app stessa a guidare l'utente nella creazione di un messaggio che sia il più semplice e significativo possibile. Per questo e per altri motivi, come verrà spiegato nella relativa sezione XXX del prossimo capitolo, tale tipo di messaggi sarà considerato con una diversa rilevanza.

### 3.2.2 I dati in input

Quello che faremo è analizzare il tipo di informazione fornita dal social network Twitter e dall'app mobile in caso di emergenza. La natura generale di tale informazione può essere due tipi:

- Estrinseca: è un tipo di informazione chiara e riferita al contesto, contenente dati precisi e puntuali e non necessita quindi manipolazioni (tipicamente quella proveniente dall'app mobile);
- Intrinseca: è un tipo di informazione che necessita di essere estratta e interpretata mediante una serie di deduzioni che è possibile effettuare sulla sintassi del messaggio (tipicamente quella proveniente da Twitter).

Perciò, per prima cosa, va considerato il fatto che non sempre l'informazione sociale può essere dedotta e immediatamente registrata, ma spesso tale operazione deve essere preceduta da una fase di manipolazione.

Inoltre, se ci troviamo davanti ad informazione intrinseca, è ovvio che essa non sempre può essere dedotta e registrata, soprattutto da algoritmi informatici.

Per esempio, si è deciso che per il sistema TORCIA le coordinate della geolocalizzazione del tweet sono sostituite, se presenti, da coordinate intrinseche nella semantica del tweet stesso: ovvero le coordinate di un tweet diventano quelle di un qualsiasi indirizzo che viene scritto nel testo di 140 caratteri disponibili (es: “Disagi per allagamento in Piazzale Leonardo a Milano”).



## 3.2 Il contesto specifico di analisi del progetto

---

L'informazione, poi, può essere:

- Completa: permette la registrazione di più dati, specifici ad un evento e quindi maggiormente efficaci per coloro che dovranno utilizzarli;
- Incompleta: questo tipo di informazione è comunque utile e interessante, soprattutto se può essere integrata o confrontata con quella proveniente da altre segnalazioni.

Nonostante entrambi i tipi di informazioni siano utili, ci occuperemo di dedurre e registrare soprattutto le informazioni complete, proprio perché sono quelle più efficaci poi in fase operativa.

### 3.2.3 Scelta del contesto

Al fine di fare una migliore analisi all'interno del progetto, è stata scelta un'unica tipologia di evento.

La prima cosa da considerare è il fatto che l'evento deve produrre un certo volume di traffico in termini di messaggi e quindi deve potersi verificare con una elevata frequenza. Inoltre esso deve far sì che vengano prodotti messaggi sia nella fase di emergenza sia nelle fasi precedenti e successive. Le conseguenze dell'evento devono, poi, essere di molteplice natura in modo da poter coinvolgere enti diversi e produrre uno spontaneo "chiacchericcio" su Twitter da parte dei cittadini.

Basandosi su queste considerazioni, Andrea Palla, tesista presso il Politecnico di Milano e collaboratore del progetto, ha scelto come tipologia di evento di riferimento quella delle alluvioni e allagamenti [14].

Ecco le motivazioni da lui fornite in merito a tale scelta:

- sono fenomeni che si possono verificare in ambito cittadino;
- sono fenomeni frequenti sull'intero territorio italiano: l'Italia dal 1980 al 2002 è stato il paese europeo con il più alto numero di morti e danni causati dalle alluvioni;
- sono argomento di discussione frequente sui social media;
- giornali e organi di informazione se ne occupano pienamente, anche in previsione di un'emergenza;

## Il Progetto TORCIA

---

- cause, effetti, prevenzione e prevedibilità sono argomenti di discussione alla portata dei cittadini in quanto non richiedono una conoscenza specifica dell'argomento.

### 3.3 I partner del progetto

I partner di progetto presentano tra loro competenze diverse e complementari che favoriranno il raggiungimento degli obiettivi.

Alcatel-Lucent (ALU) coordinerà da un punto di vista amministrativo il progetto, grazie anche alle consolidate strutture aziendali di cui dispone in virtù della sua storia e dimensione. Dal punto di vista delle attività, contribuirà al progetto per la creazione di infrastrutture di rete resilienti e innovative, sulle quali innestare le applicazioni sviluppate nell'ambito di TORCIA.

Il coordinamento scientifico del progetto è affidato a POLIMI, che ha partecipato e partecipa a numerosi progetti di ricerca regionali, nazionali ed europei in collaborazione con imprese di tutte le dimensioni. Il gruppo di ricerca di POLIMI che parteciperà al progetto svolge inoltre ricerca da anni su temi di qualità dei dati e analisi semantica del Web 2.0, sviluppando così competenze messe a disposizione del team di progetto.

ACT Solutions offrirà al progetto la sua consolidata esperienza in ambito di ricerca operativa per la messa a punto del modello decisionale.

BETA80 ha una consolidata esperienza nello sviluppo di soluzioni di business intelligence, cruscotti di monitoraggio e integrazione di applicazioni. La precedente esperienza di BETA80 in progetti regionali in collaborazione con altri enti, uniti alla sua dimensione, la rendono un partner fondamentale per la realizzazione del cruscotto direzionale di TORCIA, il relativo applicativo mobile e per tutta l'integrazione di progetto.

Visto l'importante legame dello strumento prodotto da TORCIA con il territorio e la necessità di un supporto cartografico, le competenze di BETA80 saranno integrate dalle competenze specifiche di VIDIEMME Services in ambito di GIS (Geographical Information Systems) e cartografia.

La Fondazione Politecnico di Milano (FPM), infine, offrirà al consorzio di progetto la sua consolidata esperienza in termini di disseminazione, gestione delle relazioni con i media e gli stakeholder e validazione dei risultati. Forte

della sua esperienza nel campo, FPM supporterà ALU nel coordinamento del progetto.

#### 3.3.1 Alcatel-Lucent

E' il leader del progetto e fornisce a TORCIA la rete ottica backbone che collega i diversi server. Inoltre fornisce un sistema di messaggistica innovativo, per informare istantaneamente i cittadini e le autorità del pericolo imminente in corso: il framework UMB (Universal Message Broadcaster).

Per quanto concerne il settore delle comunicazioni ottiche saranno presi in esame due tipi d'innovazione. Il primo riguarda il possibile utilizzo dell'accesso in fibra ottica per portare le informazioni rilevanti ai centri di elaborazione del sistema. Il secondo si occupa dell'utilizzo della rete ottica ad alte prestazioni per interconnettere in cloud le risorse informatiche del sistema. Infine, per quanto riguarda il framework UMB, il valore aggiunto del progetto TORCIA risiede proprio negli aspetti di integrazione in un particolare contesto applicativo che prevede lo sviluppo di nuove funzionalità e nuove interfacce che ne abilitino l'integrazione.

Per quanto riguarda la piattaforma UMB, l'introduzione di nuove funzionalità e l'estensione delle interfacce, consentirà di accelerare i tempi di consegna e il time-to-market in quanto il tempo di integrazione e di eventuale personalizzazione del sistema potrà avvalersi degli sviluppi e dell'esperienza acquisita tramite il progetto TORCIA.

Le estensioni delle funzionalità e del campo applicativo dei sottosistemi che ALU considererà all'interno del progetto TORCIA sono elementi che contribuiranno all'ampliamento del mercato indirizzabile. Si evidenzia l'aspetto di integrazione che permetterà di indirizzare modelli di business sempre più rivolti ai servizi e alle applicazioni e a più alto valore aggiunto rispetto a quelli più standard delle infrastrutture di rete.

La dimensione della struttura conta più di 800 addetti e non è influenzata in termini numerici dall'attività del progetto TORCIA. Viceversa va evidenziata la significativa acquisizione di nuove competenze dal personale coinvolto dal progetto specie in riferimento ai nuovi aspetti applicativi e alle attività di integrazione peculiari del sistema.

## **Il Progetto TORCIA**

---

Le caratteristiche dell'infrastruttura del Sistema TORCIA possono beneficiare di considerazioni legate all'ottimizzazione del consumo energetico. In particolare la rete di accesso in fibra ottica, anche come possibile aggregatore di terminali radio risulta particolarmente efficiente nel rapporto consumo/banda. Allo stesso modo la distribuzione geografica delle risorse informatiche usate in modalità cloud consente un elevato livello di affidabilità senza la necessità di duplicazione delle risorse di calcolo. L'efficienza energetica della soluzione risulta di interesse rispetto alle tematiche di impatto ambientale.

### **3.3.2 ACT Solutions**

L'azienda è specializzata nella concezione di prodotti e nella fornitura di soluzioni basate su modelli matematici per il controllo di processi e il supporto alle decisioni. Le finalità di gestione di eventi catastrofici del progetto, che richiedono di operare decisioni in tempo reale in un contesto poco predicibile, costituiscono una sfida per ACT.

Le funzionalità introdotte contribuiranno alla crescita e all'ampliamento del portafoglio prodotti e del mercato indirizzabile anche all'estero. Le attività del progetto consentiranno di ampliare e potenziare i moduli ottimizzazione che lavorano su dati cartografici ed in ambito stocastico. Alcune delle nuove conoscenze maturate potranno rappresentare la base per innovare i prodotti esistenti tramite cross-fertilization in altri mercati verticali. L'esperienza maturata conduce ad una maggiore produttività del team impegnato in progetti che richiedono conoscenze simili quali ad esempio quelli relativi a distribuzioni merci cittadine a mobilità. L'attività trova sinergia anche a livello tecnologico relativamente ai prodotti/algoritmi su cui ACT è attiva. Si stima una sinergia delle attività in corso e quelle programmate di sviluppo prodotto tale da impattare con una riduzione dei costi di sviluppo del 20% relativamente agli algoritmi di ottimizzazione realtime associati all'elaborazione di geo-dati.

I nuovi mercati geografici da servire previsti nel business plan 2010-2015 di ACT presentano delle opportunità di mercato verticale che grazie all'esperienza del progetto potrebbero essere anticipate.

Dopo la produzione (e il consumo) di energia, il trasporto è il settore che contribuisce a bloccare la fuoriuscita del calore del sole attraverso i gas inquinanti generati, alterando gli equilibri dell'ecosistema. L'architettura introdotta

dal progetto TORCIA e gli algoritmi in carico ad ACT potrebbero trovare applicazione per un governo in real time dei flussi di traffico.

#### 3.3.3 Beta 80 Group

Essendo il leader in Italia nella realizzazione di Centrali Operative soprattutto in alcuni ambiti come Sanità, Protezione Civile e Vigili del Fuoco, essa fornisce il cruscotto di monitoraggio del progetto.

Beta80 ha una propria piattaforma adottata, EMMA (EMergency MAnagement), in grado di gestire sia le emergenze sanitarie sia emergenze legate a rischi naturali ed antropici (Vigili del Fuoco e Protezione Civile). Grazie al progetto TORCIA si prevede l'integrazione di due nuovi moduli EMMA-RO, destinato alla ricerca operativa ed EMMA-CROW, destinato alla raccolta delle informazioni dai social network. Tale azione consentirà di aumentare in modo significativo la competitività del prodotto estendendone le capacità di identificazione delle emergenze in termini più rapidi, sicuri ed accurati, nonché fornire informazioni più efficaci nella successiva gestione dell'evento per quanto riguarda gli aspetti logistici.

La commercializzazione di soluzioni finalizzate alla Gestione delle Emergenze è particolarmente complicata. Questo sia a causa della complessità della soluzione sia a causa del fatto che la soluzione realizzata si rivolge ad interlocutori pubblici le cui procedure di acquisto sono lunghe e laboriose. In questo senso, TORCIA può rappresentare una soluzione, dal momento che le funzionalità che saranno introdotte rappresentano elementi distintivi di cui i concorrenti sono privi e che consentiranno, per questo motivo, un'accelerazione delle procedure di acquisto.

#### 3.3.4 Vidiemme Services

Mediante il progetto TORCIA, Vidiemme Services intende accrescere in modo significativo le proprie competenze nel settore del GIS (Geographic Information System) e del Mobile Solutions, eseguendo il monitoring del traffico della zona osservata e fornendo informazioni sulle vie preferenziali per le evacuazioni e soccorsi.

Si prevede che la partecipazione a TORCIA avrà come risultato la possibilità di realizzare una soluzione di cruscotto di monitoraggio che, perfettamente

## **Il Progetto TORCIA**

---

integrata con la componente cartografica, garantirà un prezioso aumento del know-how e l'allargamento dell'offerta attuale.

Vidiemme ha esperienza nella realizzazione di Sistemi Informativi Territoriali utilizzando Software GIS, Applicazioni, Database e Ambienti di sviluppo dei principali vendor di mercato (Oracle, ESRI, etc). Quello che si vuole realizzare nell'ambito del progetto TORCIA però consiste nello slegare il cruscotto di monitoraggio da qualsiasi licenza di natura commerciale, compiendo nel contempo un significativo passo in avanti verso il mondo Open Source che, specie nell'ambito GIS, garantisce un'alta affidabilità grazie anche a una vasta community di utenti e sviluppatori. In questo modo si prevede un netto incremento qualitativo in termini di offerta da parte di una PMI come Vidiemme che si tradurrà in un aumento di competitività sul mercato della Pubblica Amministrazione Locale cui si rivolgono principalmente le offerte di servizi e progetti dell'azienda.

Il mercato di riferimento di Vidiemme è legato per lo più alla Pubblica Amministrazione Locale dove la capacità di spesa, specie per licenze software e relativi canoni di manutenzione, è andata scemando nel corso degli anni, accelerata dalla mancanza di risorse economiche erogate dal governo centrale. Pertanto la possibilità di offrire ai propri clienti soluzioni dove il costo delle Licenze software risulta nullo o quantomeno fortemente limitato, garantisce un elevato grado di competitività alla PMI e nel contempo la possibilità di accedere in tempi più rapidi a una conclusione contrattuale, essendo innegabile il vantaggio economico di un ente pubblico nel dotarsi di una soluzione interamente basata su piattaforma Open Source.

Grazie al progetto TORCIA, vedrà accrescere il proprio know how nell'utilizzo di prodotti Open Source Desktop e Server. Inoltre, il progetto, garantirà l'apprendimento e l'introduzione di nuove metodologie di sviluppo, esperienza che potrà essere utilizzata sia per la realizzazione di progetti simili, sia in contesti differenti che implicano, però, tutti la presenza di un'informazione cartografica di base. Ciò significa che allargando il panel dei prodotti offerti e di conseguenza il target dei potenziali clienti i costi iniziali legati all'acquisizione delle nuove tecnologie saranno ammortizzati via via nel tempo, senza contare il fatto che il ricorso a prodotti Open Source svincolerà Vidiemme dall'utilizzare e mantenere ancora soluzioni basate su software di mercato, abbattendo di conseguenza il costo dei propri prodotti e servizi.

#### 3.3.5 Dipartimento di Elettronica, Informazione e Bioingegneria del Politecnico di Milano

I ricercatori del Politecnico di Milano, e in particolare quelli del Dipartimento di Elettronica, Informazione e Bioingegneria, uno dei maggiori dipartimenti universitari in Europa per dimensioni e qualità della ricerca, forniscono il motore semantico Web 2.0, realizzando una struttura Web che possa essere facilmente letta dal sistema in progetto e integrata nello stesso.

Lo sviluppo di un motore semantico per il web 2.0 si inserisce in un filone di ricerca già avviato da tempo al Dip. di Elettronica e Informazione del Politecnico di Milano (POLIMI), che quindi potrà beneficiare del progetto per poter approfondire ricerche già in atto, nonché per accrescere la propria visibilità internazionale tramite pubblicazioni accademiche.

Le caratteristiche di innovazione e di complessità metodologica e tecnologica del progetto rendono indispensabile la partecipazione del Dipartimento di Elettronica e Informazione del Politecnico di Milano come parte del raggruppamento. Il Dipartimento di Elettronica e Informazione (DEI) è uno dei maggiori dipartimenti di ICT d'Europa per dimensioni e produzione scientifica. La ricerca è l'obiettivo principale delle attività dipartimentali e viene condotta secondo gli standard internazionali più elevati. In questo campo, il DEI vanta competenze consolidate nelle aree dell'automazione, dell'elettronica, dell'informatica e delle telecomunicazioni.

Il Dipartimento di Elettronica e Informazione del Politecnico di Milano parteciperà al progetto con il gruppo di ricerca di Sistemi Informativi. La ricerca del gruppo di Sistemi Informativi si focalizza sulla progettazione architeturale di sistemi informativi, con particolare attenzione alla relazione tra requisiti e progettazione, all'impatto organizzativo delle scelte di progettazione, all'uso di componenti a livello concettuale per il design di sistemi informativi e ad aspetti tecnologici per lo sviluppo degli stessi.

Le competenze messe a disposizione apparterranno soprattutto nell'ambito dell'analisi semantica, della data quality e dei Web service. In quest'ambito l'Organismo di ricerca vanta un'esperienza decennale, con risultati di ricerca di rilevanza internazionale e oltre venti progetti di ricerca nazionali e internazionali. In particolare, nell'ambito della progettazione architeturale dei sistemi informativi e della data quality il Dipartimento di Elettronica e Informazio-

## Il Progetto TORCIA

---

ne del Politecnico di Milano vanta una leadership europea e rappresenta un centro di eccellenza internazionale, come testimoniato dal livello qualitativo e dall'impatto della produzione scientifica.

Tali competenze verranno messe a disposizione del progetto TORCIA, con i seguenti fondamentali obiettivi:

1. Coordinamento scientifico del progetto – L'organismo di ricerca sarà responsabile del coordinamento scientifico di tutto il progetto. L'obiettivo del coordinamento scientifico è la valutazione delle scelte metodologiche e progettuali dal punto di vista della loro:
  - Innovatività
  - Validità scientifica in termini di replicabilità
  - Generalizzabilità e applicabilità in più contesti di gestione di emergenze e, se possibile, in altri domini applicativi diversi dalla gestione delle emergenze, ma con analoghi requisiti di tempo reale
  - Flessibilità
  - Interoperabilità.
2. Esecuzione di task metodologici – L'organismo di ricerca si occuperà in particolare della definizione delle linee guida metodologiche per il crowdsourcing, che richiede competenze metodologiche di progettazione architeturale di sistemi informativi.

Più nel dettaglio il Polimi si è occupato della realizzazione di un Tool di Classificazione delle informazioni ricevute, basato sull'algoritmo ideato da Claudio Carcaci [5], e di un Tool di Entity Discovery per la rilevazione delle emergenze, basato sull'algoritmo di Giulia Cesana e Adolfo D'Ettore [7]. I due tool sono stati poi da noi riadattati ed inglobati in un unico software, chiamato Motore Semantico, che è stato a sua volta integrato nel sistema TORCIA, con l'obiettivo di individuare e segnalare preventivamente le emergenze.

### 3.3.6 Fondazione Politecnico di Milano

La Fondazione, voluta dal Politecnico e da altre aziende per favorire le collaborazioni dei dipartimenti universitari con le aziende e la pubblica amministrazione



ne, ha il compito di curare le attività di diffusione e affiancare Alcatel-Lucent nel coordinamento e nella gestione del progetto.

## 3.4 Le attività core del sistema

Le attività del progetto si occupano principalmente dello sviluppo di un sistema che sia in grado di monitorare, interpretare e comunicare le informazioni provenienti dai social network.

### Monitoraggio

Il monitoraggio dovrà avvenire in tempo reale analizzando sia i dati provenienti dai canali tradizionali sia quelli raccolti tramite *crawling* sui social network, garantendo la presenza di un cruscotto cartografico operativo.

### Interpretazione

Sarà necessario interpretare immediatamente le informazioni relative a ciò che sta accadendo, in modo da avere indicazioni sull'impatto che questo avrà sull'infrastruttura di trasporto locale e indicare le azioni da intraprendere per gestire l'evento critico.

### Comunicazione

Dovranno essere fornite tutte le informazioni necessarie, sia a livello di istituzioni che a livello di cittadinanza, diffondendo i dati in ottica multicanale per poter rendere immediate le iniziative che devono essere intraprese.

Al fine di rendere il sistema utilizzabile e funzionante anche in condizioni di maxi-emergenza ed eventi catastrofici, dovranno essere studiati metodi innovativi per rendere resiliente l'infrastruttura mediante *cloud computing* distribuito.

Queste attività rientrano all'interno del modello di gestione delle emergenze che, per la pianificazione di tali procedure e l'erogazione dei corrispondenti servizi, prevede le seguenti attività cicliche, riportate nella figura 3.1:

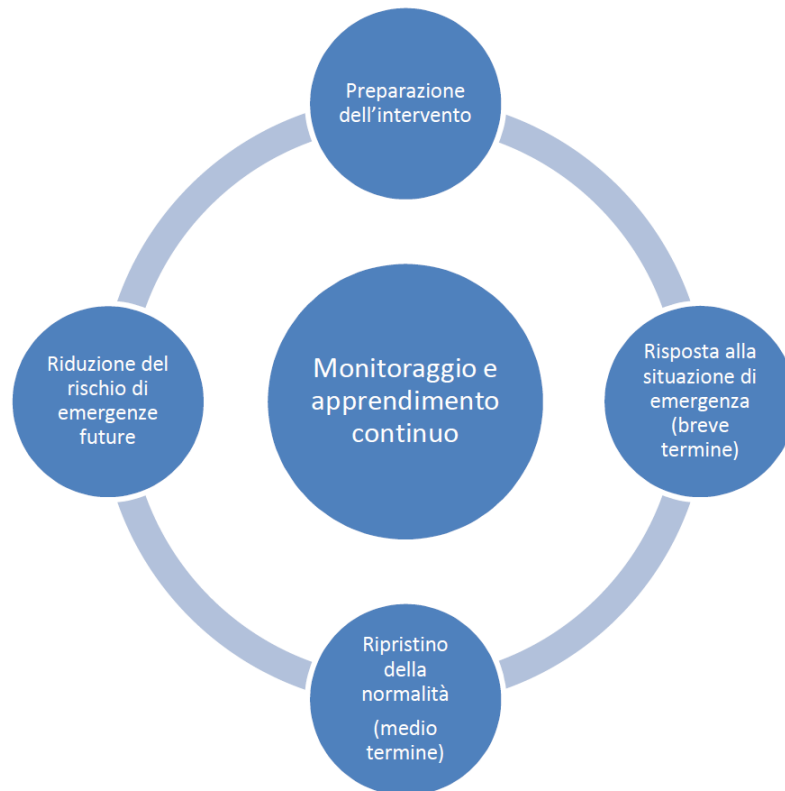


Figura 3.1: Il flusso della attività svolte dal sistema TORCIA.

1. Pianificazione dell'intervento
2. Monitoraggio
3. Gestione dell'emergenza
4. Post-Emergenza
5. Mitigazione del rischio di emergenze future

Il progetto TORCIA, pur lambendo tutti gli aspetti precedentemente considerati, si focalizza principalmente su:

- Fase di monitoraggio, che consentirà di avere a disposizione una “rete di sensori virtuali”, rappresentata dagli stessi cittadini e dai messaggi che loro lasciano nell'ambito dei social network. Tale rete virtuale sarà focalizzata principalmente sulla resilienza dell'infrastruttura stradale.

### **3.5 Implementazione del progetto TORCIA: piano di attività**

---

- Comunicazioni legate alla fase di gestione di emergenza, che consentono una comunicazione tempestiva agli stessi cittadini che hanno inviato le informazioni. Il sistema sarà in grado di predisporre messaggi sulla base dei dati rilevanti ed inviarli sia ai cittadini sia alle altre istituzioni coinvolte. Tali comunicazioni potranno avvenire real-time.
- Fase di pianificazione, che si focalizza sugli interventi e sull'utilizzo dell'infrastruttura stradale, grazie alla disponibilità di analisi what-if combinando i dati storici, i dati provenienti dalle reti tradizionali e le reti di sensori virtuali.

## **3.5 Implementazione del progetto TORCIA: piano di attività**

La realizzazione del sistema TORCIA si è articolata in diverse fasi successive, secondo un Workflow operativo ben scandito che ha permesso di ottimizzare i tempi di lavoro e l'efficienza degli agenti coinvolti. La divisione del lavoro ha così permesso ad ogni azienda di focalizzarsi sul proprio campo di forza, integrando poi i lavori finali. Anche per questo le fasi, pur essendo distinte, non sono state percorse sequenzialmente, ma si è potuto parallelizzarne le attività.

### **3.5.1 WP1 – Requisiti del sistema, architettura e analisi degli scenari**

Il WP1 si pone l'obiettivo di definire l'architettura del sistema, identificandone i sottosistemi e lo schema di funzionamento generale, di identificare i requisiti base del sistema e di ipotizzare i possibili scenari di utilizzo.

Si tratta della fase di identificazione dei requisiti e degli obiettivi del sistema TORCIA, sia nel suo complesso che per gli aspetti specifici delle varie tecnologie coinvolte. Sono quindi valutate le prestazioni attese del sistema, definite le metriche necessarie alla loro verifica.

Nell'architettura si possono identificare due macro-aree:

- l'infrastruttura del sistema, composta dalla rete di accesso e dalla rete core

## **Il Progetto TORCIA**

---

- gli applicativi di elaborazione per l'analisi delle informazioni e la distribuzione dell'informazione

Gli scenari di utilizzo del sistema sono volti a valutare la gestione della viabilità in concomitanza di eventi catastrofici. Si ritiene che il sistema si possa facilmente prestare ad estensioni del proprio utilizzo che verranno valutate.

### **3.5.2 WP2 – Progettazione e realizzazione del sistema raccolta dati**

La raccolta dati ha due sorgenti: le tradizionali (Radio, Tom Tom) che rappresentano la fonte di partenza con cui effettuare le analisi dell'utilizzo e della congestione delle infrastrutture e le fonti provenienti dal mondo Web 2.0, raccolte mediante crawling.

Successivamente vi è stata una fase di identificazione del formato e dello standard da utilizzare per raccogliere i dati elementari. Grazie a questa attività sarà possibile ricondurre le sorgenti a un formato comune, in modo che possano essere interpretate.

Si sono poi identificate tutte le fonti da cui si pensa di poter effettuare la raccolta dei dati e le modalità di storage dei dati stessi, per poi passare alle attività di trasformazione dei dati, necessarie per rendere questi interpretabili dal motore di analisi.

### **3.5.3 WP3 – Progettazione del motore di analisi**

L'obiettivo del WP3 è la progettazione del motore di analisi delle informazioni fornite dagli utenti.

Questo permetterà la categorizzazione delle conversazioni su social media in base a un'ontologia progettata specificamente per la gestione di emergenze basata su di una classificazione di criticità rilevanti per le procedure di intervento dei diversi attori coinvolti. L'analisi permetterà la geolocalizzazione e risponderà a requisiti di tempo reale o quasi reale.

La business intelligence permetterà di individuare tutte le informazioni utili alla riduzione del rischio di future emergenze. Sarà basta su tecniche di data mining semantico, che consentono l'individuazione di eventi (per es.,

### **3.5 Implementazione del progetto TORCIA: piano di attività**

---

picchi conversazionali su una specifica criticità) legati a un'inadeguatezza (non voluta e non prevedibile) delle procedure di gestione dell'emergenza.

#### **3.5.4 WP4 – Progettazione e realizzazione del motore di calcolo per l'indirizzamento ed il supporto alle decisioni**

L'obiettivo del WP4 è la realizzazione del motore di ottimizzazione che a partire da informazioni contenute nei piani di intervento e quelle in ricevute dai social network stabilisce i suggerimenti sulle vie di fuga e vie di accesso ai diversi punti.

Si formalizzano per questo le informazioni necessarie per il motore di calcolo, che permettono di definire preferenze o indicazioni preventive che il sistema deve considerare, la mappatura di zone e di punti di interesse e la caratterizzazione delle risorse e funzioni che dovranno intervenire nei diversi tipi di intervento.

Si formalizzano poi anche le informazioni che descrivono lo stato dell'emergenza (in real-time) nelle diverse zone e punti, come pure degli input che consentono ad un operatore della sala controllo di esprimere degli indirizzamenti che il motore di calcolo dovrà eseguire.

#### **3.5.5 WP5 – Progettazione e realizzazione del cruscotto di monitoraggio e framework di distribuzione delle informazioni con supporto cartografico**

L'obiettivo del WP5 è stata la progettazione e la successiva realizzazione del cruscotto di monitoraggio che sarà in grado di monitorare le condizioni di sicurezza e le corrette condizioni di trasporto. Allo stesso tempo il WP5 si è occupato dell'integrazione della piattaforma di delivery dei messaggi UMB, dotata delle opportune estensioni di funzionalità e di adattamento delle interfacce.

Il Cruscotto Operativo realizzato tiene conto delle condizioni in cui operano normalmente gli addetti di Protezione Civile e Polizia Locale, potenziali utilizzatori dello stesso. Il sistema include una parte cartografica ed una parte di presentazione di messaggi.

## **Il Progetto TORCIA**

---

Contemporaneamente c'è stata un'analisi e una progettazione dettagliata della piattaforma UMB, dei componenti necessari alle estensioni funzionali della piattaforma e la definizione delle interfacce per la successiva integrazione. La realizzazione è in grado di diffondere i messaggi richiesti dagli scenari selezionati e lavorare in modo integrato con gli altri componenti del sistema.

### **3.5.6 WP6 – Applicativo mobile web 2.0**

Il WP6 si concentrerà sulla componente mobile di TORCIA. Questa sarà installabile su tutte le tipologie di device e sarà in comunicazione costante con il sistema. Il componente ha due obiettivi: consentire le segnalazioni al sistema che si aggiungeranno a quelle raccolte dai social network e rendere disponibili le informazioni proposte da TORCIA all'utente in funzione della situazione di utilizzo

### **3.5.7 WP7 – Progettazione infrastruttura di rete**

L'obiettivo del WP7 è l'identificazione delle caratteristiche e dell'architettura delle reti che costituiscono l'infrastruttura portante del sistema.

In particolare si possono individuare tre sottosistemi di rete:

1. Rete di accesso per il recupero delle informazioni di input al sistema
2. Rete ottica resiliente per l'interconnessione dell'infrastruttura cloud
3. Rete di distribuzione degli output del sistema

Oltre all'utilizzo di tecnologie esistenti, che comunque sono inquadrate all'interno di una proposta architetture innovativa del sistema complessivo e caratterizzate in termini di prestazioni, ci si pone l'obiettivo di svolgere attività di ricerca sulla rete di accesso ottico a larga banda e sull'integrazione della rete ottica con i sistemi di cloud.

### **3.5.8 WP8 – Integrazione, dimostrazione e validazione**

L'obiettivo del WP8 è l'integrazione complessiva della soluzione TORCIA nell'infrastruttura di rete resiliente e la successiva dimostrazione e validazione del funzionamento del sistema.

### **3.5 Implementazione del progetto TORCIA: piano di attività**

---

L'integrazione applicativa è stata realizzata negli ambienti di sviluppo di BETA80. Il task è focalizzato sulla verifica della correttezza del dialogo tra le componenti.

L'integrazione nell'infrastruttura di rete resiliente è invece predisposta negli ambienti forniti da ALU.

#### **3.5.9 WP9 – Disseminazione**

Gli obiettivi del WP9 sono la definizione e l'attuazione di un piano di comunicazione che possa adeguatamente diffondere i risultati raggiunti, con anche lo scopo di sensibilizzare gli stakeholder.

Il piano include:

- La creazione di un sito web di progetto per permettere anche di raccogliere l'opinione dei visitatori; in accordo con le tematiche sviluppate nel progetto, il sito è dotato di strumenti collaborativi di interazione con l'utente in ottica Web 2.0 (forum, mailing list, wiki, ecc.)
- La realizzazione di un workshop di mezza giornata per il kick-off del progetto e di un convegno finale per la presentazione dei risultati
- L'elaborazione di comunicati stampa e gestione dei rapporti con le principali testate generaliste e di settore
- Azioni di sensibilizzazione presso gli stakeholder
- Presentazione in Polonia **(DA METTERE A POSTO CON NUOVE INFORMAZIONI CHE TROVEREMO)**

I risultati di volta in volta raggiunti e verificati durante le riunioni di coordinamento verranno inoltre presentati sia ai soggetti istituzionali che alla comunità scientifica tramite pubblicazioni su riviste scientifiche, partecipazione ed organizzazione di congressi da parte dei membri dei team. Le attività di disseminazione sono coordinate dalla Fondazione Politecnico di Milano, che svolge da diversi anni questo ruolo per progetti di collaborazione pubblica-privata coinvolgenti il Politecnico di Milano e ha quindi avuto occasione di sviluppare un significativo network di stakeholder e utenti potenzialmente interessati ai servizi sviluppati in TORCIA.

### 3.6 L'architettura core del sistema

In figura 3.2 è indicato uno schema al alto livello dell'architettura del sistema considerato.

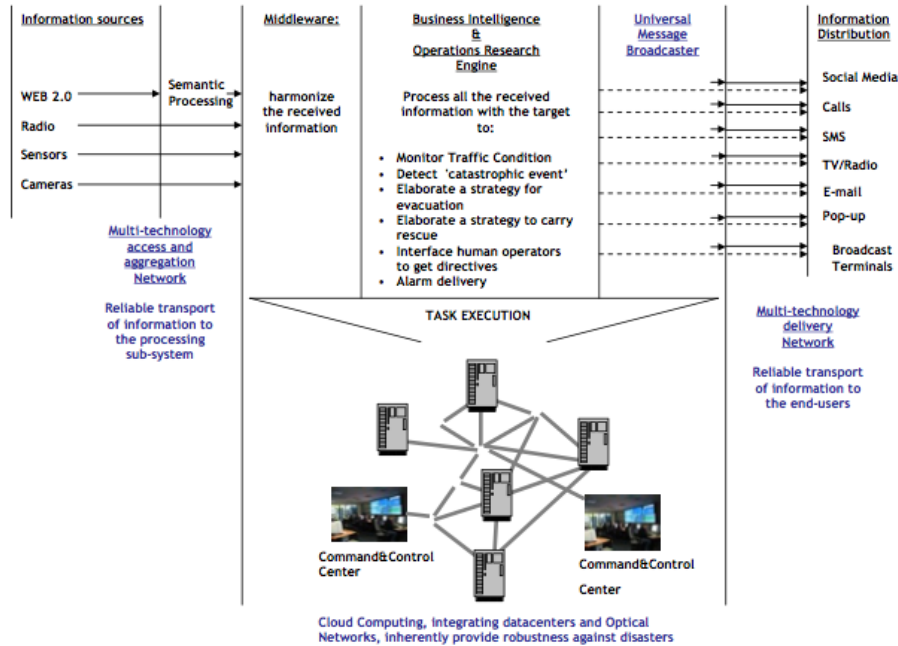


Figura 3.2: Architettura del sistema TORCIA.

#### 3.6.1 Input del sistema

Il sistema riceve in ingresso informazioni di natura diversa:

- Messaggi indiretti: comprendono tutti quei messaggi non diretti alla centrale operativa ma estratti da altre fonti, che normalmente non trattano informazioni relative alle emergenze. In questa categoria ritroviamo tutti i messaggi estratti dai social network Web-based che dovranno essere approfonditamente analizzati al fine di estrarre informazione strutturata.
- Messaggi diretti: comprendono tutti i messaggi indirizzati direttamente alla centrale operativa, inviati sia da fonti istituzionali che da persone private, e comprendono tutte le chiamate agli operatori di emergenza e i messaggi inviati tramite applicazioni mobili apposite.
- Dati provenienti da dispositivi elettronici: si tratta di dati provenienti da sensori e telecamere, strumenti già installati e preposti al controllo del



## 3.6 L'architettura core del sistema

---

territorio. Questo tipo di informazione aggiuntiva verrà utilizzata sia per permettere un'immediata localizzazione degli eventi, sia per valutarne la criticità.

### 3.6.2 L'analisi semantica

Il core del sistema racchiude tutte le operazioni volte all'analisi, la comprensione e l'utilizzo delle informazioni giunte in input. Il Politecnico di Milano è responsabile dello sviluppo del motore di analisi che dovrà rispondere alle seguenti esigenze:

- **Analisi semantica:** permetterà una categorizzazione dei messaggi prelevati dai social media mediante un'ontologia, creata specificatamente da Claudio Carcaci per la gestione delle emergenze [5]. L'analisi verrà eseguita mediante un tool di Classificazione che sfrutta la sopra citata ontologia, e mediante un tool di Entity Discovery, che sfrutta l'algoritmo di Giulia Cesana e Adolfo D'Ettore, per rilevare le emergenze [7]. I due tool sono stati da noi estesi, e integrati all'interno del sistema TORCIA.
- **Business Intelligence semantica:** si occuperà di individuare informazioni utili al fine di migliorare l'efficienza del sistema e ridurre il rischio di future emergenze. Si baserà su tecniche di data mining semantico, permettendo l'individuazione di eventi.
- **Crowdsourcing e analisi dei piani di emergenza:** si forniranno soluzioni per la comunicazione tra le autorità competenti e l'aggiornamento collaborativo dei piani di emergenza.

ACT Consulting, invece, si occuperà di sviluppare il motore di calcolo per la definizione dei percorsi e delle vie di fuga, minimizzando i tempi di percorrenza e la pericolosità dei percorsi considerati.

### 3.6.3 Output del sistema

L'output del sistema dovrà essere utilizzato a supporto delle decisioni dai diversi enti preposti alla salvaguardia del cittadino. Esso permetterà una migliore comprensione dello scenario di emergenza e, di conseguenza, fornirà un valido aiuto per la definizione delle azioni necessarie.

## **Il Progetto TORCIA**

---

Infatti, grazie al cruscotto di monitoraggio, sarà possibile geolocalizzare le richieste di intervento, leggere tutte le informazioni associate alla segnalazione e visualizzarne le immagini/fotografie. Saranno, inoltre, sviluppate apposite applicazioni per smartphone che permettano un invio di messaggi diretti al sistema.

Le funzionalità di output sfrutteranno l'UMB, sistema sviluppato da Alcatel-Lucent, che provvederà all'invio di messaggi in broadcast attraverso diverse le piattaforme: social network, email, sms, allarmi, etc.

### **3.6.4 Destinatari del sistema**

I destinatari dei messaggi generati dal sistema TORCIA appartengono a categorie diverse:

- Autorità, come il Sindaco, il Presidente della Provincia, il Prefetto, etc.
- Squadre di emergenza, come Vigili del Fuoco, Pronto Soccorso, Protezione Civile
- Cittadini

Ovviamente i messaggi di output del sistema saranno personalizzati in base alla categoria a cui sono rivolti: per esempio i cittadini riceveranno solo informazioni relative all'evento in atto.

## **3.7 Complessità di progetto e impieghi futuri**

Il progetto ha caratteristiche di notevole complessità sia metodologica, sia tecnologica.

Un primo rilevante elemento di complessità è l'approccio orientato al crowdsourcing, che rappresenta una metodologia di cooperazione ancora molto nuova e per la quale non esistono linee guida di progettazione consolidate e condivise. Da questo punto di vista, la letteratura fornisce numerose indicazioni e una ricca case history sulle difficoltà connesse a una efficace messa in opera di meccanismi di coordinamento basati sul crowdsourcing. Viene spesso constatato che molti casi di insuccesso sono legati alla erronea assunzione che

### 3.7 Complessità di progetto e impieghi futuri

---

gli individui mettano a disposizione la propria conoscenza in modo spontaneo se vengono dati loro gli strumenti tecnologici per farlo. Al contrario, la tecnologia è un fattore abilitante per un crowdsourcing di successo, ma non garantisce la sua efficacia, che dipende piuttosto dai meccanismi di incentivazione e dal valore aggiunto che gli utenti percepiscono nei servizi messi a loro disposizione a compenso della loro collaborazione. La piattaforma deve essere quindi implementata in modo che si stimoli la collaborazione, fornendo la giusta compensazione a chi contribuisce alla generazione di conoscenza.

Nel particolare contesto della gestione delle emergenze, un ulteriore elemento di complessità tecnologica è la necessità di operare in tempo reale. È noto come le tecnologie semantiche richiedano tipicamente analisi complesse e abbiano di conseguenza tempi di risposta non brevi. Nel Motore Semantico si enfatizzerà quindi la necessità di progettare algoritmi semantici agili e verticalizzati nel settore della gestione delle emergenze, con un'ontologia fortemente personalizzata sulle esigenze dello specifico dominio.

Inoltre la necessità di informazioni di geolocalizzazione e di supporti cartografici su dispositivo mobile pone il problema di progettare un'architettura complessiva particolarmente efficiente, anche in considerazione della necessità di far cooperare una molteplicità di attori distinti. Con lo strumento cartografico e gestionale reso disponibile, sarà possibile sia effettuare l'attività di monitoraggio utilizzando le informazioni previste, sia effettuare le attività a supporto alle decisioni, nonché effettuare l'elaborazione delle informazioni in fase di emergenza e veicolare l'informazione in modo indistinto verso i social network.

Un ulteriore e non trascurabile elemento di complessità tecnologica è poi l'infrastruttura di rete che deve garantire caratteristiche di resilienza che permettano alle informazioni di continuare a fluire anche in situazioni di emergenza. Si deve quindi progettare un'infrastruttura resiliente a supporto dell'intero sistema. In particolare sono prese in considerazione da un lato le nuove tecnologie di accesso in fibra ottica a supporto di quello che possiamo definire l'I/O di TORCIA, dall'altro l'integrazione della rete ottica con i sistemi di calcolo, che rappresenta uno degli argomenti di massimo interesse della ricerca come testimoniano i numerosi progetti internazionali che lavorano sul tema.

## Il Progetto TORCIA

---

L'elevato contenuto di ricerca e la multidisciplinarietà del progetto hanno come obiettivo principale la creazione di un servizio innovativo per la gestione di eventi catastrofici. I risultati ottenuti puntano dunque alla definizione di nuove soluzioni nelle aree tecnologiche di competenza dei partner di progetto. Vestono un ruolo di particolare importanza l'aspetto d'integrazione dei sottosistemi che compongono la soluzione. Tale approccio punta dunque a un miglioramento della capacità di innovare delle rispettive organizzazioni di R&S, con un conseguente incremento della competitività dei singoli prodotti e al miglioramento dei processi di collaborazione tra PMI e grandi imprese.

Nell'ottica di utilizzo futuro, il progetto TORCIA si pone come obiettivo di sviluppare strumenti di supporto per la gestione delle infrastrutture critiche di trasporto in contesti di emergenza, sia dovuti a fenomeni naturali (per es. terremoti, alluvioni, incendi) che sociali (per es. manifestazioni popolari, grandi eventi come l'EXPO 2015). Attualmente ci si è focalizzati sul primo di questi aspetti, problemi legati a fenomeni naturali.

TORCIA vuole favorire lo sviluppo di processi collaborativi per poter gestire al meglio le situazioni di crisi e aumentare la sicurezza dei cittadini, sfruttando in particolare la collaborazione degli stessi cittadini analizzando i loro contributi in ambito Web 2.0, tramite innovative tecnologie di crowdsourcing.

La sicurezza del territorio, la digitalizzazione e l'attenzione a eventi della portata di EXPO 2015 sono peraltro priorità regionali esplicitamente menzionate nel Piano Regionale di Sviluppo (PRS) della IX Legislatura. Il crowdsourcing rappresenta una metodologia nuova e ancora largamente inesplorata per la realizzazione di processi collaborativi che vedono coinvolti una molteplicità di attori istituzionali e un numero potenzialmente elevato di individui. Questo tipo di processi presentano alcuni aspetti chiave che li contraddistinguono come critici, in particolare in relazione alla progettazione di una piattaforma tecnologica di supporto unica e integrata. Da un lato, la molteplicità di attori istituzionali coinvolti crea il problema dell'integrazione di procedure e sistemi fra più organizzazioni, operazione sempre difficile dal punto di vista organizzativo e costosa dal punto di vista tecnico. Dall'altro lato, la natura stessa delle emergenze, che comportano inevitabilmente situazioni eccezionali e difficilmente prevedibili a priori, implica requisiti di grande flessibilità sia organizzativa sia tecnologica.

Da questo punto di vista il Web 2.0 fornisce per la prima volta nella storia

dell'evoluzione dell'informatica un insieme di tecnologie e applicazioni per l'erogazione di servizi con le necessarie caratteristiche di interoperabilità e flessibilità. Il Web 2.0 mette infatti a disposizione una serie di servizi già usati dagli utenti, specialmente in mobilità, che possono essere integrati nella piattaforma di gestione delle emergenze sia per raccogliere informazione sull'emergenza stessa, sia per erogare servizi da parte degli attori istituzionali. Si pensi per esempio agli eventi di cronaca dei mesi scorsi, in cui manifestazioni di protesta popolari in alcune grandi città italiane che hanno avuto pesantissimo impatto sulla sicurezza cittadina sono stati coordinati tramite l'utilizzo di social network, che hanno rappresentato uno dei principali mezzi di comunicazione fra i manifestanti. Considerare e analizzare il Web 2.0 in previsione di eventi di massa quali Expo 2015 è quindi cruciale per poter monitorare le città e le infrastrutture critiche.

A differenza del passato, il coinvolgimento degli utenti finali e l'integrazione di servizi erogati da attori diversi sono possibili a costi non eccessivi, permettendo quindi un continuo aggiornamento della piattaforma secondo un processo di apprendimento continuo. Il progetto TORCIA si propone di trarre vantaggio dalla disponibilità delle tecnologie e delle applicazioni del mondo Web 2.0 per l'implementazione di una piattaforma integrata di gestione delle emergenze. Questo permette di integrare tutti gli attori coinvolti nella gestione delle emergenze nel territorio, fornendo un'infrastruttura tecnologica flessibile che può essere adattata alla gestione di diversi tipi di emergenze in molteplici aree territoriali al servizio di raggruppamenti di attori di volta in volta distinti.

### 3.8 Il nostro lavoro di tesi

Quello visto fin ora è una descrizione a livello generale di tutta la struttura del sistema TORCIA. Come abbiamo già più volte menzionato, il Politecnico di Milano è responsabile della realizzazione del Motore Semantico e della sua integrazione all'interno del progetto.

Il nostro lavoro di tesi, in particolare, si è focalizzato sull'implementazione di un software che racchiude le funzionalità di Classificazione, che sfrutta l'algoritmo di Claudio Carcaci [5], e di Entity Discovery, basata invece sull'algoritmo di Giulia Cesana e Adolfo D'Ettore [7]. Per raggiungere tale scopo,

## **Il Progetto TORCIA**

---

entrambi gli algoritmi originali sono stati da noi ripresi e significativamente estesi, come descritto nel prossimo capitolo.

Infine ci siamo occupati della parte di integrazione, implementando i Web Services necessari allo scopo, in modo da poter comunicare con il sistema TORCIA complessivo.

Nel corso del nostro lavoro, abbiamo sviluppato anche alcuni altri tool che ci sono serviti da supporto per il monitoraggio e il testing sia del software finale, che di alcune idee avute nel corso d'opera.

Nel prossimo capitolo scenderemo nel dettaglio, spiegando struttura, caratteristiche e funzionalità del Motore Semantico, nonché le tappe percorse per arrivare alla sua realizzazione finale.

# Capitolo 4

## Il Motore Semantico

Il Motore Semantico è uno dei componenti fondamentali del progetto TORCIA.

Le sue criticità e funzionalità risiedono nella capacità del software di analizzare messaggi provenienti dal Web e di riconoscere se l'informazione portata è rivelatrice di una potenziale emergenza imminente, in modo da trasmettere un segnale di allarme al resto del sistema e allertarne gli utilizzatori con anticipo.

In questo capitolo si parlerà approfonditamente del Motore Semantico e, in particolare:

La sezione 4.1 spiega come opera il software a grandi linee;

La sezione 4.2 è un'introduzione alle parti principali che compongono il Motore Semantico e che verranno meglio descritte in seguito;

La sezione 4.3 analizza nello specifico il Web Service usato per la comunicazione col Crawler;

La sezione 4.4 presenta il componente Classificatore, il suo algoritmo di base e la sua integrazione nel sistema;

La sezione 4.5 presenta il componente Entity Discovery, il suo algoritmo di base e la sua integrazione nel sistema;

La sezione 4.6 mostra le caratteristiche e scopi del Tool di Monitoraggio da noi sviluppato;

La sezione 4.7 è una riflessione su possibili modifiche, innovazioni o cambiamenti che potrebbero essere applicati al Motore Semantico.

Per semplicità di trattazione, col termine “tweet” ci riferiremo, da qui in poi, sia ai messaggi strutturati provenienti dall’ap mobile che ai tweet veri e propri, in quanto il loro formato, come vedremo nella successiva sezione 4.3, è quasi del tutto identico. Nel caso si renda necessario, verrà specificato chiaramente se ci si sta riferendo a una delle due tipologie di messaggio in particolare.

### 4.1 Funzionamento Generale

Come già accennato, l’obiettivo fondamentale del nostro Motore Semantico è di ricevere e analizzare messaggi provenienti dal Web per rilevare eventuali situazioni di emergenza in corso o, meglio ancora, che stanno per verificarsi.

A questo scopo, il software utilizza tre moduli principali e integrati tra loro:

- Web Service
- Classificatore
- Entity Discovery

Queste parti poi, si appoggiano a un Database, appositamente creato per conservare le informazioni, i tweet grezzi. Il DB serve anche come struttura di appoggio per permettere una loro successiva rielaborazione.

Il Web Service riceve dei file, in un formato XML prestabilito, contenenti i dati acquisiti da Beta80. In una fase successiva, il Classificatore li analizza e, mediante una prima valutazione semantica del contenuto del testo salvato, assegna una o più Classi a ogni nuovo tweet significativo, scelte tra le sette disponibili di:

1. Dove
2. Segnalazione
3. Conseguenze
4. Responsabilità
5. Viabilità



6. Warning

7. Post (Postemergenza)

Ognuna di queste corrisponde a una fase o a un attributo fondamentale legato all'emergenza in corso e, nell'insieme, risultano ampiamente significative non solo nella fase di Emergenza in Corso, ma anche e soprattutto in fase di Emergenza Conclusa, come verrà spiegato nella sezione 4.4 relativa al Tool di Classificazione.

Una volta che i tweet sono stati classificati, il programma di Entity Discovery analizza, sempre con tecniche semantiche, tutte le informazioni raccolte e classificate in un lasso di tempo definito (al momento settato sulle ultime 24 ore). Il suo scopo è di ricercare delle correlazioni tra questi tweet e di controllare la presenza di parole chiave (da qui in poi Keyword), come ad esempio "ALLUVIONE" o "NUBIFRAGIO", legate ad aree geografiche italiane. Se vi è un riscontro, l'Entity Discovery crea una Segnalazione contenente tutti i tweet da cui si è generata, in un formato XML analogo a quello di ricezione, e il Web Service si preoccupa di inoltrarla al Web Service di Beta80, da cui la segnalazione continuerà il suo iter all'interno del sistema TORCIA.

Oltre alla segnalazione di emergenze in corso, abbiamo voluto dare un carattere di prevenzione effettiva al Motore Semantico, creando un algoritmo in grado di rilevare, con un certo anticipo, l'arrivo di una potenziale emergenza. Questa funzionalità è svolta dal Tool di Entity Discovery mediante la ricerca di Allerte.

Per poter agevolmente gestire il nostro Motore Semantico, come anche per tenerne sempre sotto controllo lo stato, abbiamo inoltre realizzato un Tool di Monitoraggio esterno che permette, tra le funzioni più rilevanti, di avviare/fermare il Motore, controllare i log delle esecuzioni e degli errori, visualizzare grafici storici e attuali sui livelli di tweet ricevuti e sulle segnalazioni rilevate.

Come già accennato nell'introduzione al capitolo, ciascuno di questi moduli verrà analizzato nel particolare nei paragrafi seguenti.

## Il Motore Semantico

---

Si cercherà di spiegare in modo chiaro come ogni parte è stata ideata, quali sono state le motivazioni dietro ad alcune scelte implementative, come sono poi state realizzate ed integrate e quale è il risultato finale ottenuto.

### 4.2 Architettura del Motore Semantico

L'architettura del nostro sistema è costituita essenzialmente da due blocchi funzionali, uno dei quali, a sua volta, raccoglie due processi di elaborazione che sono il core dell'intero Motore Semantico.

In particolare, con riferimento alla figura 4.1 della nostra architettura, emergono:

- due Web Service : per l'interfacciamento e l'integrazione con il Crawler di Beta80.
- Motore Semantico :
  - Classificatore : per la classificazione dei tweet.
  - Entity Discovery : per la rilevazione e segnalazione di allerte e/o emergenze.

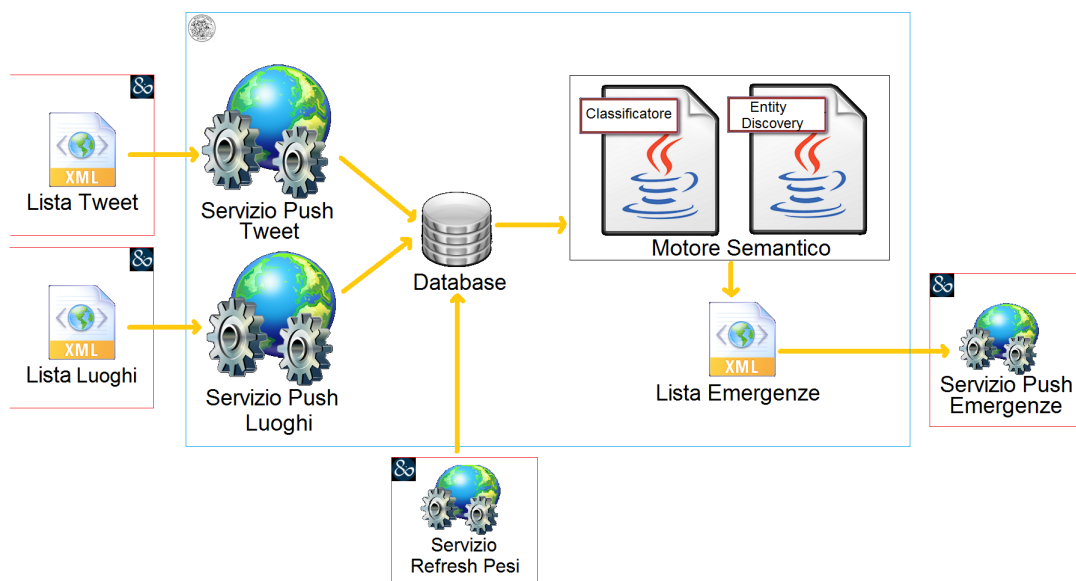


Figura 4.1: Architettura del Motore Semantico.

### 4.2.1 Il Web Service

Come già anticipato nel capitolo precedente, Beta80 raccoglie i tweet dal web tramite il proprio Crawler per poi inviarli al Motore Semantico del Politecnico.

Per realizzare questo interfacciamento è stato deciso di utilizzare un Web Service di tipo REST, che accetta in input un file XML contenente la lista di tutti i tweet raccolti in un certo arco temporale. Il servizio andrà automaticamente ad inserire i tweet nel database del Politecnico.

Oltre a quello appena descritto, abbiamo introdotto un Web Service che, con gli stessi meccanismi, consente a Beta80 di inviare al Database del Politecnico una lista di “Luoghi Notevoli”, ovvero quei punti geografici che hanno precedenza rispetto agli altri nell’analisi semantica (ne parleremo meglio nella relativa sezione 4.5.3.5).

Nella sezione 4.3 descriveremo nel dettaglio la struttura e le funzionalità di questi Web Service da noi sviluppati nel linguaggio di programmazione Java.

### 4.2.2 Database

Uno dei nostri compiti è stato anche quello di progettare e realizzare il Database di supporto al Motore Semantico.

La nostra scelta è caduta su un database MySQL, perché supportato sia da sistemi Unix che Windows. Inoltre esiste più di un driver di interfacciamento con il linguaggio Java, da noi utilizzato per tutto il progetto. Per la gestione del database abbiamo utilizzato il software MySQL Workbench.

Per la progettazione del Database abbiamo tenuto conto:

- Dei documenti CSV che Carcaci, Cesana e D’Ettorre utilizzavano come input e output nell’esecuzione dei loro algoritmi; abbiamo quindi tradotto in tabelle i file che contenevano le parole della Blacklist, le Keyword indicanti le possibili fonti di emergenze e quelli contenenti le meta-informazioni per la classificazione [5][7].
- Del formato dei documenti XML che vengono scambiati per l’interfacciamento con Beta80; in particolare abbiamo ricalcato l’XSD di quei documenti per la realizzazione della tabella dei tweet e delle segnalazioni.

## Il Motore Semantico

Queste considerazioni, unite ad alcune aggiunte dettate dall'esigenza di regolare il flusso delle informazioni e di registrare le attività svolte dal sistema, hanno portato alla progettazione del seguente database:

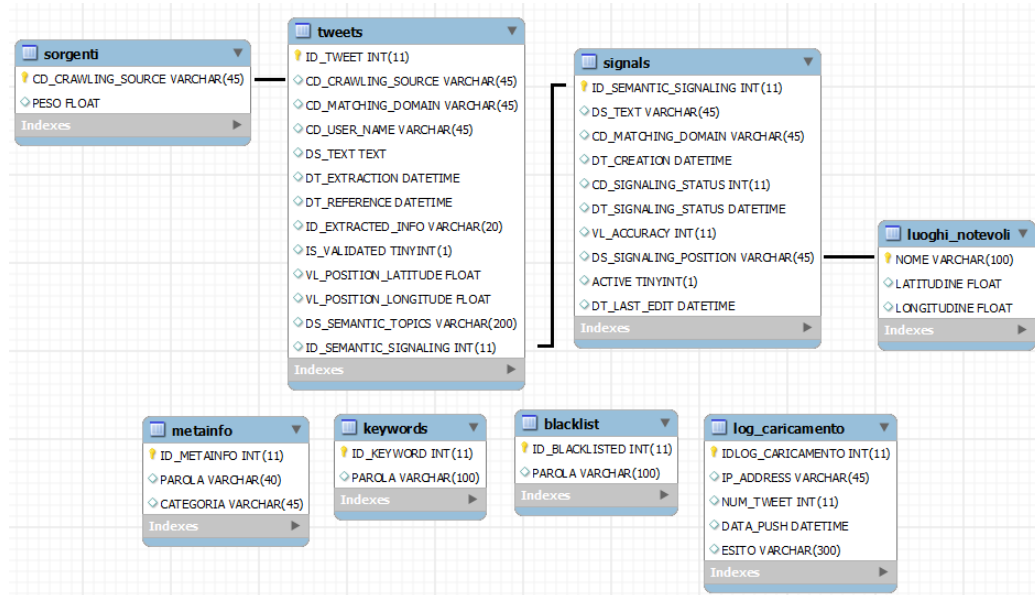


Figura 4.2: Struttura del Database del Motore Semantico.

Le tabelle *tweets* e *signals* contengono quasi solamente informazioni corrispondenti a quelle presenti nei documenti XML per l'interfacciamento tra i vari blocchi funzionali di TORCIA. Questi campi saranno descritti nel dettaglio nelle prossime sezioni, dove andremo ad illustrare l'XSD scelto in fase di analisi da Politecnico e Beta80.

Di seguito facciamo una descrizione di quelli che sono i campi contenenti quelle informazioni che invece vengono aggiunte dal nostro sistema.

ACTIVE e DT\_LAST\_EDIT sono due campi della tabella *signals* che hanno una specifica utilità interna al Motore Semantico.

In particolare, ACTIVE descrive in ogni istante lo stato di un'emergenza rilevata dal sistema, e può essere:

- ACTIVE = 0 : l'emergenza è terminata.
- ACTIVE = 1 : l'emergenza è attiva.
- ACTIVE = 2 : si tratta in realtà di un'allerta e non di un'emergenza vera e propria.

## 4.2 Architettura del Motore Semantico

---

Il campo `DT_LAST_EDIT` conterrà invece la data dell'ultima volta in cui il Motore Semantico rileva che si è parlato su twitter di quell'emergenza. La vera e propria funzionalità di questi campi e la loro evoluzione durante l'esecuzione del Motore Semantico saranno descritte nel dettaglio nella sezione 4.5 riguardante l'Entity Discovery.

La tabella *meta-info* raccoglie le categorie per la classificazione e le parole chiave associate ad ognuna delle categorie, i *Metadata*. L'utilità di questa tabella e soprattutto il delicato compito di stabilirne il contenuto saranno ampiamente illustrate nella sezione 4.4 riguardante il Classificatore.

La tabella *keywords* e la tabella *blacklist* contengono liste di parole che:

- Devono essere considerate per la rilevazione di eventuali emergenze: nel caso delle Keyword.
- NON devono essere considerate per la rilevazione di eventuali emergenze: nel caso della Blacklist.

Anche in questo caso rimandiamo la descrizione dettagliata alla sezione 4.5 sull'Entity Discovery.

La tabella *log\_caricament* costituisce, di fatto, un registro degli eventi per quando Beta80 interroga il servizio Push Tweet. Viene salvato l'indirizzo IP da cui provengono i tweet, la data, il numero di tweet salvati correttamente e l'esito dell'interfacciamento, sia esso positivo che negativo.

La tabella *luoghi\_notevoli* è stata introdotta per dare una priorità di ricerca a determinate zone di interesse. Ogni luogo è identificato dal nome e dalle sue coordinate geografiche, ed è inserito da Beta80 tramite il servizio Push Luoghi, che introdurremo nella sezione 4.3.6. Le modalità con cui questa tabella viene consultata dal Tool di Entity Discovery durante l'analisi semantica saranno descritte nel dettaglio nella sezione 4.5.

Infine la tabella *sorgenti* indica con quale peso conteggiare un certo tweet, a seconda della fonte da cui proviene. Per ora il sistema TORCIA può avvalersi solo delle due sorgenti:

## Il Motore Semantico

---

- generico utente di Twitter
- utente registrato all'app mobile

ma nel futuro nulla impedisce che nuove fonti possano essere aggiunte, ad esempio le sonde di rilevamento poste sugli argini dei fiumi. Il nostro sistema è infatti predisposto ad accettare messaggi da qualunque fonte di cui sia stato inserito anche un peso, senza alcun problema. Questa tabella viene periodicamente sincronizzata con i pesi stabiliti dall'Operatore che utilizza il Cruscotto di Beta80 che possono anche variare, quindi, in tempo reale per meglio adattarsi a una specifica emergenza.

Nel corso del nostro lavoro di tesi abbiamo potuto testare l'interfacciamento al database da un sistema Unix, da Windows e pure creando un database su CloudBees. Quest'ultimo è un PaaS (Platform as a Service) provvisoriamente utilizzato da noi per gestire, eseguire e testare tutti i servizi web, prima della migrazione sulla macchina TORCIA su cui girerà la versione definitiva del Motore Semantico.

### 4.2.3 Motore Semantico

#### 4.2.3.1 Classificatore

Il Classificatore è di fatto un software scritto in Java, che si occuperà di fornire un valore aggiunto all'informazione stessa, classificandola in diverse categorie (le Classi), sfruttando i *Metadata*, di cui si parlerà meglio nella sezione 4.4. In tal modo, l'informazione sarà consultabile con maggior facilità sia per gli operatori e gli enti preposti alla gestione delle emergenze, sia per i cittadini che successivamente la potranno esaminare all'interno dell'applicazione.

L'importanza del classificatore sta nel fatto che permette l'analisi e la comprensione dei messaggi in maniera completamente automatica, aumentando ancora di più l'efficienza dell'utilizzo di tale informazione nella gestione delle emergenze. Inoltre, esso fa da filtro, eliminando tutti i messaggi non inerenti al nostro contesto ma ugualmente prelevati dal Crawler e, quindi, inutili e dannosi per il conseguimento dei nostri obiettivi.

Il funzionamento di tale classificatore verrà spiegato in dettaglio nella sezione 4.4.

### 4.2.3.2 Entity Discovery

Un ulteriore valore aggiunto viene dato alle informazioni mediante l'utilizzo di un altro software, l'Entity Discovery. Esso, per prima cosa, tenterà di aggregare tutte le informazioni riferite allo stesso evento, analizzando i tweet, uno per uno, e rintracciando in essi un'Entity comune. Così facendo, come il Classificatore, permetterà una migliore visualizzazione delle informazioni, aumentando ancora di più il loro valore e l'importanza del loro utilizzo.

Inoltre, eseguirà una funzione di alerting: infatti esso cercherà di rilevare, nel più breve tempo possibile, nuovi eventi, in modo da notificarli immediatamente a Beta80 contattando il servizio Push Emergenze da loro implementato. A sua volta Beta80 attraverso il proprio Cruscotto e la propria App potrà alertare enti di competenza e cittadini.

Rimandiamo alla lettura della sezione 4.5 per i dettagli funzionali e implementativi di tale tool, anch'esso sviluppato con il linguaggio Java.

## 4.3 Web Service

In questa sezione andremo ad esporre nel dettaglio la struttura e le funzionalità dei due Web Service da noi creati per il progetto TORCIA, i servizi PUSH TWEET e PUSH LUOGHI.

Introdurremo il concetto di Web Service e confronteremo le due tecnologie esistenti in ambito di servizi web, ovvero SOAP e REST.

Quindi parleremo del servizio da noi sviluppato per il progetto TORCIA, soffermandoci sulla logica che ci sta dietro e sulle difficoltà riscontrate in fase di implementazione.

### 4.3.1 Cos'è un Web Service

In informatica un Web Service (servizio web), secondo la definizione data dal World Wide Web Consortium (W3C), è un sistema software progettato per

## Il Motore Semantico

---

supportare l'interoperabilità tra diversi elaboratori su di una medesima rete, ovvero in un contesto distribuito.

Tale caratteristica si ottiene associando all'applicazione un'interfaccia software (descritta in un formato automaticamente elaborabile quale, ad esempio, il Web Services Description Language) che espone all'esterno il servizio/i associato/i e utilizzando la quale altri sistemi possono interagire con l'applicazione stessa attivando le operazioni descritte nell'interfaccia (servizi o richieste di procedure remote) tramite appositi "messaggi" di richiesta. Tali messaggi di richiesta sono inclusi in una "busta" (la più famosa è SOAP), formattati secondo lo standard XML, incapsulati e trasportati tramite i protocolli del Web (solitamente HTTP), da cui appunto il nome web service.

Proprio grazie all'utilizzo di standard basati su XML, tramite un'architettura basata sui Web Service (chiamata, con terminologia inglese, Service oriented Architecture - SOA) applicazioni software scritte in diversi linguaggi di programmazione e implementate su diverse piattaforme hardware possono quindi essere utilizzate, tramite le interfacce che queste "espongono" pubblicamente e mediante l'utilizzo delle funzioni che sono in grado di effettuare (i "servizi" che mettono a disposizione), per lo scambio di informazioni e l'effettuazione di operazioni complesse (quali, ad esempio, la realizzazione di processi di business che coinvolgono più aree di una medesima azienda) sia su reti aziendali come anche su Internet.

La possibilità dell'interoperabilità fra diversi linguaggi di programmazione (ad esempio, tra Java e Python) e diversi sistemi operativi (come Windows e Linux) è resa possibile dall'uso di standard "aperti". Il consorzio OASIS (Organization for the Advancement of Structured Information Standards) ed il W3C sono i principali responsabili dell'architettura e della standardizzazione dei Web Service; per migliorare l'interoperabilità tra le diverse implementazioni dei Web Service l'organizzazione WS-I sta inoltre sviluppando una serie di "profili" per meglio definire gli standard coinvolti.

Alcuni dei vantaggi che è possibile ottenere con l'utilizzo dei Web Service sono i seguenti:



- Permettono l'interoperabilità tra diverse applicazioni software su diverse piattaforme hardware
- Utilizzano standard e protocolli “open”; i protocolli ed il formato dei dati è, ove possibile, in formato testuale, cosa che li rende di più facile comprensione ed utilizzo da parte degli sviluppatori
- Mediante l'uso di HTTP per il trasporto dei messaggi, i Web Service non necessitano, normalmente, che vengano effettuate modifiche alle regole di sicurezza utilizzate come filtro sui firewall
- Possono essere facilmente utilizzati, in combinazione l'uno con l'altro (indipendentemente da chi li fornisce e da dove vengono resi disponibili) per formare servizi “integrati” e complessi
- Consentono il riutilizzo di infrastrutture ed applicazioni già sviluppate e sono (relativamente) indipendenti da eventuali modifiche delle stesse

Di contro vi sono i seguenti aspetti da considerare:

- Attualmente non esistono standard consolidati per applicazioni critiche quali, ad esempio, le transazioni distribuite
- Le *performance* legate all'utilizzo dei Web Service possono essere minori di quelle riscontrabili utilizzando approcci alternativi di distributed computing quali Java RMI, CORBA, o DCOM
- L'uso dell'HTTP permette ai Web Service di evitare le misure di sicurezza dei firewall (le cui regole sono stabilite spesso proprio per evitare le comunicazioni fra programmi “esterni” ed “interni” al firewall).

La ragione principale per la creazione e l'utilizzo di Web Service è il “disaccoppiamento” che l'interfaccia standard, esposta dal Web Service, rende possibile fra il sistema utente ed il Web Service stesso: modifiche ad una o all'altra delle applicazioni possono essere attuate in maniera “trasparente” all'interfaccia tra i due sistemi; tale flessibilità consente la creazione di sistemi software complessi costituiti da componenti svincolati l'uno dall'altro e consente una forte riusabilità di codice ed applicazioni già sviluppate.

Proprio per questo disaccoppiamento i partner di TORCIA hanno scelto di realizzare i propri blocchi funzionali appoggiandosi su database separati,

per poi scambiarsi informazioni attraverso i servizi web.

I Web service hanno inoltre guadagnato consensi visto che, come protocollo di trasporto, possono utilizzare HTTP “over” TCP sulla porta 80; tale porta è, normalmente, una delle poche (se non l’unica) lasciata “aperta” dai sistemi firewall al traffico di entrata ed uscita dall’esterno verso i sistemi aziendali poiché, su tale porta, transita il traffico HTTP dei web browser: ciò consente l’utilizzo dei Web Service senza modifiche sulle configurazioni di sicurezza dell’azienda (un aspetto che se da un lato è positivo, solleva però preoccupazioni concernenti la sicurezza).

Un’ultima ragione che ha favorito l’adozione ed il proliferare dei Web Service è la mancanza, prima dello sviluppo di SOAP, di interfacce realmente pratiche per l’utilizzo di funzionalità distribuite in rete: EDI, RPC, ed altri tipi di API (Application Programming Interface) erano e rimangono meno conosciute e di più difficile utilizzo rispetto all’architettura dei Web Service.

### 4.3.2 SOAP e REST

Allo stato attuale esistono due approcci alla creazione di Web Service:

- Un approccio è basato sul protocollo standard SOAP (*Simple Object Access Protocol*) che, per lo scambio di messaggi per l’invocazione di servizi remoti, si prefigge di riprodurre in ambito Web un approccio a chiamate remote, *Remote Procedure Call*, tipico di protocolli di interoperabilità come CORBA, DCOM e RMI.
- Un secondo approccio è ispirato ai principi architettureali tipici del Web e si concentra sulla descrizione di risorse, sul modo di individuarle nel Web e sul modo di trasferirle da una macchina all’altra. Questo approccio prende il nome di REST (*REpresentational State Transfer*).

Anche se l’obiettivo dei due approcci è pressoché identico, cioè l’adozione del Web come piattaforma di elaborazione, la loro visione e la soluzione suggerita sono totalmente differenti. Risulta utile fare un confronto tra i due approcci per comprendere queste differenze.

La prima evidente differenza tra i due tipi di Web Service è la visione del Web proposta come piattaforma di elaborazione. REST propone una visione del Web incentrata sul concetto di risorsa mentre i SOAP Web Service mettono in risalto il concetto di servizio.

- Un Web Service RESTful è custode di un insieme di risorse sulle quali un client può chiedere le operazioni canoniche del protocollo HTTP
- Un Web Service basato su SOAP espone un insieme di metodi richiamabili da remoto da parte di un client

L'approccio dei SOAP Web service ha mutuato un'architettura applicativa denominata SOA, *Service Oriented Architecture*, a cui si è recentemente contrapposta l'architettura ROA, *Resource Oriented Architecture*, ispirata ai principi REST.

Il protocollo SOAP definisce una struttura dati per lo scambio di messaggi tra applicazioni, riproponendo, in un certo senso, parte di quello che il protocollo HTTP faceva già. SOAP utilizza HTTP come protocollo di trasporto, ma non è limitato nè vincolato ad esso, dal momento che può benissimo usare altri protocolli di trasporto.

A differenza di HTTP, però, le specifiche di SOAP non affrontano argomenti come la sicurezza o l'indirizzamento, per i quali sono stati definiti standard a parte, nello specifico *WS-Security* e *WS-Addressing*.

È evidente che l'approccio adottato dai Web Service basati su SOAP è derivato dalle tecnologie di interoperabilità esistenti al di fuori del Web e basato essenzialmente su chiamate di procedura remota, come DCOM, CORBA e RMI.

In sostanza questo approccio può essere visto come una sorta di adattamento di queste tecnologie al Web. SOAP non sfrutta a pieno il protocollo HTTP, utilizzandolo come semplice protocollo di trasporto.

REST invece sfrutta HTTP per quello che è, un protocollo di livello applicativo, e ne utilizza a pieno le potenzialità.

L'approccio REST tende a conservare e ad esaltare le caratteristiche intrinseche del Web, evidenziandone la predisposizione ad essere una piat-

## Il Motore Semantico

---

taforma per l'elaborazione distribuita. Quindi, non è necessario aggiungere nulla a quanto è già esistente sul Web per consentire ad applicazioni remote di interagire.

Inoltre i Web Service basati su SOAP prevedono lo standard WSDL, *Web Service Description Language*, per definire l'interfaccia di un servizio. Questa è un'ulteriore evidenza del tentativo di adattare al Web l'approccio di interoperabilità basato su chiamate remote. Infatti il WSDL non è altro che un IDL (*Interface Description Language*) per un componente software.

Da un lato l'esistenza di WSDL favorisce l'uso di tool per creare automaticamente client in un determinato linguaggio di programmazione, ma allo stesso tempo induce a creare una forte dipendenza tra client e server.

REST non prevede esplicitamente nessuna modalità per descrivere come interagire con una risorsa. Le operazioni sono implicite nel protocollo HTTP. Qualcosa di analogo a WSDL è WADL, *Web Application Definition Language*, un'applicazione XML per definire risorse, operazioni ed eccezioni previsti da un Web Service di tipo REST.

WADL è stato sottoposto al W3C per la standardizzazione nel 2009, ma allo stato attuale non ci sono piani per la sua discussione ed eventuale approvazione. In realtà esso non ha avuto una accoglienza molto favorevole da parte della comunità REST, in quanto offre una visione statica di un Web Service, contraddicendo il principio HATEOAS che pone nella presenza di collegamenti all'interno della rappresentazione di una risorsa la definizione di un contratto con il client, con una visione quindi molto più dinamica e un debole accoppiamento tra client e server.

In conclusione, i Web service basati su SOAP costruiscono un'infrastruttura prolissa e complessa al di sopra del Web, per fare cose che il Web è già in grado di fare. Il vantaggio di questo tipo di servizi è che in realtà definisce uno standard indipendente dal Web e l'infrastruttura può essere basata anche su protocolli diversi.

REST invece intende ripristinare il Web ad architettura per la programmazione distribuita, senza aggiungere sovrastrutture non necessarie.

Nell'ambito del progetto TORCIA, i vari partner hanno deciso in fase di analisi che l'integrazione tra i diversi blocchi funzionali verrà implementata con dei Web Service di tipo REST.

### 4.3.3 Architettura di PUSH TWEET e PUSH LUOGHI

Nella Figura 4.3 è mostrato il dettaglio dell'architettura dei servizi da noi sviluppato, già anticipato nella Figura 4.1 dell'architettura generale del Motore Semantico.

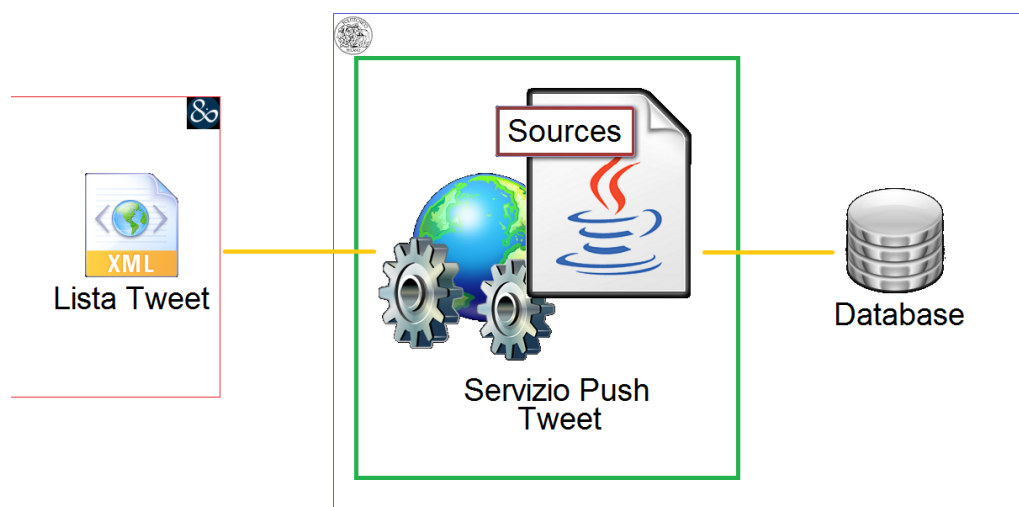


Figura 4.3: Architettura del servizio PUSH TWEET.

Il deploy di entrambi è eseguito sul contenitore servlet Apache Tomcat. Quest'ultimo fornisce una piattaforma software per l'esecuzione di applicazioni web sviluppate in Java, includendo nella sua distribuzione standard anche le funzionalità di Web Server tradizionale. I servizi sono stati realizzati nel linguaggio Java e, grazie alle funzionalità di Apache Tomcat, potranno essere sempre in ascolto per ricevere dati in input.

Per quanto riguarda i tweet, i dati provengono dal Crawler di Beta80 all'interno di un documento XML, in un formato ben preciso che descriveremo meglio nella prossima sezione ??.

L'invio dei tweet da parte di Beta80 avviene quando accade la prima delle seguenti situazioni:

- È passato un limite massimo di tempo dall'ultimo invio

## Il Motore Semantico

---

- La finestra di tweet messi in coda dal Crawler è piena

I valori di tempo massimo e dimensione della coda sono stabiliti da Beta80, ma il nostro servizio è sempre in ascolto e pronto a ricevere, quindi non è rilevante per noi conoscerli.

Per il servizio PUSH LUOGHI invece, i dati possono provenire non solo da Beta80, ma anche e soprattutto dai cartografici. Sono questi ultimi infatti i più indicati e interessati a riempire il Database con nuovi luoghi e geolocalizzazioni specifiche, per rendere il più efficiente e semplice possibile il loro compito di creazione delle mappe.

Anche in questo caso le informazioni sono fornite al nostro servizio per mezzo di un documento XML, in un formato prestabilito che verrà descritto nella sezione 4.3.6.

In entrambi i casi, i dati vengono quindi processati dal vero e proprio software Java, che costituisce il core del Web Service, per poi essere salvati all'interno del Database del Politecnico. Una volta inserite le informazioni nel database, i servizi torneranno in ascolto, pronti per le prossime chiamate.

### 4.3.4 Funzionalità del servizio PUSH TWEET

Il servizio riceve in input un documento XML che deve rispettare le regole del documento XSD mostrato in Figura 4.4:

```

<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xs="http://www.w3.org/2001/XMLSchema" attributeFormDefault="unqualified" elementFormDefault="qualified">

  <xs:element name="EXTRACTED_INFO_LIST">
    <xs:complexType>
      <xs:sequence>

        <xs:element name="EXTRACTED_INFO" maxOccurs="unbounded" minOccurs="0">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="CD_CRAWLING_SOURCE" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="CD_MATCHING_DOMAIN" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="CD_USER_NAME" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="DS_TEXT" type="xs:string" minOccurs="0" maxOccurs="1" nillable="false" />
              <xs:element name="DT_EXTRACTION" type="xs:dateTime" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="DT_REFERENCE" type="xs:dateTime" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="ID_EXTRACTED_INFO" type="xs:unsignedShort" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="IS_VALIDATED" type="xs:unsignedByte" minOccurs="1" maxOccurs="1" nillable="false" />
              <xs:element name="VL_POSITION_LATITUDE" type="xs:decimal" minOccurs="0" maxOccurs="1" nillable="false" />
              <xs:element name="VL_POSITION_LONGITUDE" type="xs:decimal" minOccurs="0" maxOccurs="1" nillable="false" />
              <xs:element name="DS_SEMANTIC_TOPICS" type="xs:string" minOccurs="0" maxOccurs="1" nillable="false" />
            </xs:sequence>
          </xs:complexType>
        </xs:element>

      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>

```

Figura 4.4: Struttura XSD per la trasmissione dei Tweet.

Di seguito riportiamo l'esempio di un possibile documento XML di input, composto da due tweet (il secondo dei quali geolocalizzato):

```

<EXTRACTED_INFO_LIST>
<EXTRACTED_INFO>
  <CD_CRAWLING_SOURCE>TWITTER</CD_CRAWLING_SOURCE>
  <CD_MATCHING_DOMAIN>ALLAGAMENTO</CD_MATCHING_DOMAIN>
  <CD_USER_NAME>chiccokb24</CD_USER_NAME>
  <DS_TEXT>allagamento in viale Filippo Turati, Lecco</DS_TEXT>
  <DT_EXTRACTION>2014-01-07T10:00:19</DT_EXTRACTION>
  <DT_REFERENCE>2014-01-07T10:00:19</DT_REFERENCE>
  <ID_EXTRACTED_INFO>297</ID_EXTRACTED_INFO>
  <IS_VALIDATED>1</IS_VALIDATED>
</EXTRACTED_INFO>
<EXTRACTED_INFO>
  <CD_CRAWLING_SOURCE>TWITTER</CD_CRAWLING_SOURCE>
  <CD_MATCHING_DOMAIN>ALLAGAMENTO</CD_MATCHING_DOMAIN>
  <CD_USER_NAME>ciummi89</CD_USER_NAME>
  <DS_TEXT>Aiuto! E' allagato viale Filippo Turati a Lecco!</DS_TEXT>
  <DT_EXTRACTION>2014-01-07T10:01:19</DT_EXTRACTION>
  <DT_REFERENCE>2014-01-07T10:01:19</DT_REFERENCE>
  <ID_EXTRACTED_INFO>298</ID_EXTRACTED_INFO>
  <IS_VALIDATED>1</IS_VALIDATED>
  <VL_POSITION_LATITUDE>45.860943</VL_POSITION_LATITUDE>
  <VL_POSITION_LONGITUDE>9.388724</VL_POSITION_LONGITUDE>
</EXTRACTED_INFO>
</EXTRACTED_INFO_LIST>

```

Figura 4.5: Esempio valido di un XML contenente due Tweet.

Il Servizio Push Tweet, una volta ricevuto il documento, lo apre e ne estrae ogni elemento:

## Il Motore Semantico

---

- `CD_CRAWLING_SOURCE` : indica qual è la sorgente di quell'informazione; nel caso di TORCIA sono stati scelti inizialmente solo Twitter ed il futuro applicativo mobile di Beta80.
- `CD_MATCHING_DOMAIN` : è il dominio di applicazione dell'informazione. Ricordiamo che Beta80 preleva dalla rete dei tweet che appartengono ad un determinato contesto, o meglio che contengano determinate parole chiave che li fanno risalire al contesto in esame (es. ALLAGAMENTO, NUBIFRAGIO). Sarà poi compito del Motore Semantico capire se il tweet in questione è effettivamente relativo ad un'emergenza oppure no.
- `CD_USER_NAME` : identificativo dell'utente di twitter che ha cinguettato.
- `DS_TEXT` : il testo del tweet.
- `DT_EXTRACTION` : la data e l'ora in cui Beta80 ha prelevato il tweet dalla rete.
- `DT_REFERENCE` : la data e l'ora in cui il tweet è stato digitato.
- `ID_EXTRACTED_INFO` : è il numero che identifica il tweet tra i vari partner TORCIA. E' possibile che i singoli partner diano ai tweet degli ID interni univoci nel loro database, ma quando l'informazione deve circolare tra tutti gli attori allora deve essere identificata mediante questo campo.
- `IS_VALIDATED` : indica se l'informazione che è stata ricevuta è validata o meno dall'ipotetico Operatore che lavora sul Cruscotto di Beta80; questo valore non deve essere in alcun modo da noi modificato.
- `VL_POSITION_LATITUDE` e `VL_POSITION_LONGITUDE` : sono le coordinate date dalla geolocalizzazione del tweet. Questi sono campi opzionali, in quanto non tutti i tweet contengono una propria geolocalizzazione quando vengono postati. Per sopperire all'assenza di queste coordinate è stata introdotta la funzionalità di Geolocalizzazione Semantica del Motore Semantico, che verrà descritta meglio nella sezione 4.5 riguardante l'Entity Discovery.



- DS\_SEMANTIC\_TOPICS : questo campo non verrà mai inviato da Beta80, in quanto sarà compito del Motore Semantico calcolare le Classi da inserirvi, nel formato “classe1,classe2,...” (tutto ciò sarà spiegato meglio nella sezione 4.4 relativa al Classificatore). Nonostante questo però, il formato XML deve già contenere l’indicazione della sua esistenza, poiché l’XSD dei tweet è definito unico e fisso per ogni parte di progetto.

Come già citato nella descrizione generale dell’architettura del Motore Semantico, il database che abbiamo progettato ricalca nei suoi campi quelli che sono i nomi degli elementi ricevuti nel file XML, come si può notare anche dalla Figura 4.6 sottostante.

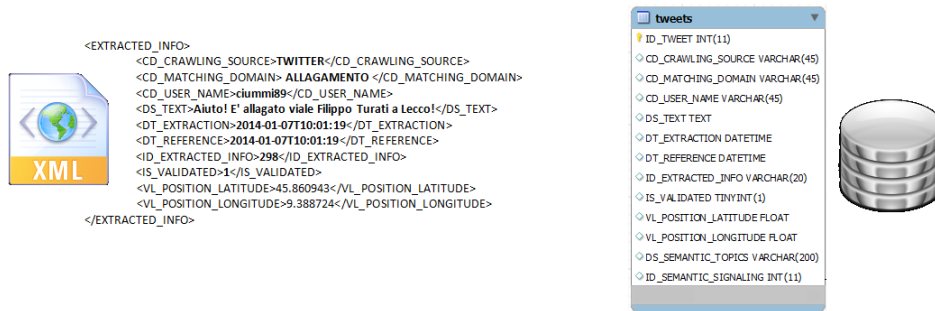


Figura 4.6: Confronto tra XML di un tweet e tabella tweets del Database.

La funzionalità del Servizio Push Tweet è quindi quella di aprire il documento XML, prendere i dati relativi ad ogni tweet ed inserirli nel Database del Politecnico.

Nell’eseguire il suo compito il servizio restituisce una risposta a Beta80 sottoforma di documento XML.

Esistono tre tipi di output che il nostro servizio può restituire:

1. Presenza di un errore che indica un utilizzo scorretto del formato del file XML ricevuto (nell’esempio si può intuire l’assenza di uno dei tag di chiusura).

```
<>xml version='1.0' encoding='UTF-8'>>
<POLIMI_RESPONSE>
  <ERROR_MESSAGE>The element type "EXTRACTED_INFO" must be terminated by the matching end-tag "</EXTRACTED_INFO".".
</ERROR_MESSAGE>
</POLIMI_RESPONSE>
```

In queste situazioni il servizio respinge la richiesta di Push dei Tweet, non inserendo nulla nel database.

## Il Motore Semantico

---

2. Il servizio ha riscontrato dei problemi nel momento dell'esecuzione della query per l'inserimento dei dati nel Database. Ciò vuol dire che il formato XML è valido, ma alcuni elementi contenevano caratteri non compatibili con il corrispondente campo del Database (nell'esempio c'erano delle lettere nell'elemento DT\_EXTRACTION).

```
<?xml version='1.0' encoding='UTF-8'?>
<POLIMI_RESPONSE>
  <RESPONSE_DESCRIPTION>Caricati 1 tweet</RESPONSE_DESCRIPTION>
  <sql_error_list>
    <sql error>SQL Exception: Data truncation: Incorrect datetime value: 'aaa' for column 'DT_EXTRACTION' at row 1</sql error>
    <sql error>SQL Exception: Data truncation: Incorrect datetime value: 'aaa' for column 'DT_EXTRACTION' at row 1</sql error>
  </sql_error_list>
</POLIMI_RESPONSE>
```

In questi casi il servizio non esegue l'inserimento del tweet che ha dato problemi, ma continua con l'inserimento degli altri, riportando il numero totale dei tweet caricati.

3. Come si può facilmente intuire dall'immagine, in questo caso l'operazione di Push è andata a buon fine.

```
<?xml version='1.0' encoding='UTF-8'?>
<POLIMI_RESPONSE>
  <RESPONSE_DESCRIPTION>Caricati 3 tweet</RESPONSE_DESCRIPTION>
</POLIMI_RESPONSE>
```

Il servizio è riuscito ad aprire correttamente il documento XML e ad inserire i dati nel Database del Politecnico. La risposta contiene il numero totale dei tweet che sono stati ricevuti e memorizzati.

L'esito di ogni chiamata al Servizio Push Tweet viene memorizzato nell'apposita tabella del Database interno del Politecnico chiamata *log\_caricamento*. In questo modo, oltre all'esito dell'operazione, vengono memorizzate anche la data e l'indirizzo IP di chi ha fatto la richiesta.

### 4.3.5 Sviluppo

Prima di essere attivato sul server TORCIA, il Servizio è stato provato sul PaaS CloudBees. In questo modo è stato possibile testare i tempi di risposta quando esso viene chiamato in remoto.

L'esito del test è stato da noi considerato positivo, in quanto per caricare 1000 tweet il servizio ha impiegato circa 1 secondo. Si aggiunga che il

Crawler di Beta80 risiederà sullo stesso server TORCIA in cui risiederanno il Web Service e il Motore Semantico e che, a regime, verranno integrati in locale. Tutto questo permetterà un'ulteriore riduzione dei tempi di esecuzione.

Il fatto che il Crawler di Beta80 e il Database del Politecnico risiederanno sullo stesso server sembra rendere inutile la presenza del Servizio da noi implementato ma, come già espresso nelle sezioni precedenti, il progetto TORCIA prevede un disaccoppiamento dei blocchi funzionali. Tale disaccoppiamento ha portato a scegliere di realizzare il Web Server anche per la comunicazione tra blocchi residenti sulla stessa macchina. Inoltre se nel futuro si volessero disaccoppiare anche fisicamente le parti che compongono TORCIA, il nostro Motore Semantico sarebbe già pronto.

La più grande difficoltà incontrata in fase di implementazione è stata sicuramente quella dovuta alla presenza di caratteri speciali nei tweet. Per un qualsiasi sistema informatico risulta critico e fondamentale filtrare gli input degli utenti. Per la maggior parte dei software è possibile dare dei limiti ai caratteri utilizzabili o al formato del testo inserito, ma nel nostro caso l'input è costituito da una serie di tweet scritti da persone potenzialmente provenienti da ogni parte del mondo e per questo non controllabili a priori.

Quando il Servizio Push Tweet va ad aprire il documento XML ricevuto da Beta80, deve riuscire a decodificare tutti caratteri presenti al suo interno. Per questo è stato scelto di utilizzare lo standard di codifica ISO-8859-1 (anzichè UTF-8) per il parsing dell'XML di input. ISO-8859-1 permette infatti la conversione di un maggior numero di caratteri, tra i quali, soprattutto, le vocali accentate, non supportate invece da UTF-8.

### 4.3.6 Funzionalità del servizio PUSH LUOGHI

Il servizio riceve in input un documento XML che deve rispettare le regole del documento XSD mostrato in Figura 4.7 :

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:xsd="http://www.w3.org/2001/XMLSchema"
xmlns:xs="http://www.w3.org/2001/XMLSchema" attributeFormDefault="unqualified" elementFormDefault="qualified">

<xs:element name="LISTA_LUOGHI">
  <xs:complexType>
    <xs:sequence>

      <xs:element name="LUOGO" maxOccurs="unbounded" minOccurs="1">
        <xs:complexType>
          <xs:sequence>

            <xs:element name="NOME" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false" />
            <xs:element name="LATITUDINE" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false" />
            <xs:element name="LONGITUDINE" type="xs:string" minOccurs="1" maxOccurs="1" nillable="false" />

          </xs:sequence>
        </xs:complexType>
      </xs:element>

    </xs:sequence>
  </xs:complexType>
</xs:element>

</xs:schema>
```

Figura 4.7: Struttura XSD per la trasmissione dei Luoghi Notevoli.

Come facilmente intuibile, la funzionalità di questo servizio è quella di permettere a Beta80 di inserire nel Database del Politecnico una lista di Luoghi Notevoli. Per ognuno dei luoghi dovrà essere indicato il NOME, la LATITUDINE e la LONGITUDINE.

Il servizio riprende in generale la struttura di quello per il push dei tweet descritto nella sezione precedente, soprattutto per quanto riguarda gli standard di codifica utilizzati e le possibili risposte.

Le modalità con cui questi luoghi notevoli entrano in gioco durante l'analisi semantica verrà descritta ampiamente nel capitolo 4.5 sull'Entity Discovery.

## 4.4 Motore Semantico: Classificatore

Si tratta di un algoritmo sintattico che esegue una categorizzazione dei messaggi prelevati dai social media mediante un'ontologia appositamente creata. Tale procedura è stata ideata da Claudio Carcaci ed è da noi sfruttata per la realizzazione del Classificatore del sistema TORCIA [5]. Al fine di adattare il Tool di Classificazione alle nostre esigenze abbiamo apportato alcune modifiche ma, prima di procedere, richiamiamo gli elementi fondamentali di tale algoritmo.

### 4.4.1 Algoritmo di Carcaci

Secondo quanto detto dall'autore nel suo elaborato di tesi “*l'applicativo realizzato ha come obiettivo quello di fornire un'analisi guidata totalmente dalla sintassi che permetta di classificare messaggi, post, tweet provenienti dai social network*”. L'analisi sintattica proposta è un processo volto ad analizzare uno stream continuo in input, in modo da determinarne la struttura grammaticale grazie all'utilizzo di una grammatica formale e una conseguente costruzione di un albero sintattico.

Le entità di riferimento su cui l'algoritmo agisce sono:

- testi
- termini
- classi

Quello che la procedura cerca di fare è di delineare le loro interrelazioni, secondo lo schema riportato in Figura 4.8:

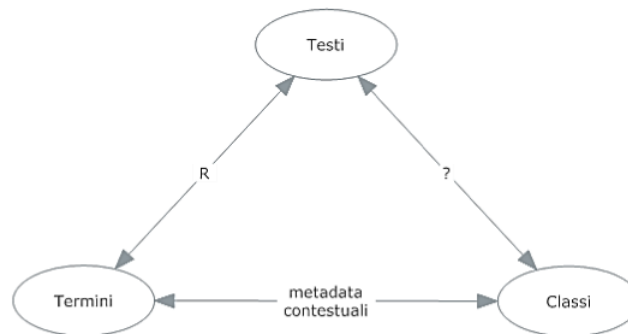


Figura 4.8: Relazione tra le entità nell'algoritmo di classificazione.

Nel dettaglio, abbiamo:

**Relazione tra testi e termini.** Viene espressa mediante una matrice di rilevanza nel quale vengono inseriti dei valori, corrispondenti all'importanza che un dato termine ha all'interno del testo.

**Relazione tra termini e classi.** Questa relazione è definibile mediante i *Metadata*, preventivamente definiti, che identificano una classe. I *Metadata* sono i termini che, opportunamente correlati all'interno di un testo, permettono al Classificatore di assegnare una o più categorie alla frase in esame.

**Relazione tra testi e classi.** Si tratta di ciò che l'algoritmo deve riuscire a definire, l'obiettivo ultimo che si pone.

In sostanza, quello che l'algoritmo intende fare è suddividere i messaggi in diverse categorie, le classi, sulla base di termini, metadata, presenti in essi.

### 4.4.2 Modifiche alle Classi

Ai fini del progetto TORCIA sono state identificate 7 classi caratterizzanti. Esse sono state costruite da Andrea Palla nel proprio lavoro di tesi e sono prettamente pensate per il contesto di analisi da noi utilizzato [14]. Le classi, con alcuni metadata caratterizzanti, sono le seguenti:

Categorie	<i>Metadata</i> che le compongono
Dove	Città, Centro, Via, Metro, Nord, Spagna, Venezia, Posto, Località, Zona
Segnalazione	Allagamenti, Pioggia, Esondato, Fiume, Acqua, Tempesta, Nubifragio
Conseguenze	Morti, Salvo, Frana, Evacuata, Inghiottiti, Situazione, Danni, Crollata
Responsabilità	Polizia, Magistrato, Pattuglia, Autopompa, Ministri, Comune, Tilt, Capo
Viabilità	Libera, Percorribile, Ripulita, Sgombra, Chiusa, Traffico, Percorsi, Frana
Warning	Fogne, Intasato, Argini, Pericoloso, Previsti, Rischio, Allerta, Flusso
Post	Risarcimento, Lamentela, Soldi, Assegni, Fondi, Denunce, Interventi

Tabella 4.1: Composizione delle categorie pre modifica.

Rimandiamo alla consultazione della tesi di Andrea Palla per tutti i dettagli sulla metodologia adottata per la costruzione di tali classi e per la definizione dei metadata ad esse associati [14].

Le modifiche apportate riguardano soprattutto l'introduzione di nuovi metadata, col fine di ottimizzare i valori delle metriche adottate per la valutazione della qualità del classificatore.

Esse si baseranno sulla definizione della "tabella di contingenza", nella quale si associano le decisioni prese dall'algoritmo alla corretta classificazione desiderata (Tabella 4.2).

#### 4.4 Motore Semantico: Classificatore

	Realtà: appartiene	Realtà: non appartiene
Algoritmo: appartiene	a (veropositivo)	b (falsopositivo)
Algoritmo: non appartiene	c (falsonegativo)	d (veronegativo)

Tabella 4.2: Tabella di contingenza.

In questo modo, data una classe e un test set, è possibile contare il numero di testi classificati/non classificati correttamente, conoscendo la reale appartenenza o meno ad una classe e ottenendo i quattro valori numerici: a, b, c, d in  $\mathbb{N}^+$ .

A partire da questi valori è possibile poi definire le metriche per la valutazione:

- precisione: è il numero testi classificati correttamente sul numero di testi classificati come appartenenti alla classe

$$P = \frac{a}{a + b} \in [0, 1]$$

- recall: è il numero di testi classificati correttamente sul numero di testi che appartengono nella realtà alla classe

$$R = \frac{a}{a + c} \in [0, 1]$$

- errore: è il numero dei testi erroneamente classificati sul numero totale di testi classificati

$$E = \frac{b + c}{a + b + c + d} \in [0, 1]$$

L'obiettivo principale del processo di miglioramento è quello di modificare i termini che compongono le classi, i metadata, per migliorare i parametri di valutazione.

In particolare il processo si articola in cinque fasi e assumerà un andamento ciclico fino a quando i risultati ottenuti non saranno soddisfacenti:

## Il Motore Semantico

---

1. Calcolo delle metriche di valutazione di partenza per il test set;
2. Analisi manuale dei messaggi contenuti nel test set eseguito specificatamente per ogni classe;
3. Modifica dei metadata, mediante l'aggiunta o la rimozione di alcuni termini;
4. Calcolo delle metriche di valutazione post modifica per il test set;
5. Calcolo delle metriche di valutazione post modifica per un test set diverso rispetto a quello utilizzato per effettuare la modifica e confronto dei valori;

I passaggi centrali di questo processo sono l'inserimento di nuovi metadata utili e la rimozione di quelli dannosi. Si rende necessaria un'analisi accurata dei termini che caratterizzano i messaggi appartenenti alle classi, per capire cosa debba essere aggiunto e cosa invece vada escluso in quanto poco pertinente o fuorviante.

E' importante, durante questo processo, tenere in considerazione che si possa verificare il fenomeno dell'overfitting, cioè che il modello delle classi costruito si adatti eccessivamente ai dati utilizzati per definirlo, perdendo la sua generalità. Pertanto è necessario, come indicato nel punto 5, testare tale modello su un altro campione di dati, al fine di calcolarne le metriche di valutazione ed effettuare le opportune considerazioni sui risultati ottenuti, indipendentemente dai primi valori e dati da cui si è partiti.

### 4.4.3 Risultati Ottenuti

Tutte le migliorie effettuate seguono l'iter sopra descritto. Di seguito ne presentiamo un esempio realizzato utilizzando un campione di dati di 5000 tweet, relativi alla specifica segnalazione del Nubifragio di Catania di Febbraio 2013, di cui parleremo meglio in seguito nel capitolo 5.3 sul testing:

#### **Calcolo metriche di valutazione di partenza per il test set**

Come primo passo sono state calcolate le metriche di valutazione di partenza.



## 4.4 Motore Semantico: Classificatore

Nel caso specifico, come si vede nella tabella in Tabella 4.3, i valori di partenza mostrano risultati scarsi, soprattutto per quanto riguarda il parametro di precisione, sintomo di una poca generalità del modello costruito.

Metrica	Valore
Precisione	0.437719579381041
Recall	0.741342602899215
Errore	0.276563128491621

Tabella 4.3: Valori delle metriche prima delle modifiche.

### **Analisi manuale dei messaggi contenuti nel test set e modifica dei metadata**

Nella fase successiva è stata effettuata un'analisi molto accurata dei termini contenuti nei messaggi, valutando una classe alla volta. Sono stati presi in considerazione sia tutti i messaggi che secondo la corretta classificazione appartengono alla classe ma non secondo il tool, sia quelli che il tool classifica erroneamente come appartenenti ad essa.

Alcuni esempi significativi:

*Classe valutata: "Segnalazione".*

*Considerazioni effettuate: esistono molti tweet contenenti la parola "frana" e "tempesta" che il tool classifica erroneamente come appartenenti a tale classe.*

*Esempi significativi : sono tutti i tweet contenenti modi di dire come "sono una frana", "frana la terra sotto i piedi", "tempesta d'amore", "la quiete dopo la tempesta".*

*Modifiche apportate ai metadata: eliminazione dei termini "frana", "tempesta".*

*Classe valutata: "Post".*

*Considerazioni effettuate: ci sono alcuni tweet contenenti parole come "crisi", "indagati", "interpellanza", "euro" che dovrebbero appartenere a tali classe ma che il tool non classifica come tali.*

*Esempi significativi : tweet come "dopo il danno la beffa: #Lipari, disdette e #crisi dopo l'#alluvione", "cantiere ostruì le fogne, 5 indagati: Gip chiede incidente probatorio su danni durante il nubifragio", "#alluvione a #Lipari*

## Il Motore Semantico

---

*chiesto stato di calamità naturale, #danni per 30 milioni di euro, pochi ne parlano”.*

*Modifiche apportate ai metadata: aggiunta dei termini “crisi”, “disdette”, “indagato/a/i”, “interpellanza”, “euro”.*

### **Calcolo metriche di valutazione post modifica per il test set**

A questo punto sono state confrontate le metriche di valutazione iniziali con quelle post modifica per trarre le opportune considerazioni. I nuovi valori sono indicati nella tabella di Tabella 4.4.

Metrica	Valore
Precisione	0.79630522336574
Recall	0.768389271617802
Errore	0.060492139439508

Tabella 4.4: Valori delle metriche dopo le modifiche.

I miglioramenti sono palesi in tutte le categorie, Precisione soprattutto. Il modello così modificato risulta essere più Robusto.

### **Calcolo delle metriche di valutazione post modifica per un test set diverso rispetto a quello utilizzato per effettuare la modifica**

L'ultimo controllo consiste nel valutare la generalità del nuovo modello, applicandolo ad un test set diverso da quello utilizzato in partenza. A tal proposito, è stato fatto classificare al tool un insieme di 1000 tweet diversi, ottenendo i valori indicati nella tabella seguente per le metriche di valutazione.

Metrica	Valore
Precisione	0.6920932183792426
Recall	0.8569183727726524
Errore	0.0860829493087559

Tabella 4.5: Valori ottenuti per le metriche usando un diverso Test Set.

I valori ottenuti sono ovviamente un po' più bassi di quelli calcolati sul test set specifico, ma risultano essere comunque validi e indicatori di una buona generalità dell'algoritmo adottato.

### 4.4.4 Riprogettazione del Tool di Classificazione

Il tool di classificazione e il suo relativo algoritmo sono stati creati da Claudio Carcaci con scopi diversi rispetto a quelli per cui viene impiegato all'interno di TORCIA. Pertanto si sono rese necessarie alcune modifiche importanti, che ne hanno permesso una specifica personalizzazione [5].

In particolare due sono stati i cambiamenti apportati per compensare altrettanti aspetti fortemente correlati con il tipo di informazioni da classificare:

1. Poichè Twitter fornisce la funzionalità di retweet, il primo cambiamento effettuato è stato quello di far sì che il tool non li scartasse, ma che li classificasse come un qualunque messaggio normale. Questo aspetto è poi fondamentale per il corretto funzionamento dell'Entity Discovery, di cui parleremo in seguito
2. Il tool, oltre ad eseguire la funzione di classificazione, effettua anche un filtraggio di messaggi. In altre parole, esso scarta i messaggi che non rientrano in nessuna delle categorie, cioè i tweet che non hanno al loro interno termini che dimostrano la loro appartenenza ad una specifica classe. I termini che caratterizzano la categoria "Dove" sono termini molto generici e, per tal motivo, spesso alcuni messaggi vengono classificati come appartenenti a tale classe, anche se non coerenti con il nostro contesto specifico. Per questo, l'algoritmo è stato modificato in modo che vengano scartati i messaggi che, nonostante rientrino nella classe "Dove", non sono associati ad alcuna altra categoria. Nonostante possa apparire una modifica drastica, i miglioramenti, visibili dalle metriche di valutazione del tool, hanno dimostrato la sua efficacia.

### 4.4.5 Considerazioni sul Tool di Classificazione

L'importanza del Classificatore non è da sottovalutare, in quanto può contribuire notevolmente alla fruizione del servizio, sia in fase di emergenza che in fase di post-emergenza.

L'applicativo mobile che verrà reso disponibile da Beta80 alla fine del progetto infatti, prevede l'inserimento di una funzionalità di filtraggio dei tweet legati a una o più segnalazioni. Un utente avrà quindi la possibilità di scegliere

se selezionare solo i tweet di una certa classe, per avere delle informazioni più specifiche riguardo una precisa segnalazione o a tutte quelle attive al momento.

Ma il contributo forse maggiore di questo Tool sorge nella fase di post-emergenza, quando i dati raccolti durante la crisi possono essere analizzati mediante tecniche di Business Intelligence, per capire, tra le altre cose, come la situazione è stata gestita, dove ci sono stati i problemi maggiori e come fare, in futuro, opera di prevenzione. Tutto ciò è permesso grazie alle analisi di classi specifiche, come quella di “Responsabilità”, in cui si trovano, potenzialmente, tutti i tweet che un autorità potrebbe aver interesse di valutare e d esaminare.

Come mostrato in questa sezione, le performance del Classificatore sono molto buone, ma un opera di raffinamento del Database che lo ampli con nuovi termini chiave o elimini alcuni di quelli presenti ma nocivi, come mostrato negli esempi, è cruciale per un futuro miglioramento di questo Tool e quindi non solo consigliabile, ma anche necessaria.

Passiamo ora alla trattazione di quello che è il componente principale e centrale del Motore Semantico, L’Entity Discovery.

## 4.5 Motore Semantico: Entity Discovery

Una volta creata una base di dati ricca di informazioni, è assolutamente necessario elaborare il suo contenuto mediante tecniche di knowledge discovery. Di seguito illustriamo lo strumento da noi realizzato per tale fine: il Tool di Entity Discovery; inizieremo dandone una prima descrizione generale, necessaria per comprendere alcuni concetti base essenziali per capire la successiva parte in cui l’algoritmo verrà presentato in maniera più dettagliata.

### 4.5.1 Descrizione generale

L’obiettivo principale dell’Entity Discovery è quello di scoprire ed estrarre, mediante l’analisi dei dati, delle entità rappresentanti eventi critici in atto.

Il suo funzionamento parte dallo studio di informazione sociale corrispondente ad un picco di parlato sulla piattaforma social Twitter e sui messaggi provenienti dagli utenti utilizzatori dell’applicativo mobile TORCIA, al fine di ricercare e scoprire l’avvenimento che lo ha prodotto. Come è stato dedotto

## 4.5 Motore Semantico: Entity Discovery

---

dall'analisi riportata nel capitolo ?? (SARÀ NELLA STORIA DELL'ARTE), in corrispondenza di una situazione di emergenza, si crea un "trend" nel modo di riferirsi all'evento critico, grazie un rimbalzo social continuo e preciso: le espressioni adottate sono composte in prevalenza da una Keyword, indicante la natura del fatto, associata al Luogo dell'accadimento.

L'aspetto interessante risiede nel fatto che la Keyword è spesso la medesima o, più in generale, è scelta all'interno di un insieme ristretto. Per tali motivi, la ricerca dell'Entity viene effettuata per mezzo di questa associazione di termini, individuando le Keyword più utilizzate e la parola indicante il Luogo dell'evento. Il risultato finale è un entità del tipo Keyword+Luogo.

Una volta trovata l'Entity, essa viene utilizzata per aggregare i tweet riferiti all'evento, cioè quelli che hanno permesso la sua identificazione. In questo modo, viene dato un valore aggiunto ai dati: gli operatori potranno consultare solamente i tweet di interesse, cioè quelli riferiti all'accadimento in atto, al fine di intervenire in maniera più celere possibile; i cittadini, inoltre, vengono facilitati nella visione delle informazioni, poichè le potranno esaminare, filtrandole per occorrenza di evento. Per di più, avere le informazioni aggregate per evento può essere utile alle autorità in fase di *Mitigation*, in quanto sarà possibile effettuare un'analisi più mirata su ciò che è accaduto durante l'emergenza e quindi dedurre gli eventuali miglioramenti che è necessario apportare al piano di emergenza.

L'aspetto più interessante legato all'Entity Discovery è dato dalla possibilità di sfruttare questo strumento come trigger per la generazione di alerting. Infatti, mediante un'interazione indiretta con l'applicazione mobile, sarebbe possibile informare automaticamente gli operatori sull'evento critico in atto con una notifica di alerting. Inoltre, si potrebbero raggiungere direttamente anche tutti i volontari, che spesso non risultano pronti a rispondere alla chiamata, in quanto contattati via email o call list.

Inoltre automaticamente una notifica di alerting agli operatori, per mezzo di un'analisi di informazione sociale, è una tecnica molto innovativa ed efficace, soprattutto se si valuta il fatto che ciò può permettere una più celere reazione da parte delle autorità e, di conseguenza, un più incisivo intervento. Per di più, una volta notificato l'alerting, gli operatori avrebbero

a disposizione un gran numero di dati e informazioni sull'evento in atto e ciò gli permetterebbe di pianificare un piano operativo di intervento più mirato e puntuale, e di gestire al meglio le risorse a disposizione.

Dato che l'Entity identificata è formata da un'associazione di termini tra cui anche il luogo dell'evento critico, tale strumento può essere utilizzato, oltre che per la geolocalizzazione del fatto, anche per aumentare manualmente la specificità geografica, modificando, in tempo reale, i *metadata* componenti le classi. Ciò permette di avere una classificazione sempre più efficace e quindi di generare dati e informazioni maggiormente utili all'intervento delle autorità.

In questo modo il Tool di Classificazione e di Entity Discovery interagiscono in maniera indiretta tra di loro, cooperando nell'analisi e nella generazione di informazione valida ed efficiente.

### 4.5.2 Metodologia di sviluppo

In questa sezione descriveremo la metodologia seguita per lo sviluppo del tool di Entity Discovery.

L'obiettivo principale è quello di creare uno strumento che, una volta identificato un picco di parlato, analizzi i tweet che lo hanno provocato al fine di ricercare, al loro interno, un'Entity comune.

Lo sviluppo di tale strumento è partito dall'idea dei precedenti tesisti Giulia Cesana e Adolfo D'Ettore di sfruttare un'interessante proprietà riscontrata nella loro analisi del caso di Catania (di cui parleremo approfonditamente anche noi nel successivo capitolo 5 sul testing): l'elemento principale che accomuna i tweet riferiti al caso in analisi è la presenza di una delle Keyword utilizzate per il crawling, e di un termine geografico, indicante il Luogo dell'evento [7]. Infatti la maggior parte delle persone si riferiva all'evento mediante espressioni come “Nubigragio Catania” o “Alluvione Catania”.

Partendo da questa loro idea di base e da un primo abbozzo di implementazione, siamo riusciti a sviluppare un algoritmo ed un programma che funzionano in maniera efficace ed efficiente e che sono uno dei punti cardine dell'intero progetto TORCIA.

Andremo ora a vedere nel dettaglio come è costruito l'algoritmo nella sua forma finale.

### 4.5.3 Descrizione dettagliata dell'algoritmo

Il Tool di Entity Discovery è stato studiato per risultare al contempo semplice ed efficace. Come appena spiegato, è incentrato sulla valutazione del superamento di alcune soglie, al fine di individuare un eventuale picco di discussione e l'evento ad esso associato. Ogni Entity, appunto, è formata dall'accostamento di una Keyword e di un Luogo Geografico. Le Keyword sono dei termini, memorizzati nel Database, legate ai fenomeni di emergenza di cui stiamo trattando, per esempio parole come "ALLUVIONE" o "NUBIFRAGIO". L'algoritmo realizzato si compone di 3 parti principali:

- Ricerca di un'emergenza
- Trasmissione della segnalazione trovata (opzionale)
- Ricerca di un'allerta

Nella pratica il codice è unico, ma nelle seguenti sezioni descriveremo nel dettaglio tale algoritmo suddividendolo nelle sue tre aree di competenza, per farne meglio comprendere la struttura e le modalità del suo effettivo funzionamento.

#### 4.5.3.1 Schema generale per il rilevamento di un'emergenza

Mediante la Figura 4.9, vogliamo mostrare il flusso di esecuzioni che il Tool compie per identificare un'Entity, indice di un evento critico in atto.

## Il Motore Semantico

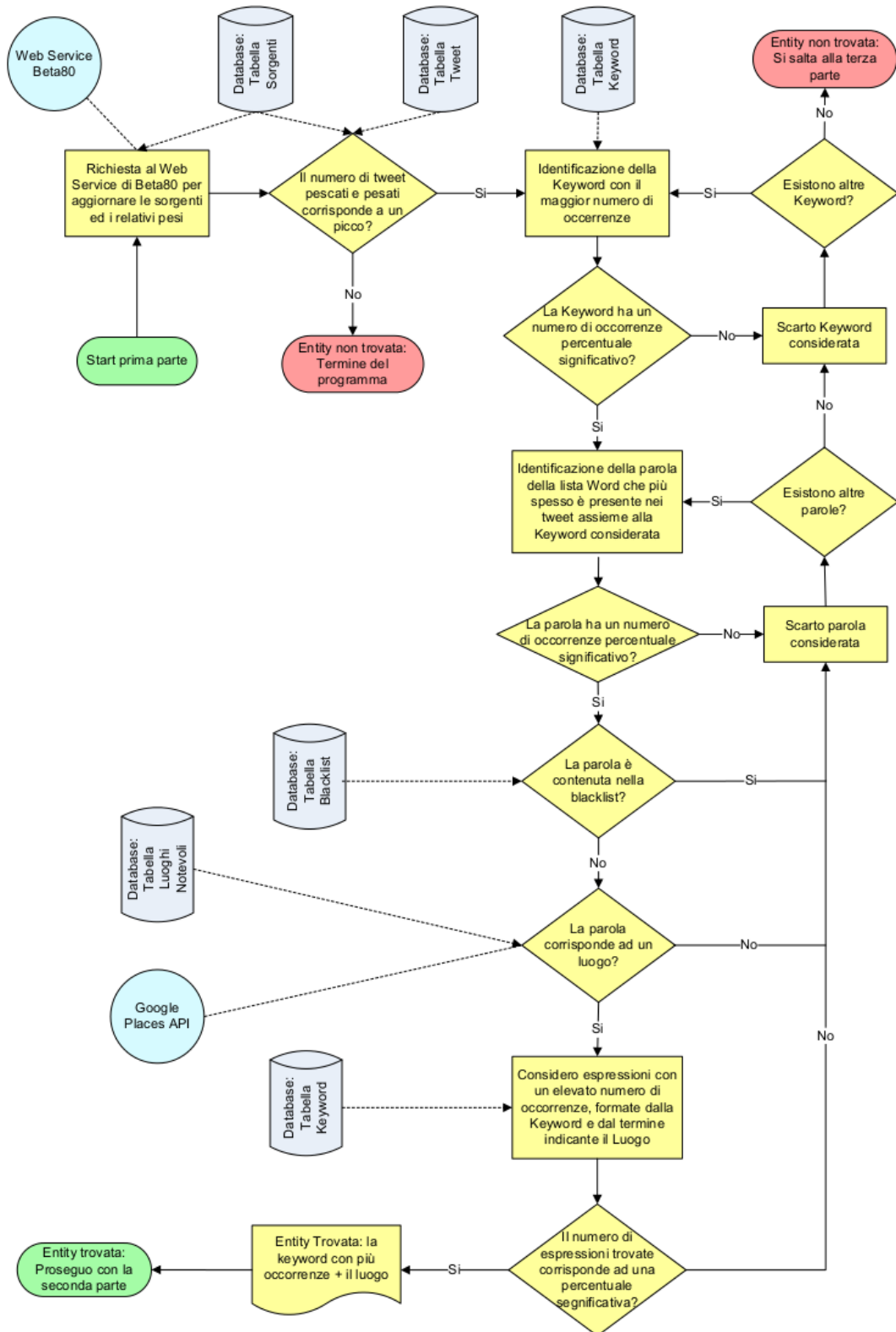


Figura 4.9: Diagramma di Flusso per il rilevamento di un'emergenza.

L'obiettivo del tool è quello di scoprire se i dati prelevati dal Database



## 4.5 Motore Semantico: Entity Discovery

---

corrispondano effettivamente ad un picco di parlato.

La primissima operazione che si compie è una chiamata al Web Service di Beta80, per ricevere la lista aggiornata delle sorgenti e dei relativi pesi. Questa operazione è compiuta ad ogni lancio del Tool di Entity Discovery in quanto è necessario che questo sia sempre aggiornato sui pesi delle varie fonti, nel caso un Operatore, per esempio, decidesse di modificarli ad emergenza in atto. In fondo a questo paragrafo verrà spiegata l'importanza e l'utilizzo dei pesi stessi.

A questo punto, la query creata pesca dal Database tutti i tweet che sono stati scritti nelle ultime 24 ore, confrontando il numero totale di dati con una soglia statica precedentemente impostata (informazioni più specifiche al riguardo si possono trovare alla sottosezione 4.5.3.4 dedicata alle soglie). Se tale numero supera la soglia, allora la presenza di questo picco può essere sintomo di un evento critico in corso.

Il campione così analizzato corrisponde all'insieme dei tweet scritti a partire dalle 24 ore precedenti fino a quel momento. In questo modo consideriamo un campione di dati abbastanza omogeneo, il cui numero non è influenzato da altri fattori esogeni come, per esempio, il fatto che il numero medio di tweet diminuisca notevolmente durante le ore notturne.

Una volta identificato un picco, il software deve stabilire se esso è dovuto ad un'emergenza in atto o se, semplicemente, il parlato su Twitter ha subito un'impennata in quel lasso di tempo per un altro fattore non rilevante per noi. Qualora vi fosse un riscontro, è compito del Tool identificare un'Entity che lo rappresenti.

Sono le Keyword ad essere lo strumento più efficace per l'individuazione di una Entity e, pertanto, il Tool analizza i dati in input ricercandole e assegnando ad ognuna di esse il numero percentuale di occorrenze. A questo punto l'Entity Discovery analizza i tweet in cui compare la keyword con il maggior numero di occorrenze e ricerca al loro interno le parole che più spesso sono associate ad essa, dove associate significa che compaiono assieme nello stesso tweet. Questo insieme di parole così definite prende il nome di Word. Una volta individuate, esse vengono esaminate per scoprire se ne esiste almeno una indicante un luogo geografico. A questo scopo vengono utilizzati:

- la Tabella Blacklist, già introdotta nella sezione 4.2.2, con cui si controlla

## Il Motore Semantico

---

se la Word è inclusa tra i termini che si è deciso di non considerare

- la Tabella dei Luoghi Notevoli, anch'essa già introdotta sempre nella sezione 4.2.2, con cui si valuta la presenza della Word tra i possibili luoghi preinseriti nella tabella
- il servizio di Google denominato “Google Places”, nel caso la Word non sia già stata trovata in una delle tabelle sopra

Se si ha un riscontro, e quindi esiste un luogo geografico tra le parole più usate in associazione alla Keyword valutata, il Tool ricerca tutti i tweet contenenti il luogo scoperto, al fine di individuare tutte le Keyword più utilizzate al loro interno. Ciò è stato pensato come perfezionamento, in quanto non è detto che le Keyword più utilizzate all'interno dell'intero campione corrispondano effettivamente a quelle più utilizzate per riferirsi all'evento.

Come conclusione del processo, si conta la percentuale di occorrenze delle espressioni composte dalle Keyword più utilizzate, non soltanto quindi la prima assoluta, e dalla parola indicante il luogo e si controlla se tale valore supera un'ulteriore soglia preimpostata. Se ciò avviene, allora è identificato l'evento di emergenza e viene creata l'Entity ad esso associata: essa sarà formata dalla Keyword più utilizzata, affiancata alla Word che denota il Luogo. Per esempio nel caso si fosse identificato un evento di tipo Alluvione, con luogo di riferimento a Milano, l'Entity creata avrebbe il nome “ALLUVIONE MILANO”.

Come già spiegato nella sezione 4.2.2 sul Database, i messaggi vengono memorizzati con un loro peso, a seconda che provengano da Twitter o dall'app mobile. Questo è stato fatto per poter valorizzare maggiormente le informazioni provenienti dall'applicativo, in quanto gli utenti saranno, in linea teorica, persone interessate e attive, che forniranno informazioni più precise e strutturate rispetto alla massa confusa dei tweet.

Al momento di prelevare le informazioni dal Database, l'Entity Discovery effettua quindi una conversione dei pesi, semplicemente moltiplicando questo dato per 100, in modo da ottenere un numero intero. In questo modo ogni tweet, a cui come già detto è dato il peso fisso basilare di 0.01, conterà come un

---

## 4.5 Motore Semantico: Entity Discovery

unità di informazione, mentre ogni messaggio realizzato con l'applicativo verrà considerato come 10, 20 o anche 50 tweet, a seconda del valore che l'Operatore avrà dato al peso di questa fonte.

La tabella delle sorgenti, infatti, potrà essere ampiamente modificata:

- aggiungendo nuove fonti, in una futura ottica di integrazione con altre sorgenti social (YouReporter per esempio) o con strumenti ad-hoc (come le sonde di rilevamento dei livelli d'acqua poste sugli argini di alcuni fiumi)
- cambiando uno o più valori dei pesi esistenti, dando così la possibilità agli Operatori di migliorare la sensibilità del Motore Semantico alle diverse fonti in base ai risultati raggiunti, mediante una continua opera di apprendimento

Nella sezione successiva spiegheremo come, a questo punto, può iniziare l'esecuzione della seconda parte di algoritmo, responsabile della gestione delle segnalazioni.

### 4.5.3.2 Schema generale per la gestione della segnalazione

Questa parte di codice viene eseguita solamente nel caso in cui il Tool rilevi la presenza di almeno un Entity. In caso contrario, l'esecuzione del software passa alla fase successiva reportando un caso tra:

- La non presenza di un picco di parlato
- La non esistenza di un Entity a fronte del rilevamento del picco

Mediante la Figura 4.10, vogliamo mostrare il flusso di esecuzioni che il Tool compie una volta identificata un'Entity, al fine di segnalarla.

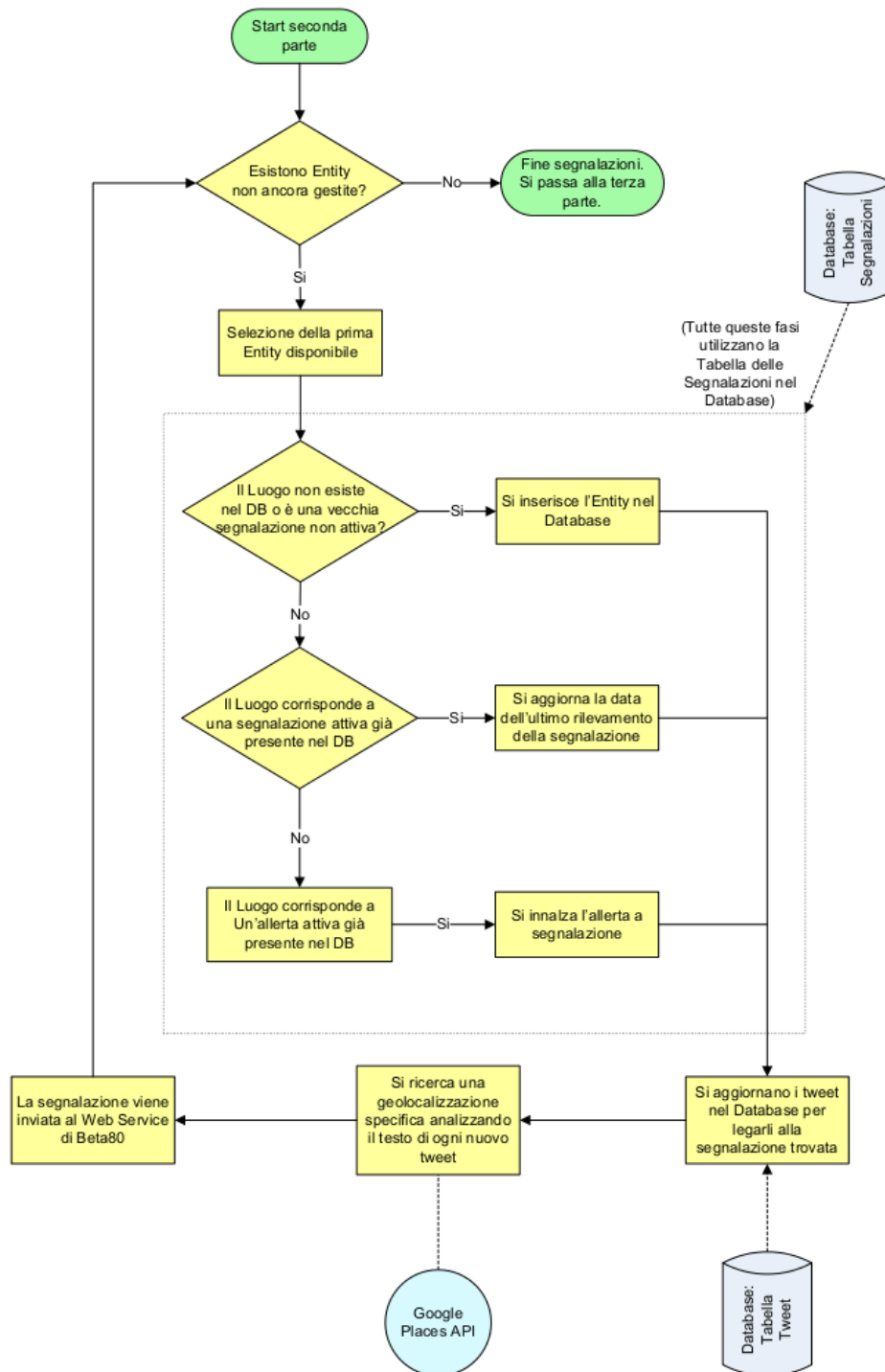


Figura 4.10: Diagramma di Flusso per la gestione delle segnalazioni.

Per ogni Entity rilevata, la prima operazione che il programma compie è controllare nel Database l'esistenza o meno di un'altra Entity attiva con lo stesso luogo geografico di interesse.

## 4.5 Motore Semantico: Entity Discovery

---

Se non ci sono dei riscontri, viene creata una nuova tupla contenente la segnalazione appena rilevata, secondo il formato precedentemente spiegato nella sezione 4.2.2.

Nel caso invece vi sia un riscontro, è necessario capire se si tratta di un effettiva segnalazione ancora attiva o di un'allerta attiva.

Nel primo caso significa che l'Entity Discovery aveva già in precedenza identificato l'emergenza e che questa è ancora in atto, la gente ne parla ancora. A questo punto il Tool deve semplicemente aggiornare il campo `DT_SIGNALING_STATUS`, contenente appunto la data e l'ora dell'ultimo rilevamento di tale Entità, e, nella fase di invio, si dovrà mandare una segnalazione avente ID uguale a quello del vecchio invio e contenente solo i nuovi tweet rilevati e ad essa collegati, vedremo più avanti come.

Nel secondo caso siamo di fronte ad uno stato di allerta (di cui parleremo a breve) che ora è diventato abbastanza importante da essere considerato un'emergenza a tutti gli effetti. Il Tool dovrà allora aggiornare sia la data che lo stato dell'allerta, per farla diventare una segnalazione.

A valle di ognuno di questi tre casi, a prescindere da quale sia il punto di partenza, il software di Entity Discovery deve performare altre tre operazioni, prima di poter passare alla fase successiva:

- Update dei Tweet per legarli alla segnalazione
- Ricerca di una geolocalizzazione specifica
- Invio della segnalazione

L'update dei tweet consiste semplicemente nella modifica del valore del campo `ID_SEMANTIC_SIGNALING` di ogni tweet che ha concorso alla creazione, o aggiornamento, della segnalazione, inserendovi, qualora non sia già presente, l'identificatore della segnalazione stessa. Come già spiegato sempre nella sezione 4.2.2 dell'architettura del sistema infatti, ogni segnalazione ha un ID univoco e in questo modo ogni tweet può essere facilmente ricollegato all'eventuale emergenza a cui fa riferimento.

## Il Motore Semantico

---

Ogni tweet che il Motore Semantico riceve dal Crawler di Beta80 può contenere o meno dei valori nei campi di Latitudine e Longitudine, a seconda se il dispositivo con cui è stato inviato aveva la funzione di localizzazione attiva o meno. Il Tool di Entity Discovery cerca però, qualora sia possibile, di migliorare ulteriormente questi dati per mezzo di una Geolocalizzazione Semantica. Viene effettuata un'analisi semantica all'interno di ogni tweet relativo alla nuova segnalazione, col fine di identificare eventuali riferimenti a vie, piazze, corsi, etc... all'interno del messaggio. Per ogni tweet in cui si ha un riscontro, i campi di latitudine e longitudine della sua tupla nel Database vengono aggiornati con le coordinate geografiche di Latitudine e Longitudine, ottenuti mediante l'utilizzo delle API di Google, di cui si parlerà meglio nella successiva sottosezione 4.5.3.5. Nel caso invece che non si trovi alcun riferimento semantico all'interno del tweet, questo mantiene semplicemente le sue coordinate originarie, se presenti.

L'invio di una segnalazione consiste nella sua notificazione al Web Service di Beta80, perennemente in ascolto, in modo che il resto del sistema TORCIA possa processare la segnalazione stessa e farle proseguire il suo iter. Questa funzionalità è realizzata mediante l'invocazione di una funzione che attiva il servizio di PUSH del Web Service di Beta80, in maniera analoga al nostro servizio descritto nella precedente sezione 4.3.

Prima di poterla inviare però, l'Entity Discovery deve creare il file XML che dovrà essere spedito. La sua struttura è la seguente, in cui non riportiamo anche lo schema della struttura dei singoli tweet in quanto già vista in precedenza sempre nella sezione 4.3:

## 4.5 Motore Semantico: Entity Discovery

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xs="http://www.w3.org/2001/XMLSchema"
attributeFormDefault="unqualified" elementFormDefault="qualified">

  <xsd:element name="AN_SIGNALING">
    <xsd:complexType>
      <xsd:sequence>

        <xsd:element name="ID_SEMANTIC_SIGNALING" type="xsd:unsignedByte" />
        <xsd:element name="DS_TEXT" type="xsd:string" />
        <xsd:element name="CD_MATCHING_DOMAIN" type="xsd:string" />
        <xsd:element name="DT_CREATION" type="xsd:dateTime" />
        <xsd:element name="CD_SIGNALING_STATUS" type="xsd:unsignedByte" />
        <xsd:element name="DT_SIGNALING_STATUS" type="xsd:dateTime" />
        <xsd:element name="VL_ACCURACY" type="xsd:decimal" />
        <xsd:element name="DS_SIGNALING_POSITION" type="xsd:string" />
        <xsd:element name="VL_SIGNALING_POSITION_LATITUDE" type="xsd:decimal" />
        <xsd:element name="VL_SIGNALING_POSITION_LONGITUDE" type="xsd:decimal" />

        <xs:element name="EXTRACTED_INFO_LIST">
          ... QUI SI TROVA LA LISTA DI TUTTI I TWEET LEGATI ALLA SEGNALAZIONE ...
        </xs:element>

      </xsd:sequence>
    </xsd:complexType>
  </xsd:element>

</xs:schema>
```

Figura 4.11: Struttura XSD per la trasmissione delle segnalazioni.

Di seguito riportiamo un esempio applicato alla struttura appena vista, completa dei relativi tweet (per praticità qui ne inseriamo solo uno):

```
<AN_SIGNALING>
  <ID_SEMANTIC_SIGNALING>1</ID_SEMANTIC_SIGNALING>
  <DS_TEXT>ALLAGAMENTO LECCO</DS_TEXT>
  <CD_MATCHING_DOMAIN>ALLAGAMENTO</CD_MATCHING_DOMAIN>
  <DT_CREATION>2014-01-30T11:14:27</DT_CREATION>
  <CD_SIGNALING_STATUS>4</CD_SIGNALING_STATUS>
  <DT_SIGNALING_STATUS>2014-01-30T11:14:27</DT_SIGNALING_STATUS>
  <VL_ACCURACY>100</VL_ACCURACY>
  <DS_SIGNALING_POSITION>LECCO</DS_SIGNALING_POSITION>
  <EXTRACTED_INFO_LIST>
    <EXTRACTED_INFO>
      <CD_CRAWLING_SOURCE>TWITTER</CD_CRAWLING_SOURCE>
      <CD_MATCHING_DOMAIN>ALLUVIONE</CD_MATCHING_DOMAIN>
      <CD_USER_NAME>Ciummi89</CD_USER_NAME>
      <DS_TEXT>allagamento in viale Filippo Turati, Lecco</DS_TEXT>
      <DT_EXTRACTION>2014-01-30T11:14:22</DT_EXTRACTION>
      <DT_REFERENCE>2014-01-30T11:14:22</DT_REFERENCE>
      <ID_EXTRACTED_INFO>297</ID_EXTRACTED_INFO>
      <IS_VALIDATED>1</IS_VALIDATED>
      <VL_POSITION_LATITUDE>45.86123</VL_POSITION_LATITUDE>
      <VL_POSITION_LONGITUDE>9.38875</VL_POSITION_LONGITUDE>
      <DS_SEMANTIC_TOPICS>CAUSA,EFFETTO</DS_SEMANTIC_TOPICS>
    </EXTRACTED_INFO>
  </EXTRACTED_INFO_LIST>
</AN_SIGNALING>
```

Figura 4.12: Esempio valido di un XML contenente una segnalazione.

Il significato degli elementi è il seguente:

## Il Motore Semantico

---

- ID\_SEMANTIC\_SIGNALING : ID della segnalazione che sta per essere inviata
- DS\_TEXT : Testo dell'Entity legata alla segnalazione
- CD\_MATCHING\_DOMAIN : Dominio dell'emergenza rilevato dall'Entity Discovery (la Keyword)
- DT\_CREATION : Indica la data in cui la segnalazione è stata creata
- CD\_SIGNALING\_STATUS : Indica lo stato della segnalazione all'interno del sistema TORCIA (Inizializzato da altre componenti del sistema, noi non lo utilizziamo o modifichiamo)
- DT\_SIGNALING\_STATUS : Indica la data in cui la segnalazione è stata aggiornata l'ultima volta
- VL\_ACCURACY : Campo che permette di inserire la precisione percentuale della segnalazione (inizializzato e utilizzato dal sistema TORCIA più avanti, noi non lo utilizziamo o modifichiamo)
- DS\_SIGNALING\_POSITION : Indica il luogo in cui l'Entity Discovery ha rilevato l'emergenza
- VL\_SIGNALING\_POSITION\_LATITUDE e VL\_SIGNALING\_POSITION\_LONGITUDE : Contengono le coordinate geografiche del Luogo in cui si rileva l'emergenza, nel caso in cui quello sia presente nel Database nella Tabella dei Luoghi Notevoli

A questo punto la segnalazione così creata è pronta per essere inviata.

Per completezza di trattazione, riportiamo inoltre due ulteriori modifiche fatte a questa parte di codice, atte a eliminare la ridondanza dei dati e ad evitare di intasare le connessioni del Web Service:

1. Dopo aver inviato la segnalazione, il valore del campo ID\_SEMANTIC\_SIGNALING di ogni tweet è reso negativo (es. da "25" a "-25") in modo che non si perda la correlazione tra tweet ed emergenza (l'id è semplicemente l'opposto), ma successivi riscontri della



## 4.5 Motore Semantico: Entity Discovery

---

stessa Entity inviino nella notifica solo i tweet appena arrivati, niente di già trasmesso

2. Nel caso il numero di tweet legati a una segnalazione sia grande, questa viene suddivisa in più parti, al massimo da 200 tweet l'una, e inviate consecutivamente. Questo permette al Web Service di non andare in Timeout per il troppo tempo impiegato nella ricezione di un singolo file XML contenente magari 3000 tweet.

L'intero processo di gestione della segnalazione viene iterato per ogni Entity trovata, fino a quando non ne restano più. Poi si passa alla fase successiva di ricerca delle allerte.

### 4.5.3.3 Schema generale per il rilevamento di un'allerta

A prescindere dal fatto che sia stata inviata o meno una segnalazione, questa parte di codice verrà eseguita.

Mediante la Figura 4.13, vogliamo mostrare il flusso di esecuzioni che il Tool compie per identificare un'Entity, indice di un evento critico in atto.

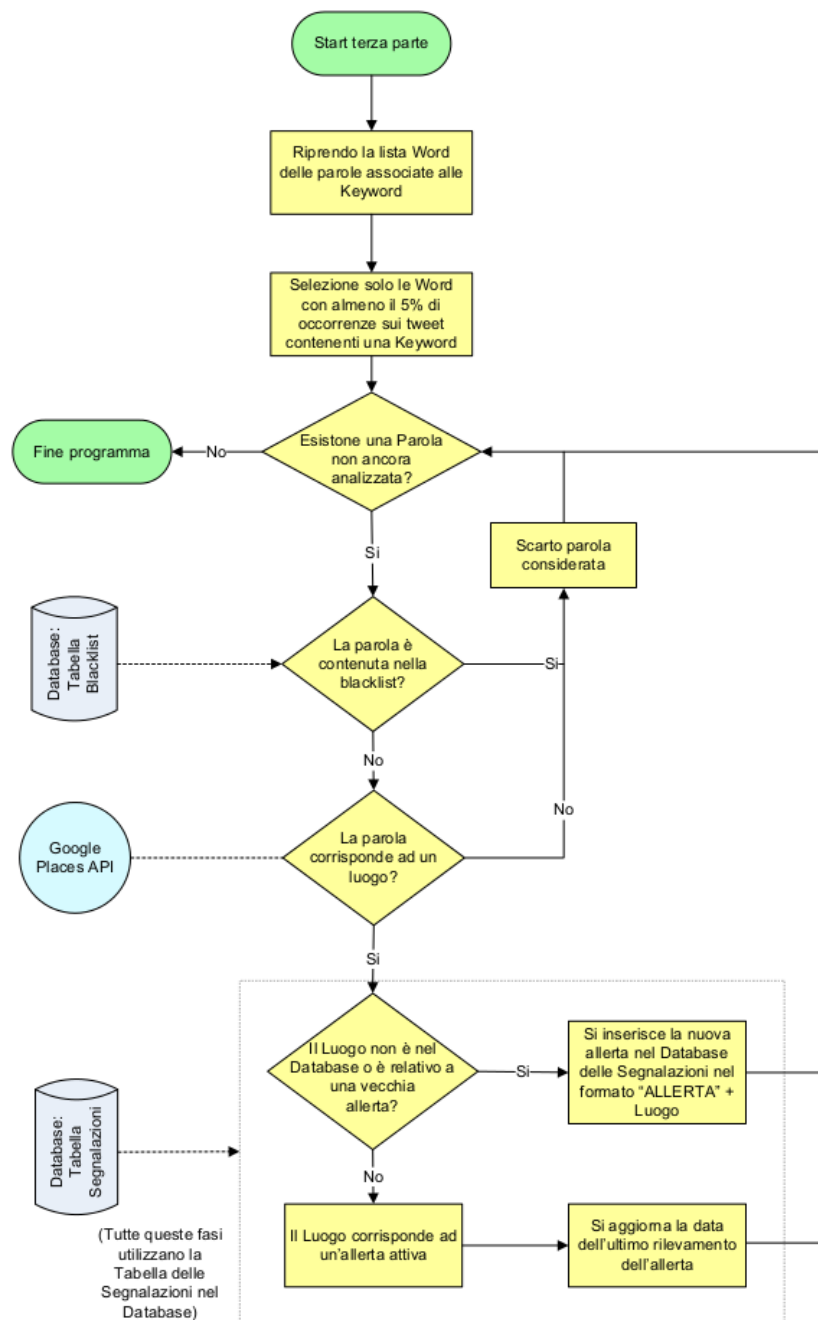


Figura 4.13: Diagramma di Flusso per il rilevamento di un'allerta.

Un'Allerta non è altro che una potenziale emergenza che non ha però ancora raggiunto un livello di parlato tale da impensierire ed essere considerata un'emergenza degna di segnalazione. Il suo formato di memorizzazione differisce, da quello di una normale segnalazione, per due aspetti:

- L'entità che viene salvata è del tipo "ALLERTA"+Luogo; ciò poiché,

## 4.5 Motore Semantico: Entity Discovery

---

in questa fase, non si sa ancora quale sarà la Keyword dominante (e se ci sarà), per cui si raggruppano tutte sotto un più generico termine “ALLERTA”.

- Il campo “ACTIVE”, presente nella tabella delle Segnalazioni, viene settato a 2, in modo da distinguerle da segnalazioni sia vecchie (ACTIVE = 0) che nuove (ACTIVE = 1).

Come già accennato in precedenza, e come verrà meglio discusso nella prossima sezione 4.5.3.4, il metodo di rilevamento del parlato si basa su delle Soglie, sia per capire se è necessario eseguire il programma, sia per capire se il parlato relativo a una Keyword è significativo da identificare un'emergenza. La ricerca delle allerte si basa, in sostanza, su una logica simile, sfruttando quella che potremmo definire come una Soglia di Allerta. Quello che si vuole fare è trovare tra le Word rimanenti un termine relativo ad un Luogo geografico la cui occorrenza superi il valore di soglia. Queste Word sono infatti le parole più ricorrenti nei tweet contenenti delle Keyword, ma che non sono ancora state usate all'interno di una recente segnalazione.

Se si ha un riscontro, si possono presentare due casistiche:

- Non è presente né un'Allerta né una segnalazione attiva con quella stessa Entity; in questo caso allora abbiamo trovato una nuova Allerta e questa viene aggiunta al Database
- Esiste una precedente Allerta attiva con quella Entity; in questo caso viene semplicemente aggiornata la data di ultima rilevazione

A questo punto il programma termina del tutto la sua esecuzione, pronto a ricominciare.

La ricerca delle potenziali allerte esula dalle richieste del progetto TORCIA. Questo tipo di rilevazioni infatti non vengono notificate alle altre parti del sistema, ma è un'idea che abbiamo avuto e realizzato noi in quanto riteniamo possa essere utile sia ai fini di un futuro miglioramento del progetto, che per gli scopi del nostro lavoro di tesi, per capire le reali potenzialità di prevenzione del sistema TORCIA. Un'Allerta memorizzata infatti, nel futuro, può sia scomparire nel nulla senza provocare alcun danno, che aumentare di

portata e diventare motivo di crisi e, per questo, è necessario tenerle sotto controllo.

Durante la nostra fase di testing, descritta approfonditamente nel capitolo 5, abbiamo potuto notare come l'anticipo dell'allerta, rispetto all'effettiva segnalazione, sia, in molti casi, notevole e sicuramente utile in un'ottica di prevenzione delle emergenze.

### 4.5.3.4 Soglie

Come si può notare dalla precedente descrizione, l'algoritmo fa uso di alcune soglie statiche numeriche o percentuali. Attualmente queste soglie sono tarate in base ad una serie di testing, effettuati su un campione di dati abbastanza ampio da ritenerlo attendibile (circa 700mila tweet).

Inizialmente si era pensato di introdurre una soglia di tipo dinamico per la valutazione dei picchi e, in effetti, era stato realizzato ed implementato un primo abbozzo di algoritmo in tal senso. Tuttavia, a seguito di test e valutazioni effettuate, è risultato che una soglia di tipo statico è più adatta all'attuale numero di tweet che devono essere analizzati e sicuramente migliore per le nostre esigenze.

Infatti, per il sistema TORCIA, è preferibile una soglia statica che esegua il programma di Entity Discovery anche molte volte di troppo, anche perché i tempi di calcolo sono minimi, piuttosto che una soglia dinamica che magari non segnali un effettiva emergenza poiché il trend dei tweet in quel periodo non era tale da permettere all'algoritmo di apprendimento dinamico della soglia di rilevarla.

Nello specifico le soglie utilizzate dall'algoritmo sono:

- Soglia per identificare un picco di “parlato” : essa è tarata sul numero di tweet minimo necessario per dimostrare la presenza di un aumento significativo di “parlato”. Nel caso in cui questa soglia non venga superata, il tool conclude il suo processo, dato che non vi è nessun segnale che testimoni la presenza di un evento in atto.
- Soglia per controllare se la Keyword e il Luogo considerati hanno una frequenza significativa : è una soglia percentuale e viene calcolata sulla base del numero di tweet prelevati in input dal tool. Il motivo di tale

## 4.5 Motore Semantico: Entity Discovery

---

scelta è il fatto che è necessario controllare che la Keyword, o il termine nel caso si stia cercando il Luogo, che stiamo analizzando sia frequente all'interno dell'insieme globale di tweet esaminato. Se questa soglia non viene valicata, allora la Keyword, o il termine, non è abbastanza frequente nei tweet per essere utile nella ricerca dell'Entity.

- Soglia per constatare l'effettiva scoperta della Entity : è la soglia percentuale definitiva ed è anch'essa calcolata sul numero di tweet prelevati in input dal tool. In pratica si controlla che le espressioni, che possono indicare la presenza di un evento, abbiano un'occorrenza percentuale significativa all'interno di tutti e soli i tweet che sono considerati per la generazione della segnalazione stessa. È la soglia più importante in quanto determina la scoperta o la non scoperta dell'Entity e, pertanto, ha richiesto un'accurata analisi per la sua definizione. Qualora la soglia non venga oltrepassata, sarà ricercata una nuova parola da associare alla Keyword analizzata, indirizzando la ricerca verso questa nuova direzione.
- Soglia per la ricerca di potenziali allerte : è la soglia percentuale utilizzata dopo la valutazione delle possibili entità, per la ricerca di quelle che sono le allerte, potenziali segnalazioni. Viene calcolata sulla base del numero di tweet che contengono almeno una Keyword. In tutti questi, si selezionano solo i termini che hanno una presenza percentuale significativa rispetto agli altri e che saranno quindi valutati per capire se si tratta di un Luogo geografico in cui si consiglia, a quel punto, di identificare un'allerta.

Indichiamo nella seguente tabella 4.6 il valore di tali soglie, definiti a seguito di vari test su casi reali. La prima è definita come “settabile” in quanto il valore non è fissato all'interno del codice, ma gestibile da un operatore esterno mediante un file di configurazione di cui si parlerà più avanti nella sezione 4.5.3.7.

## Il Motore Semantico

---

Soglia	Valore
Soglia per identificare un picco di “parlato”	Impostabile dall’operatore
Soglia per controllare se la keyword e il luogo considerati hanno una frequenza significativa	10%
Soglia per constatare l’effettiva scoperta della entity	15%
Soglia per la ricerca di potenziali allerte	5%

Tabella 4.6: Soglie statiche del Tool di Entity Discovery.

### 4.5.3.5 Gestione della Localizzazione

La Geolocalizzazione dei tweet è un aspetto allo stesso tempo importante e complicato del Motore Semantico.

Per tal motivo, spiegheremo ora nel dettaglio il funzionamento dei due strumenti fondamentali a cui il Tool di Entity Discovery fa ricorso per lo svolgimento del suo compito di localizzazione, ovvero la tabella dei Luoghi Notevoli e le API di Google Places. Nell’ultimo paragrafo faremo poi un riassunto di tutte le informazioni fornite riguardo la Geolocalizzazione in questa e nelle sezioni precedenti, per aiutare ad avere una visione migliore a livello generale.

**Luoghi Notevoli** Per come è costituito l’algoritmo, il tool deve poter decidere se un termine identifica o meno un Luogo geografico. L’algoritmo compie questa analisi in due forme:

- durante la ricerca delle emergenze e delle allerte, per capire se una delle Word corrisponde a una regione o una città (o cittadina, paese, ecc...)
- durante l’analisi semantica atta ad identificare, all’interno di ogni tweet legato ad una segnalazione, la presenza del riferimento a una geolocalizzazione specifica (una via, piazza, corso, ecc...)

Mentre nel secondo caso l’unico metodo è l’utilizzo delle Google Places API, di cui parleremo nel prossimo paragrafo, per il primo caso, prima di effettuare una richiesta a Google, il tool utilizzerà la Tabella dei Luoghi Notevoli.

Come già spiegato in precedenza, si tratta di una semplice tabella in cui sono riportati un nome identificativo per il Luogo e le sue coordinate di Latitudine e Longitudine. Nel caso il tool riscontri che la Word in esame

## 4.5 Motore Semantico: Entity Discovery

---

è contenuta in questa lista, non sarà necessario effettuare richieste a servizi online, ma il termine sarà assunto come Luogo e utilizzato per la costruzione dell'Entity, mentre le sue coordinate verranno inserite negli appositi campi nell'XML della segnalazione VL\_SIGNALING\_POSITION\_LATITUDE e VL\_SIGNALING\_POSITION\_LONGITUDE.

L'importanza dell'introduzione e utilizzo di questa tabella non è, come si potrebbe pensare ad una prima analisi, quella di risparmiare tempo non effettuando delle chiamate a risorse web. Si parla infatti di un'unica singola richiesta e i tempi di risposta del servizio offerto da Google sono irrilevanti.

La vera potenzialità dell'uso dei Luoghi Notevoli è la possibilità di inserire nella tabella le localizzazioni geografiche relative a fiumi, laghi e, in generale, a tutte quelle zone che potrebbero essere a rischio in caso di un'emergenza come un'alluvione (per esempio zone limitrofe a delle dighe) che non possono essere geolocalizzate dal sistema Google Places in quanto non corrispondenti a delle Regioni o Città.

Riferendoci al caso dell'area milanese, per esempio, abbiamo osservato come su Twitter, in caso di forti piogge, gli utenti parlino spesso di "ESONDAZIONE SEVESO", in riferimento al fiume Seveso che scorre a Nord della città. Senza l'utilizzo di questa tabella, il tool di Entity Discovery avrebbe rilevato la potenziale segnalazione identificando però il Luogo "*Seveso*" come il comune di Seveso, località poco fuori Milano, ma non collegata all'emergenza. La tabella dei Luoghi Notevoli ha permesso invece di associare al termine "*Seveso*" delle coordinate ben precise relative all'Area del Niguarda, dove il fiume Seveso è solito appunto esondare durante le forti piogge, provocando dei disagi e rischi per i cittadini.

La possibilità di utilizzare una tabella contenente questi Luoghi "caldi" è una forte innovazione e le potenzialità sono grandi. Il Motore Semantico viene infatti così dotato di una sorta di tecnica di autoapprendimento, in quanto agli Operatori che utilizzeranno il software è permessa una continua aggiunta di geolocalizzazioni all'interno del Database, col fine di migliorare la specificità delle segnalazioni per zone che si ritiene essere spesso e/o molto a rischio, inserendone sempre di nuove o modificando le coordinate delle esistenti, in

## Il Motore Semantico

---

modo da essere sempre più precisi nel tracciare poi mappe e grafici. Anche in questo caso, come per il caso dei *metadati* del Tool di Classificazione, si tratta di un continuo lavoro di affinamento che deve continuare parallelo all'utilizzo del Software anche dopo il suo futuro rilascio.

**Google Places API** Nel caso la tabella dei Luoghi Notevoli non dia riscontri, o nel caso si stiano ricercando le Geolocalizzazioni Semantiche trattate nella sezione 4.5.3.2, il Tool di Entity Discovery deve affidarsi, nella sua ricerca, a dei servizi web esterni: le Google Places API. Google, mediante queste API, fornisce una serie di servizi per recuperare informazioni riguardanti luoghi, tramite richieste HTTP.

Tra tutti questi servizi, la nostra scelta è ricaduta sulla API Google Places Autocomplete, nella quale, attraverso una richiesta appropriata, è possibile stabilire se una stringa corrisponda ad una città. In particolare, tale richiesta deve essere impostata con alcuni specifici attributi che permettano di restringere la ricerca, come per esempio il fatto che la stringa debba corrispondere al nome di una città italiana. Tale restringimento è stato effettuato in quanto il contesto di riferimento in cui si inserisce tale tool è prettamente italiano.

Di seguito gli esempi delle due tipologie di richieste presenti nell'Entity Discovery:

---

---

### Richiesta per la ricerca del Luogo (città o regione):

```
https://maps.googleapis.com/maps/api/place/autocomplete/xml?input=
"+NOME_LUOGO+"&types=(regions)&components=country:it&
language=it&sensor=true&key="+KEY_GOOGLE"
```

---

---

---

### Richiesta per la ricerca di una localizzazione specifica nel Luogo:

```
http://maps.googleapis.com/maps/api/geocode/xml?address=
"+LOC_SPECIFICA+", "+NOME_LUOGO+"&components=country:it&
language=it&sensor=false
```

---

Questi tipi di servizi sono gratuiti, anche se è necessario, al fine di utilizzare il primo dei due, creare un account e un corrispondente progetto all'interno della Google APIs Console. In tal modo è possibile ottenere una key propria da



## 4.5 Motore Semantico: Entity Discovery

---

inserire nella richiesta del servizio al fine di ottenerlo gratuitamente. L'utilizzo del servizio viene purtroppo limitato ad un numero di richieste giornaliere, per questo è necessaria anche una prima serie di filtraggi delle parole, per evitare di inviare troppe richieste a Google e raggiungere il limite, che è comunque molto alto. Questo è compito della funzione di blacklisting di cui tratteremo a breve.

All'interno del progetto TORCIA comunque, è stato deciso che gli incaricati di mappe e cartografia dovranno fornire delle loro API, sostitutive a quelle di Google, per fare richieste in maniera analoga, ma senza limiti di utilizzo. In questo modo il problema del numero di richieste giornaliere non sussisterà più e basterà solo interfacciare questa parte dell'Entity Discovery col loro servizio.

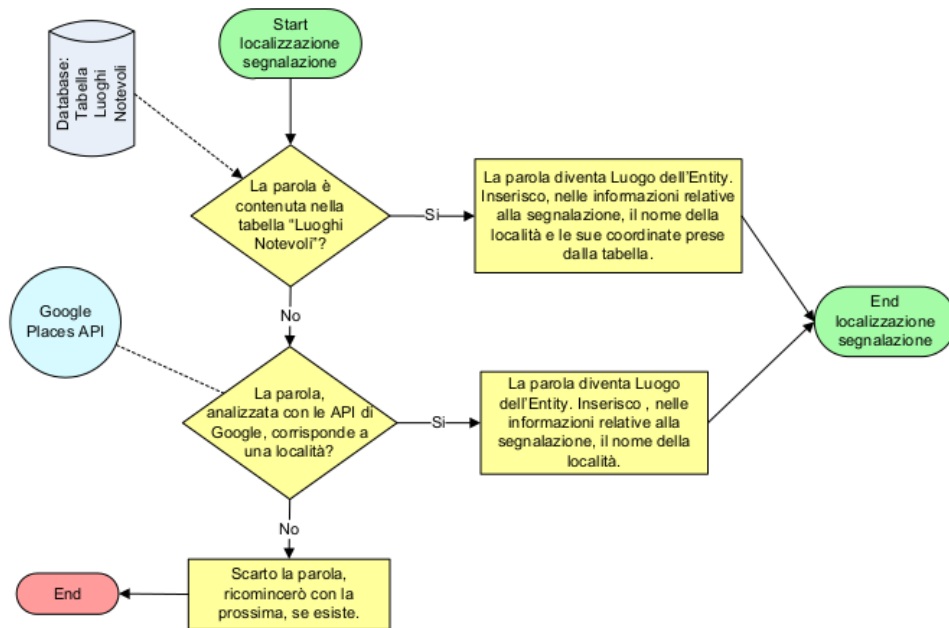
L'output del servizio viene fornito in formato XML, o alternativamente in formato JSON, e contiene al suo interno la risposta alla richiesta effettuata più una serie di informazioni aggiuntive. Per come è costituito, è stato necessario parlarlo, al fine di stabilire se realmente la stringa inviata corrispondeva ad un Luogo nella forma cercata. Infatti, il servizio fornisce un autocompletamento automatico della stringa al fine di trovare un luogo che le corrisponda.

Questa caratteristica è utile per la ricerca di uno specifico riferimento geografico all'interno del testo dei tweet (come vie o piazze), ma rappresenta un problema per il tipo di funzionalità che era necessaria nella ricerca delle segnalazioni, durante l'identificazione di città e regioni.

Tale problema è stato risolto, in questo secondo caso, parsando l'output e controllando che la stringa restituita, rappresentante il nome della città o regione trovata, corrispondesse realmente alla stringa fornitagli nella richiesta.

**Schema riassuntivo sulla Geolocalizzazione** I diagrammi di flusso in Figura 4.14 forniscono una visione generale del funzionamento del sistema di localizzazione del motore semantico:

## LIVELLO SEGNALAZIONE



## LIVELLO TWEET

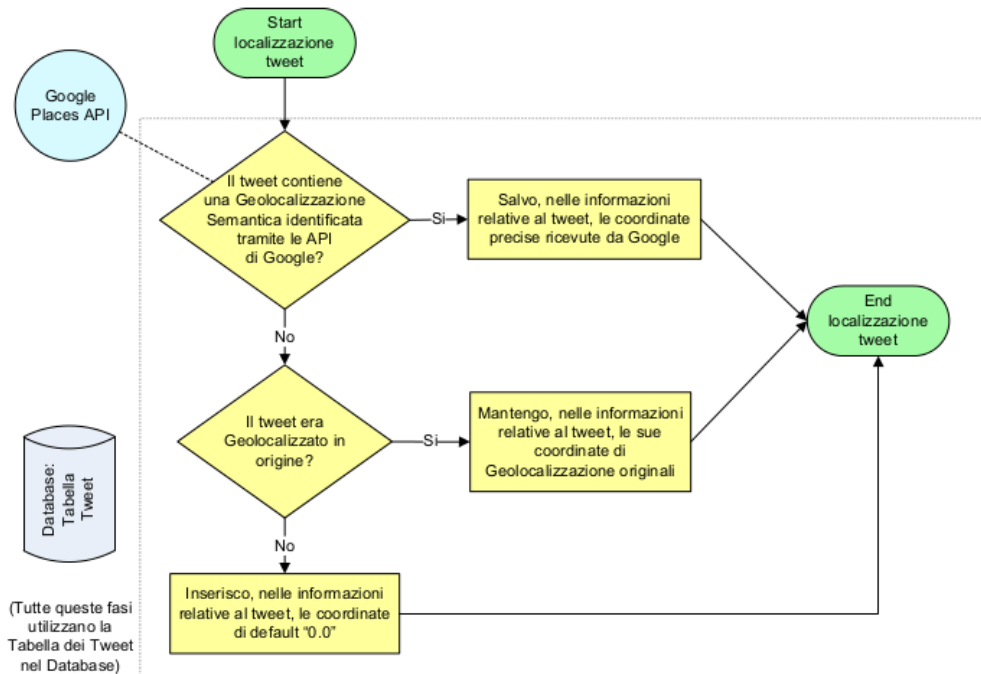


Figura 4.14: Schemi di alto livello per la Geolocalizzazione.

Come si nota, la Geolocalizzazione si può dividere su due livelli:

- Livello di Segnalazione

## 4.5 Motore Semantico: Entity Discovery

---

- Livello di Tweet

### A livello della segnalazione

Il Motore Semantico analizza i termini ricorrenti nei tweet per identificare la presenza di una “Località”, intesa come Regione, Città, Cittadina, Paese, etc...

Per farlo utilizza:

- Tabella dei “Luoghi Notevoli”
- API di Google

Come prima cosa il MS ricerca se il termine in esame si trova all’interno della tabella. Nel caso vi sia un riscontro, a livello segnalazione si comunicherà:

- nel campo DS.SIGNALING\_POSITION il nome del Luogo Notevole
- nei campi VL.SIGNALING\_POSITION\_LATITUDE e VL.SIGNALING\_POSITION\_LONGITUDE le coordinate del relativo Luogo Notevole che sono presenti nella tabella

Nel caso non vi sia un riscontro coi Luoghi Notevoli, la richiesta è girata alle API di Google. Nel caso si identifichi una Località, a livello segnalazione si comunicherà:

- nel campo DS.SIGNALING\_POSITION il nome della Località
- nei campi VL.SIGNALING\_POSITION\_LATITUDE e VL.SIGNALING\_POSITION\_LONGITUDE le coordinate di default 0.0

### A livello dei Tweet

Il MS analizza tutti e soli i tweet che hanno contribuito a generare la segnalazione e ricerca, al loro interno, un’indicazione semantica di una via, piazza, zona, corso, etc...

Per farlo utilizza le API di Google. Questa operazione può produrre 2 esiti distinti:

- Non esistono Geolocalizzazioni Semantiche all’interno del tweet
- La localizzazione specifica è trovata con le API di Google

## Il Motore Semantico

---

Nel primo caso, se non esistono Geolocalizzazioni Semantiche all'interno di un messaggio, il MS, a livello del tweet, comunicherà, nei campi VL\_POSITION\_LATITUDE e VL\_POSITION\_LONGITUDE, le coordinate di Geolocalizzazione del tweet, nel caso di tweet geolocalizzato, o le coordinate di default 0.0, nel caso di tweet non geolocalizzato.

Nel secondo caso il MS, interrogando le API di Google, trova un riscontro positivo e, quindi, identifica una Geolocalizzazione Semantica nel tweet stesso. A livello di quest'ultimo comunicherà, nei campi VL\_POSITION\_LATITUDE e VL\_POSITION\_LONGITUDE, le coordinate ricevute da Google.

Il risultato finale è che

- a livello segnalazione si ha un riferimento semantico, ed eventualmente anche coordinate, del luogo dell'emergenza
- a livello del singolo tweet si ha la sua eventuale geolocalizzazione (derivante da geolocalizzazione semantica o meno)

### 4.5.3.6 Utilizzo di Blacklist

Uno dei problemi sorti nell'utilizzo delle Google API nel tool di entity discovery consiste nel fatto che esistono molti termini comuni che corrispondono a nomi di città. Perciò, nel caso in cui tali termini siano spesso presenti all'interno di tweet contenenti la keyword in analisi, il tool, dopo aver effettuato la richiesta al servizio e ottenuta una risposta positiva, li considererebbe erroneamente come possibili nomi di luoghi. Inoltre è necessario evitare di richiedere troppe valutazioni di termini frequentemente utilizzati dalla gente, ma che si è sicuri non indicativi di un Luogo, per evitare di raggiungere la soglia massima di richieste.

Per risolvere questo problema è stata introdotta una tabella *blacklist* nel Database, contenente alcuni termini comuni corrispondenti a nomi di città e parole utilizzate oltre la norma. In questo modo il Tool, prima di effettuare la richiesta al servizio, controlla che la parola considerata non sia contenuta in essa e, nel caso in cui lo fosse, la scarta, indirizzando l'analisi verso il termine successivo.

## 4.5 Motore Semantico: Entity Discovery

L'utilizzo della *blacklist* non crea problemi nell'esecuzione dell'algoritmo. Infatti, è poco probabile che accada un evento nei luoghi identificati da termini in essa presenti, in quanto spesso si tratta di piccoli paesi o frazioni di città di dimensioni molto limitate e che su twitter non sarebbero utilizzati come Luogo di riferimento, prediligendo invece, come minimo, il comune di cui fanno parte.

Come già accennato, al fine di migliorare le prestazioni del Tool di Entity Discovery, sono state inseriti nella *blacklist* anche termini che non corrispondono a luoghi ma che spesso sono presenti all'interno dei tweet. L'esempio più emblematico viene dato dal "termine" `http`, il cui numero di occorrenze è molto elevato, data la numerosa presenza di link all'interno dei tweet.

Mostriamo nella tabella 4.7 esempi di termini contenuti nella *blacklist*.

Blacklist	
Termini comuni che corrispondono a nomi di città	Calcio, Amore, Vita, Ora, Tromba, Fallo, Piscina, Pero, Noci, Pigna, Gallo, Purgatorio
Termini spesso presenti all'interno dei tweet	Http, Scuola, Nave, Serva, Cmq, Xkè, Con

Tabella 4.7: Esempi di termini contenuti nella blacklist.

### 4.5.3.7 File di Configurazione

Per consentire all'amministratore del server TORCIA di apportare piccole modifiche a parametri fondamentali dell'Entity Discovery senza dover modificare alcun codice Java, abbiamo creato un file XML che, esterno al Tool, offre questa possibilità.

Ogni volta che il software è avviato, questi inizializza parte dei suoi valori base andando a leggerli dai relativi campi presenti all'interno di questo file, chiamato *ms-configuration.xml*.

Riportiamo qui sotto il file con valori di esempio.

## Il Motore Semantico

---

```
<?xml version="1.0" encoding="UTF-8"?>
<configuration_params>

  <db_address>localhost</db_address>
  <db_name>torcia</db_name>
  <db_username>username</db_username>
  <db_password>password</db_password>
  <numero_giorni_tweet>1</numero_giorni_tweet>
  <static_threshold>300</static_threshold>
  <cadenza_motore_semantico_minuti>30</cadenza_motore_semantico_minuti>
  <web_service_signal>http://localhost:9090/torcia/signal</web_service_signal>
  <web_service_pesi>http://localhost:9090/torcia/pesi</web_service_pesi>
  <basket_tweet_inviati>200</basket_tweet_inviati>
  <intervallo_max_segnalazione>7</intervallo_max_segnalazione>

</configuration_params>
```

Figura 4.15: Esempio del file ms-configuration.xml.

Come si vede, i parametri settabili sono:

- **DB\_ADDRESS** : indirizzo del Database in cui si trovano tutte le tabelle di appoggio per il funzionamento del Motore Semantico e a cui si vuole accedere.
- **DB\_NAME** : nome del Database.
- **DB\_USERNAME** : nome utente con cui si accede.
- **DB\_PASSWORD** : password necessaria per l'accesso.
- **NUMERO\_GIORNI\_TWEET** : indica il numero di giorni passati che l'Entity Discovery prende in considerazione al momento dell'esecuzione.
- **STATIC\_THRESHOLD** : permette di impostare a piacimento il valore della prima soglia controllata dall'Entity Discovery, che identifica o meno la presenza di un picco di parlato.
- **CADENZA\_MOTORE\_SEMANTICO\_MINUTI** : come verrà spiegato nella successiva sezione 4.6, è possibile avviare il Motore Semantico in modo che si esegua automaticamente dopo una determinata quantità di tempo, impostabile grazie a questo parametro.
- **WEB\_SERVICE\_SIGNAL** : indirizzo con cui contattare il web service di Beta80 per l'invio delle segnalazioni.
- **WEB\_SERVICE\_PESI** : indirizzo con cui contattare il web service di Beta80 per la ricezione della lista aggiornata di sorgenti e relativi pesi.

- `BASKET_TWEET_INVIATI` : indica il numero massimo di tweet inviati in una sola volta all'interno di una segnalazione. Un numero maggiore fa sì che la stessa segnalazione venga inviata in più parti.
- `INTERVALLO_MAX_SEGNALAZIONE` : intervallo temporale, in giorni, durante il quale una segnalazione avvenuta rimane attiva, anche senza causare del parlato degno di nota. La scelta di non disattivare subito le segnalazioni è dovuta alla possibilità di una momentanea perdita di interesse degli utenti verso l'accaduto in favore di un altro e della necessità di raccogliere informazioni relative a un'emergenza anche nell'immediato lasso temporale successivo alla stessa, come già sottolineato nella sezione 4.4 sul Classificatore.

## 4.6 Tool di Monitoraggio

Web Service e Motore Semantico sono entrambe applicazioni predisposte per essere eseguite in background. Il Servizio PUSH TWEET è un vero e proprio servizio web che rimane in ascolto sulla porta 8080 del server TORCIA, mentre i sorgenti Java del Motore Semantico sono stati wrappati per poter essere eseguiti come un Servizio di Windows. Per farlo è stato utilizzato il tool di Apache Commons denominato Procrun, grazie al quale il Motore Semantico verrà lanciato all'avvio del server e rimarrà in esecuzione proprio come un Servizio Windows standard.

Per entrambi i due blocchi funzionali ci sono dei meccanismi di log che consistono in una tabella del Database, per quanto riguarda il Servizio Push Tweet, e in un file di log *ms-log.log*, per il Motore Semantico. Per rendere più pratica la consultazione di tali registri, così come per poter agevolmente fermare e riavviare il Motore Semantico, abbiamo pensato di realizzare un software di monitoraggio chiamato Torcia Monitor.

Il software è un'applicazione Windows con interfaccia grafica sviluppata con il linguaggio C# e, oltre alle funzionalità appena citate, comprende dei grafici riguardanti statistiche sui dati ricevuti e sui risultati ottenuti dal Motore Semantico.

## Il Motore Semantico

In questa sezione descriveremo l'architettura di Torcia Monitor, le sue funzionalità per ogni pannello dell'interfaccia e proveremo a dare un elenco di suggerimenti per suoi possibili sviluppi futuri.

### 4.6.1 Architettura di Torcia Monitor

Uno degli aspetti fondamentali di Torcia Monitor è il suo totale disaccoppiamento con il codice sorgente degli altri blocchi funzionali.

Come si può notare dalla Figura 4.16, il Tool di Monitoraggio interagisce con i registri del Servizio Push Tweet e del Motore Semantico, restando completamente distaccato dai codici sorgenti dei due. Inoltre per lanciare o arrestare il Motore Semantico, Torcia Monitor esegue una nuova Console e da essa inserisce i comandi adatti per l'operazione desiderata. Ogni altra operazione di prelievo di informazioni dal Database, per effettuare delle statistiche, non implica alcuna interazione tra l'esecuzione del Web Service e del Motore Semantico.

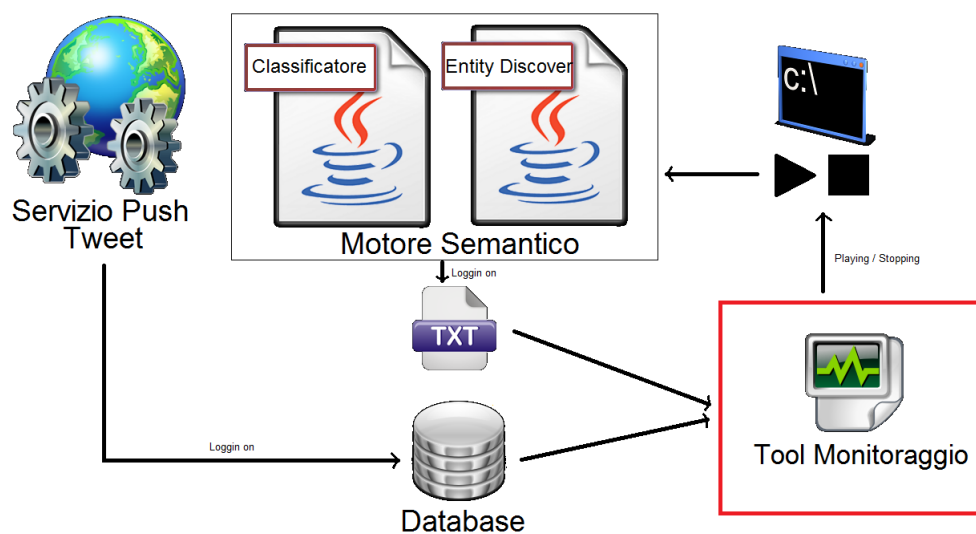


Figura 4.16: Schema architetturale di Torcia Monitor.

### 4.6.2 Monitoraggio del Motore Semantico

L'interfaccia principale del software di monitoraggio è sicuramente quella in cui viene visualizzato il registro delle attività del Motore Semantico.

Come si può notare in figura 4.17, l'utente che accede al server Torcia può visualizzare la lista dei messaggi registrati dal Motore Semantico, anche in tem-



## 4.6 Tool di Monitoraggio

po reale. I messaggi comprendono sia i dettagli della classificazione eseguita, sia quelli delle eventuali segnalazioni o allerte generate. Ovviamente in questa finestra sarà possibile visualizzare anche possibili messaggi di errore, dati da problemi che il Motore Semantico può incontrare durante la sua esecuzione.



Figura 4.17: Parametri di monitoraggio del Motore Semantico.

I comandi che si trovano in alto a destra dell'interfaccia di Torcia Monitor sono quelli per regolare l'esecuzione del Motore Semantico e la sua configurazione. Come si può vedere dalla figura 4.17 abbiamo:

1. La spia di ON/OFF, che indica in ogni istante se il Motore Semantico è avviato o meno.
2. L'identificativo del processo (PID). E' l'identificativo del processo del Motore Semantico, vale -1 se il Motore Semantico è spento.
3. Il tasto PLAY per lanciare il Motore Semantico eseguendolo automaticamente da una console.
4. Il tasto STOP per fermare il Motore Semantico, ovvero per interrompere il processo identificato dal PID.
5. Il tasto CLEAR LOG per svuotare il contenuto del registro del Motore Semantico. Questa è una funzionalità che è tornata utile soprattutto in fase di testing.

## Il Motore Semantico

---

Più a destra troviamo quelli che sono i parametri di configurazione del Motore Semantico. Questi andranno a modificare i campi del file di configurazione *ms-configuration.xml*, già descritto nella sezione 4.5.3.7, dal quale il Motore al primo avvio andrà a recuperare le informazioni per la sua modalità di esecuzione.

Più nel dettaglio i parametri che si possono personalizzare sono:

6. GIORNI INDIETRO: ovvero il numero di giorni passati, a partire dalla data corrente, che il Motore Semantico deve considerare per analizzare i tweet.
7. SOGLIA: il numero di tweet minimo tale per cui il Motore Semantico debba iniziare a ricercare una possibile emergenza o allerta. Se in un certo arco di tempo il numero di tweet prelevati dal web non supera la soglia, il Motore non inizia nemmeno l'analisi.
8. CADENZA: i minuti di distanza tra un'esecuzione del Motore Semantico e l'altra.

Il tasto (9) serve banalmente a salvare i parametri scelti, rigorosamente quando il Motore Semantico è fermo. Se i parametri vengono modificati durante l'esecuzione, non verranno presi in considerazione prima del prossimo arresto del Motore Semantico.

### 4.6.3 Monitoraggio del Servizio PUSH TWEET

Nella Figura 4.18 vediamo come l'utente di Torcia Monitor può consultare il registro delle attività del Servizio Push Tweet.

## 4.6 Tool di Monitoraggio

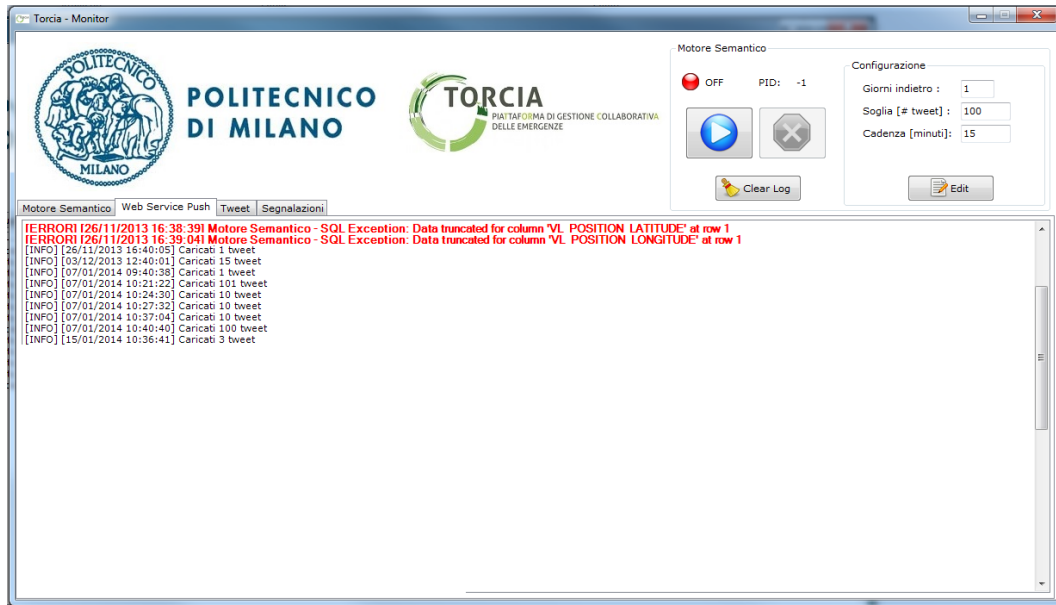


Figura 4.18: Monitoraggio del Web Service.

Il software va a recuperare i campi della tabella *log\_caricamento* e li elenca in una formattazione consona con quello di un registro degli eventi.

Si può notare come eventuali messaggi di errore del push dei tweet vengano stampati con il colore rosso, consentendo a chi consulta il registro un'immediata percezione dei problemi sorti. Anche nell'interfaccia descritta nella sezione precedente, eventuali eccezioni sollevate saranno colorate di rosso per facilitarne l'individuazione. Inoltre, così come per il registro del Motore Semantico, anche questo registro viene aggiornato in tempo reale.

### 4.6.4 Statistiche

Lo sviluppo di un'applicazione con interfaccia grafica ci ha portati a pensare che, al semplice monitoraggio delle attività, si potessero aggiungere dei pannelli per mostrare dati e statistiche interessanti riguardanti il nostro sistema. Per questo abbiamo pensato ad un paio di grafici che mostreremo nel dettaglio in questa sezione.

Con questo lavoro abbiamo gettato le basi per un Tool che può essere facilmente arricchito di peculiarità, soprattutto di carattere statistico, e che proveremo qui a suggerire per il futuro.

## Il Motore Semantico

### 4.6.4.1 Tweet

Il grafico di Figura 4.19 permette di vedere il numero di tweet presenti nel Database. E' possibile visualizzare i tweet degli ultimi 10, 20, 30, 50 o 100 giorni.

La cosa interessante di questa statistica è che saltano subito all'occhio dei picchi di parlato che sono ragionevolmente collegati a determinate emergenze. Questo grafico potrebbe essere un buon punto di partenza per poter studiare l'evoluzione del parlato, nell'ottica di realizzare una soglia di dinamica diversa da quella statica attuale.



Figura 4.19: Monitoraggio del numero di Tweet.

### 4.6.4.2 Segnalazioni

In Figura 4.20 è illustrato il grafico degli output del Motore Semantico.

L'utente può selezionare il periodo in cui vuol vedere quello che il Motore Semantico ha rilevato dal parlato di Twitter, sia in termini di allerte, sia in termini di segnalazioni.

In ascissa è rappresentato il tempo mentre in ordinata abbiamo i due possibili valori di pericolo: allerta (indicata in giallo) e segnalazione (indicata in rosso).

## 4.6 Tool di Monitoraggio



Figura 4.20: Monitoraggio di Segnalazioni ed Allerte.

Questo grafico permette di visualizzare le tempistiche con cui il Motore Semantico si accorge delle emergenze.

Si può notare come nel caso del nubifragio di Catania l'allerta sia sorta con poco anticipo rispetto alla segnalazione vera e propria; questo perché un nubifragio è un qualcosa che accade all'improvviso. Per quanto riguarda l'alluvione in Sardegna invece, il Motore Semantico ha rilevato un'allerta con più di 13 ore in anticipo rispetto all'emergenza vera e propria.

Questo poiché i vari acquazzoni, diffusi in tutta l'isola, hanno prodotto un parlato già consistente anche prima del verificarsi della vera e propria emergenza in serata.

### 4.6.4.3 Possibili sviluppi per Torcia Monitor

Torcia Monitor è un software che si presta molto bene all'aggiunta di nuove schede per il controllo di altre statistiche.

Vogliamo terminare questa sezione provando a consigliarne alcune, tutte sempre ottenibili dai dati disponibili all'interno del Database del Politecnico:

- Percentuale del numero di volte in cui il Motore Semantico esegue l'analisi senza trovare allerte o emergenze, nelle date in cui il numero di tweet supera la soglia statica prefissata.

## Il Motore Semantico

---

- Classifica degli utenti di twitter che hanno postato o retwittato il maggior numero di cinguettii per una certa emergenza.
- Categorie di classificazione più frequenti per una certa emergenza.
- Realizzazione di un pannello contenente una heat map sulla quale vengono mostrate le geolocalizzazioni dei tweet riguardanti una certa emergenza.
- Classifica dei comuni italiani in cui sono state rilevate il maggior numero di allerte o emergenze.
- Arco temporale in cui è rimasta attiva una segnalazione e, quindi, anche una classifica delle emergenze che sono durate di più.

## 4.7 Possibili nuove funzionalità

In questa sezione vogliamo proporre delle possibili funzionalità aggiuntive per uno sviluppo futuro del Motore Semantico di TORCIA. Queste idee ci sono sorte dal continuo lavoro di sviluppo e testing del nostro sistema, ma non sono state da noi portate avanti per motivi di tempo a disposizione o per il rispetto delle specifiche generali del progetto. Ci limitiamo quindi a dare dei consigli per chiunque lavorerà all'implementazione e alla manutenzione del Motore Semantico dopo di noi.

Il primo miglioramento che ci sentiamo di suggerire è l'inserimento di un'effettiva segnalazione delle Allerte rilevate, che fin ad ora non è implementata. Infatti in fase di analisi non era prevista questa categorizzazione delle segnalazioni da "ALLERTA" a "EMERGENZA". Per questo noi, per iniziare, abbiamo scritto l'algoritmo che la implementi, ma ci siamo limitati a rispettare i requisiti di progetto e a non inviare le segnalazioni targate come "ALLERTE".

Come descritto nella sezione 4.5, l'algoritmo di ricerca di un'Entity va ad interrogare un servizio di Google per ognuna delle parole più frequenti nel gruppo di tweet prelevato, nel caso non vi sia un riscontro nella tabella dei Luoghi Notevoli. Questo passaggio risulta essere critico per due motivi:

## 4.7 Possibili nuove funzionalità

---

1. Come già accennato nella sezione 4.5.3.5, Google riconosce Regioni e centri abitati italiani come “Luoghi” di una possibile emergenza, ma non riconosce, per esempio, i fiumi. Pertanto se l’esonazione di un fiume venisse denominata dal parlato di Twitter come “ESONAZIONE fiume\_in\_questione”, TORCIA sarebbe in grado di mandare una segnalazione solamente se quel determinato fiume coincide col nome di un centro abitato italiani.
2. Google dà un limite di 2500 query al giorno.

Per risolvere questi due problemi abbiamo pensato a due soluzioni:

1. Continuare a integrare la tabella *luoghi\_notevoli* con una lista molto ampia di possibili zone da analizzare, cercando di bypassare il più possibile le chiamate al servizio di Google e migliorando così sempre di più la precisione e l’efficacia del Motore semantico.
2. Utilizzare un servizio appositamente creato dai cartografi coinvolti nel progetto, che fornisca le stesse funzionalità delle API di Google, ma non abbia un limite sul numero di interrogazioni giornaliere.

Anche se il limite di interrogazioni è alto, non averne più uno sarebbe una sicurezza aggiuntiva.

Un ulteriore sviluppo riguarda poi il Classificatore. Come descritto nella sezione 4.4, il Tool di Classificazione può essere costantemente migliorato, facendo del testing sui tweet di ogni giorno e perfezionando il contenuto della tabella meta-info. Risulta necessario quindi, come per la tabella dei Luoghi Notevoli, un continuo lavoro di affinamento anche in fase di funzionamento attivo del Motore Semantico.

In fase di testing poi, ci siamo accorti che alcune segnalazioni prendono un nome univoco per tutti gli utenti di Twitter solo dopo un certo lasso di tempo. Questo significa che il Motore Semantico di TORCIA potrebbe inviare nel tempo delle segnalazioni con nome diverso, ma che in realtà sono legate alla stessa emergenza. Come vedremo nel caso dell’alluvione in Sardegna del mese di Novembre 2013, il parlato di Twitter generava delle allerte

## Il Motore Semantico

---

riguardanti la SARDEGNA e altre riguardanti OLBIA, entrambe palesemente riconducibili alla stessa emergenza. Per evitare quindi la segnalazione di “ALLUVIONE OLBIA” e successivamente quella di “ALLUVIONE SARDEGNA”, si potrebbe pensare di aggiungere una funzionalità al Motore Semantico che permetta all’ipotetico Operatore di Protezione Civile di legare le due segnalazioni rilevate per il futuro. Più tecnicamente pensiamo ad un web service interrogabile dal Cruscotto di Beta80 che modifichi i campi delle segnalazioni da noi generate.

Altra miglioria riguarda la Blacklist. Esistono molti Comuni italiani che hanno il nome che coincide con parole del linguaggio comune e questo risulta un problema quando si diffondono tweet contenenti una delle nostre keyword e una di quelle parole. La politica da noi scelta in questi casi è quella di non considerare i Comuni in questione nell’analisi semantica, aggiungendoli alla Blacklist. Si può intuire quindi che, per evitare di generare inutili segnalazioni, risulta fondamentale la qualità della compilazione della tabella blacklist. Per rendere più preciso il nostro sistema, abbiamo pensato che si potrebbe dare la possibilità all’utente TORCIA di avvisarci che la segnalazione sorta è dovuta all’ambiguità di una certa parola piuttosto che ad una reale situazione critica. Tecnicamente anche in questo caso pensiamo ad un web service interrogabile dal Cruscotto di Beta80 che possa aggiungere parole alla nostra Blacklist.

Come detto all’inizio di questo paragrafo, tutte queste possibili innovazioni sono state pensate soprattutto in fase di testing. Nel prossimo capitolo andremo ad illustrare quella che è stata la struttura e il risultato di tale testing, nel quale abbiamo eseguito un’analisi semantica sui tweet da noi raccolti da Gennaio 2013 a Febbraio 2014.



# Capitolo 5

## Verifiche Empiriche e Miglioramenti

In questo capitolo presentiamo i casi studio da noi analizzati, spiegando le modalità con cui abbiamo effettuato i test e le conclusioni che ne abbiamo tratto.

Come si vedrà meglio in seguito, ci siamo qui concentrati esclusivamente sulla parte del Motore Semantico relativa all'identificazione di allerte e segnalazioni, ovvero il Tool di Entity Discovery. Questo poiché le modalità di testing del Tool di Classificazione, con alcuni esempi significativi, sono già state trattate nel relativo capitolo 4.4 e nulla di rilevante può essere aggiunto.

In particolare abbiamo che:

La sezione 5.1 mostra i test effettuati per tutti i mesi dell'anno 2013;

La sezione 5.2 contiene le analisi nel dettaglio delle allerte e delle segnalazioni rilevanti, identificate nella sezione precedente;

Infine la sezione 5.3 è una riflessione critica sul lavoro di testing svolto, per evidenziare i principali problemi riscontrati in questa fase di progetto e le soluzioni e le modifiche da noi adottate per risolverli.

### 5.1 Analisi mensile

In questa sezione esponiamo, con una visione di livello generale, i risultati ottenuti dopo aver eseguito il nostro codice su ogni giornata dell'anno 2013.

Dal momento che il Motore Semantico è stato progettato per analizzare dati in tempo reale, per riadattarlo al nostro scopo, vale a dire un'analisi di un campione molto ampio di dati storici, siamo stati costretti ad operare una semplificazione dettata da esigenze temporali: invece che eseguire il Tool di Entity Discovery ogni 15-30 minuti, come è consigliabile, abbiamo preso intervalli di tempo giornalieri, lanciando l'analisi sui tweet delle 24 ore che intercorrono tra le 00:00 del giorno di riferimento e le 00:00 del giorno successivo.

I dati raccolti dimostrano, tuttavia, che questa semplificazione non ha prodotto alcuna limitazione al Motore Semantico, che è stato comunque in grado di rilevare le eventuali allerte e segnalazioni occorse durante le varie giornate. Queste stesse emergenze saranno poi oggetto di un esame approfondito nella prossima sezione in cui, come vedremo, saranno analizzate eseguendo il software passo passo, a intervalli di 15 minuti.

I risultati ottenuti sono qui di seguito presentati secondo lo schema:

- **Grafico dei Tweet**; rappresenta il numero di tweet giornalieri crawlati.
- **Grafico di Allerte e Segnalazioni**; visualizza graficamente le eventuali segnalazioni (rosso) e/o allerte (giallo) attive, rilevate dal software.
- **Commenti ai grafici**; questa parte contiene riflessioni riguardo i dati raccolti nel mese, con particolare riferimento ad eventuali picchi di parlato ed ai problemi riscontrati.

### 5.1.1 Gennaio 2013

**Grafico dei Tweet:**

## 5.1 Analisi mensile



Figura 5.1: Numero di Tweet nelle giornate di Gennaio 2013.

### Grafico di Allerte e Segnalazioni:

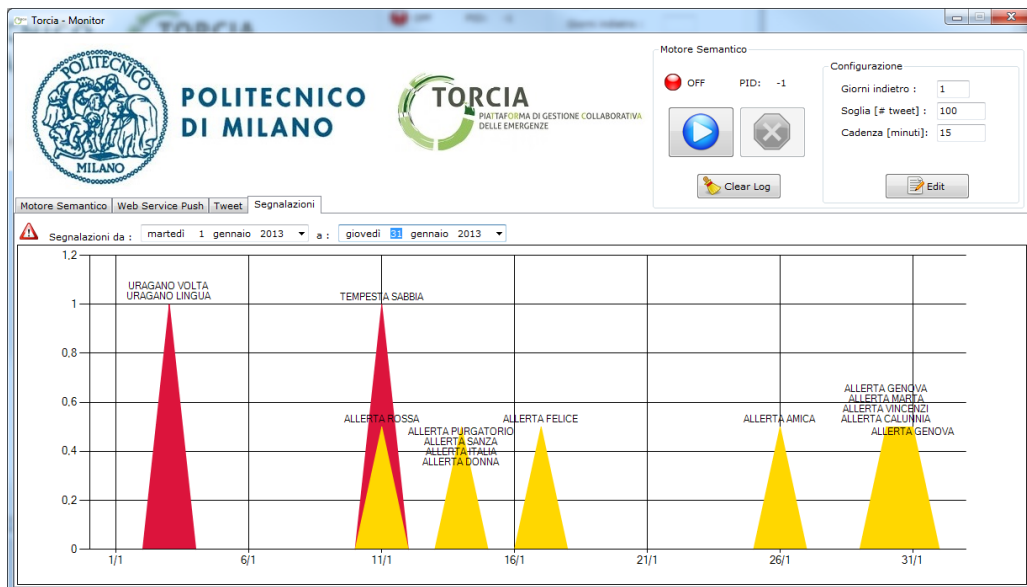


Figura 5.2: Allerte e Segnalazioni di Gennaio 2013.

### Commenti ai Grafici:

#### 3 Gennaio 2013

Il picco di parlato, e la relativa segnalazione, è dovuto ad un massiccio retweet (circa 500 cinguettii) di un messaggio dell'utente @harryismyoxygen

## Verifiche Empiriche e Miglioramenti

---

contenente il termine “*Uragano*” e le parole “*Lingua*” e “*Volta*”. Queste ultime due sono state segnalate dall’Entity Discovery in quanto *Lingua* corrisponde ad una località delle isole Lipari, mentre *Volta* si trova in provincia di Lecco. L’azione correttiva da noi intrapresa è stata il blacklistaggio di questi due termini, in quanto ambigui.

Blacklist: *Lingua* (Isole Lipari), *Volta* (Lecco)

### 10-11 Gennaio 2013

Segnalazione ed allerta sono dovute a due grosse tempeste di sabbia. La prima con origine a Gerusalemme, la seconda (una tempesta di sabbia rossa) in Australia. Giustamente il nostro motore semantico non ha segnalato un'emergenza in questi due luoghi, essendo fuori dall'Italia, ma ha avuto dei riscontri con “*Sabbia*” e “*Rossa*”.

Blacklist: *Sabbia* (Vercelli), *Rossa* (Vercelli)

### 14 Gennaio 2013

Il picco di allerte è dovuto all’ingente retweet di una delle frasi più popolari di Dante Alighieri su Twitter, tratta dal sesto canto del Purgatorio: “*Ahi serva Italia, di dolore ostello, nave senza nocchiere in gran tempesta, non donna di province, ma bordello!*”. I riscontri sono sempre dovuti al fatto che i termini creano ambiguità con località italiane, salvo la parola “*Italia*” che, invece, genera riscontro in quanto fa semplicemente riferimento al nostro stato intero.

Blacklist: Purgatorio (Trapani), *Sanza* (Salerno), *Donna* (Alessandria), *Italia*

### 15-16 Gennaio 2013

Picco di parlato dovuto a grandinate diffuse in tutta Italia, specialmente a Roma ed in Sicilia. Il parlato non ha però prodotto allerte o segnalazioni specifiche di un luogo in quanto i tweet con riferimento ad una precisa località erano molto pochi (ad esempio i tweet contenenti “*Roma*” sono inferiori al 4%).

### 17-18 Gennaio 2013

Più di 100 retweet di una frase dell’utente @aryanistic contenente le parole “*Uragano*” e “*Felice*”. Come sempre il riscontro è dovuto all’ambiguità col nome di una località italiana.

## 5.1 Analisi mensile

Blacklist: Felice (Torino)

### 26 Gennaio 2013

Allerta generata dal massiccio retwittaggio di una frase dell'utente @bjebersbed contenente le parole “*Allagamento*” e “*Amica*”.

Blacklist: Amica (Cosenza).

### 30-31 Gennaio 2013

Allerte dovute alla diffusione della notizia dell'apertura dell'indagine per calunnia e falsa testimonianza nei confronti dell'ex sindaco di Genova Marta Vincenzi, in carica durante l'alluvione di Genova del 2011.

Blacklist: Marta (Viterbo), Vincenzi (Rimini), Calunnia (Biella)

## 5.1.2 Febbraio 2013

Grafico dei Tweet:



Figura 5.3: Numero di Tweet nelle giornate di Febbraio 2013.

Grafico di Allerte e Segnalazioni:

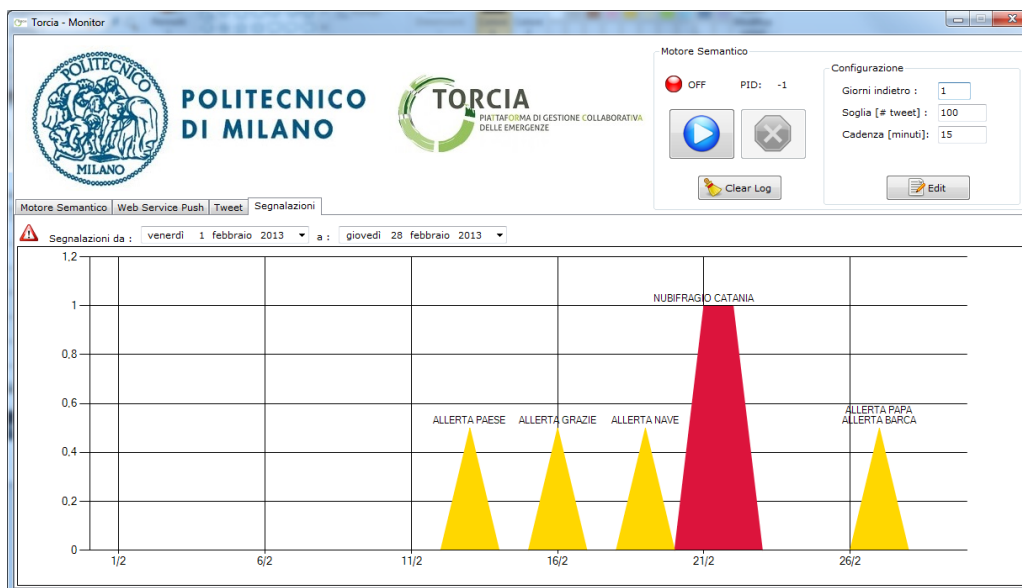


Figura 5.4: Allerte e Segnalazioni di Febbraio 2013.

### Commenti ai Grafici:

#### 7-12 Febbraio 2013

I picchi di parlato si riferiscono a numerosi tweet relativi all'imminente arrivo della tempesta di neve "Nemo" nel Nord-Est degli USA. Il Motore Semantico, a ragione, non ha individuato alcuna entità.

#### 13 Febbraio 2013

Gran numero di retweet per una dichiarazione dal mondo politico: "il paese sta franando".

Blacklist: Paese (Treviso)

#### 16 Febbraio 2013

Picco di retweet per un messaggio polemico di un ironico indignato che ringraziava le autorità politiche per essersi dimenticate degli alluvionati in maremma.

Blacklist: Grazie (Mantova)

#### 19 Febbraio 2013

## 5.1 Analisi mensile

Nuovamente la citazione di Dante già vista al 14 Gennaio. In precedenza, parte dei retweet conteneva la parola “*nave*” scritta in maniera errata e, quindi, non vi era stato un riscontro, come invece in questo caso.

Blacklist: Nave (Brescia)

### 21-22 Febbraio 2013

Identificata l’Entity “*Nubifragio Catania*”. Questa emergenza sarà oggetto del nostro studio approfondito nella relativa sezione 5.2.1.

### 21-22 Febbraio 2013

Retweet popolare di una frase del Papa contenente le parole “*Tempesta*” e “*Barca*”.

Blacklist: Papa (Varese), Barca (Alessandria)

## 5.1.3 Marzo 2013

Grafico dei Tweet:



Figura 5.5: Numero di Tweet nelle giornate di Marzo 2013.

Grafico di Allerte e Segnalazioni:

## Verifiche Empiriche e Miglioramenti



Figura 5.6: Allerte e Segnalazioni di Marzo 2013.

### Commenti ai Grafici:

#### 1 Marzo 2013

Piccola e breve ondata di maltempo sulle regioni insulari (circa 50 tweet per la Sardegna, 70 per la Sicilia).

#### 9 Marzo 2013

Allerte generate da una frase a sfondo politico di Beppe Grillo contenente la parola “*Tempesta*”. Il termine “*Mappa*” non corrisponde ad alcuna località, ma è un termine chiave di Google che restituisce la mappa del proprio stato.

Blacklist: Grillo (Como), Giallo (Brescia), Cittadina (Milano), Mappa

#### 11-12 Marzo 2013

Maltempo diffuso su tutta la penisola; in particolare il primo giorno si è riscontrata una situazione di possibile rischio nella zona di Torino, a causa di una forte grandinata.

#### 17 Marzo 2013

Dichiarazioni di sfondo politico.



Blacklist: Fondi (Latina)

18 Marzo 2013

Diffusione virale del tweet di uno studente scontento: “*La mia scuola rimane sempre aperta: pioggia, neve, grandine, terremoto, eruzione vulcanica, tsunami, pioggia di meteoriti, invasione aliena*”.

Blacklist: Aliena (Perugia)

23 Marzo 2013

Primo reale falso allarme riscontrato dal motore semantico su cui non è possibile effettuare alcun intervento. L’allarme è stato lanciato a causa del massiccio retweet di un messaggio di una fan del cantante Justin Bieber. Quest’ultimo avrebbe tenuto un concerto a Bologna in quei giorni e la fan, non potendo andarci, ha espresso la sua delusione scrivendo di come stava “*Allagando*” la casa di lacrime.

In questo caso non è possibile intraprendere alcuna azione correttiva, in quanto Bologna è una città grande e possibile parola che gli utenti di Twitter potrebbero utilizzare in caso di emergenza in zona.

25 Marzo 2013

Marcia di protesta per le vittime, e superstiti dimenticati, dell’alluvione di Albinia del Novembre 2012. Nessuna azione da intraprendere in quanto l’alluvione era effettivamente accaduta.

26-27 Marzo 2013

Secondo falso allarme riscontrato dal motore semantico e nuovamente per colpa di Justin Bieber. Parte del post retwittato: “*Bologna sta piangendo, Justin Bieber sta partendo. Verona sta per tremare, gli One Direction stanno per tornare*”. Come nel caso precedente, anche qui sono impossibili azioni correttive, in quanto Verona è una città grande e possibile parola che gli utenti di Twitter potrebbero utilizzare in caso di emergenza in zona.

30 Marzo 2013

Segnalazione dovuta ad una gran quantità di retweet (quasi 350) di una citazione letteraria contenente il termine “*Uragano*”.

## Verifiche Empiriche e Miglioramenti

Blacklist: Piano (Sondrio), Mano (Bergamo)

31 Marzo 2013

70 inspiegabili retweet della frase “*Marzo pazzerello, se esce il sole fa una grandinata che si porta via le case, manco fa rima*”.

Blacklist: Manco (Cosenza), Rima (vercelli)

### 5.1.4 Aprile 2013

Grafico dei Tweet:



Figura 5.7: Numero di Tweet nelle giornate di Aprile 2013.

Grafico di Allerte e Segnalazioni:

## 5.1 Analisi mensile



Figura 5.8: Allerte e Segnalazioni di Aprile 2013.

### Commenti ai Grafici:

#### 2-5 Aprile 2013

Picco di parlato dovuto a un inondazione su vasta scala in Argentina, giustamente non segnalata dal nostro Motore Semantico.

#### 15 Aprile 2013

Allerta per un potenziale rischio emergenza nel parmense dovuto ad alcune piccole frane provocate dal maltempo.

#### 19-20 Aprile 2013

Maltempo in tutta la penisola e, nella seconda giornata, forti piogge soprattutto in Lombardia, con rischio di esondazione del Seveso e allerta delle squadre di Protezione Civile locali.

Blacklist: Squadre (Bergamo)

#### 21 Aprile 2013

Retweet di successo contenente le keyword “*Tempesta*” e “*Uragano*”.

#### 22 Aprile 2013

## Verifiche Empiriche e Miglioramenti

Diffusione del tweet politico dell'utente @fandiben: “*se su internet un uccellino sbatte le ali, un uragano si scatena sul PD*”.

Blacklist: Uccellino (Ferrara)

28-30 Aprile 2013

Picchi di parlato dovuti a maltempo negli Stati Uniti, in Asia e, soprattutto, ad un uragano di enormi dimensioni sul pianeta Saturno.

### 5.1.5 Maggio 2013

Grafico dei Tweet:



Figura 5.9: Numero di Tweet nelle giornate di Maggio 2013.

Grafico di Allerte e Segnalazioni:

## 5.1 Analisi mensile

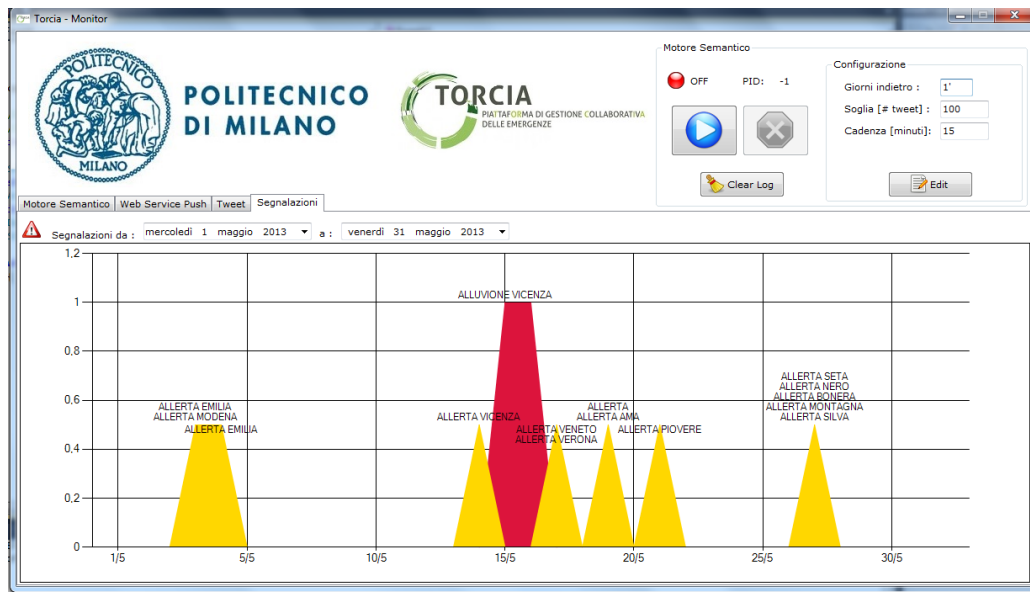


Figura 5.10: Allerte e Segnalazioni di Maggio 2013.

### Commenti ai Grafici:

#### 3-4 Maggio 2013

Allerta per il forte maltempo e le grandinate, specialmente nella zona di Modena.

#### 7 e 11 Maggio 2013

Giorni isolati caratterizzati entrambi da massicce grandinate sparse sulla penisola.

#### 14-17 Maggio 2013

Significative precipitazioni nel Veneto, con particolare riferimento alla zona di Vicenza. Il motore semantico ha, con buon anticipo, dapprima rilevato un'allerta, poi trasformatasi in segnalazione. Questa sarà nostro oggetto di studio approfondito nella relativa sezione 5.2.2.

#### 19 Maggio 2013

Impressionante diffusione virale del tweet dell'utente @andreavianel: *“Prima di essere inondato di tweet di protesta ribadisco che Rai3 ama #gazebo e non lo “chiude”. Era previsto così ma ripartirà presto, mi sa”*.

## Verifiche Empiriche e Miglioramenti

Blacklist: Era (Venezia), Ama (Bergamo)

21-22 Maggio 2013

Commento sarcastico riguardo al maltempo, condiviso dall'account dell'utente @Iddio: *“Va bene, va bene, rispetterò la legge secondo cui non può piovere per sempre: metterò un po' di grandinate come intermezzo.”*

Blacklist: Piovere (Brescia)

27 Maggio 2013

Allerte generate dalla diffusione di due tweet a sfondo calcistico scritti dall'utente @MilanistiOrld: *“Rijkaard- uragano frankie, figlio di eolo, cigno nero — Gullit- treccia nera, tulipano nero — Basten- piedi di seta, cigno diutrecht”* e *“Thiago silva- grandine nera — Nesta- tempesta perfetta — Bonera- montagna di luce — Yepes- apocalipto #soprannomi”*.

Blacklist: Seta (Reggio emilia), Nero (Torino), Bonera (Varese), Montagna (Padova), Silva (Torino)

### 5.1.6 Giugno 2013

Grafico dei Tweet:



Figura 5.11: Numero di Tweet nelle giornate di Giugno 2013.

Grafico di Allerte e Segnalazioni:

## 5.1 Analisi mensile



Figura 5.12: Allerte e Segnalazioni di Giugno 2013.

### Commenti ai Grafici:

#### 2-3 Giugno 2013

Maltempo e alluvioni in tutta l'Europa Centrale, specialmente nelle zone del Danubio e di Praga, dove un'alluvione ha provocato 5 morti.

Blacklist: Europa (Brescia), Centrale (Bergamo), Praga (Alessandria)

#### 9-10 Giugno 2013

Grandinate e maltempo in tutta Italia, con particolare riferimento a una violenta grandinata a Pescara seguita da un forte nubifragio. La situazione è però tornata alla normalità in breve tempo.

#### 12-19 Giugno 2013

I Server del Politecnico di Milano sono rimasti spenti per opere di manutenzione. Il Crawler, di conseguenza, non ha potuto raccogliere dati.

#### 20 Giugno 2013

Server e Crawler sono stati riaccessi verso sera. Essendo il numero di tweet crawlati molto esiguo, è bastato un retweet di soli 10 messaggi per generare due allerte. L'utente è di nuovo @Iddio e questa volta scrive: "Maltempo in

## Verifiche Empiriche e Miglioramenti

Francia: Lourdes allagata, acqua santa ovunque. Si consiglia ai vampiri e ai satanisti di non uscire di casa”.

Blacklist: Francia (Torino), Ginevra (Brescia)

### 24-27 Giugno 2013

Maltempo su tutta la costa romagnola, con particolare riferimento alla zona di Rimini. Nonostante ciò, in nessuno dei tre giorni si è arrivati ad una soglia di allarme e, infatti, la situazione è tornata alla normalità senza conseguenze.

### 30 Giugno 2013

Sviluppi giuridici legati al caso dell'Alluvione a Genova del 2011 già precedentemente incontrata (30 Gennaio 2013).

## 5.1.7 Luglio 2013

Grafico dei Tweet:



Figura 5.13: Numero di Tweet nelle giornate di Luglio 2013.

Grafico di Allerte e Segnalazioni:



## 5.1 Analisi mensile



Figura 5.14: Allerte e Segnalazioni di Luglio 2013.

### Commenti ai Grafici:

#### 7-8 Luglio 2013

Violento ed improvviso nubifragio sulla capitale. La stazione di Roma Termini è stata chiusa per il maltempo. Questa emergenza sarà oggetto del nostro studio nella relativa sezione 5.2.3.

Blacklist: Termini (Imperia), Chiusa (Bolzano), Stazione (Pavia)

#### 10-14 Luglio 2013

Maltempo diffuso su tutta la penisola. In particolare risaltano due grandinate avvenute rispettivamente a Perugia e a Torino, ma senza gravi conseguenze o grossi disagi.

#### 21-22 Luglio 2013

Altro violento ed improvviso, nonché altrettanto breve, nubifragio a Roma. Lo analizzeremo più nel dettaglio nella sezione 5.2.4.

#### 29 Luglio 2013

Impressionante quantità di tweet relativa al maltempo che sferza il Nord e Centro Italia, con particolari riferimenti a delle grandinate nel Piemonte e,

## Verifiche Empiriche e Miglioramenti

nello specifico, a Torino. Nonostante ciò, in nessuna località si è arrivati ad una soglia di allarme.

### 5.1.8 Agosto 2013

Grafico dei Tweet:



Figura 5.15: Numero di Tweet nelle giornate di Agosto 2013.

Grafico di Allerte e Segnalazioni:

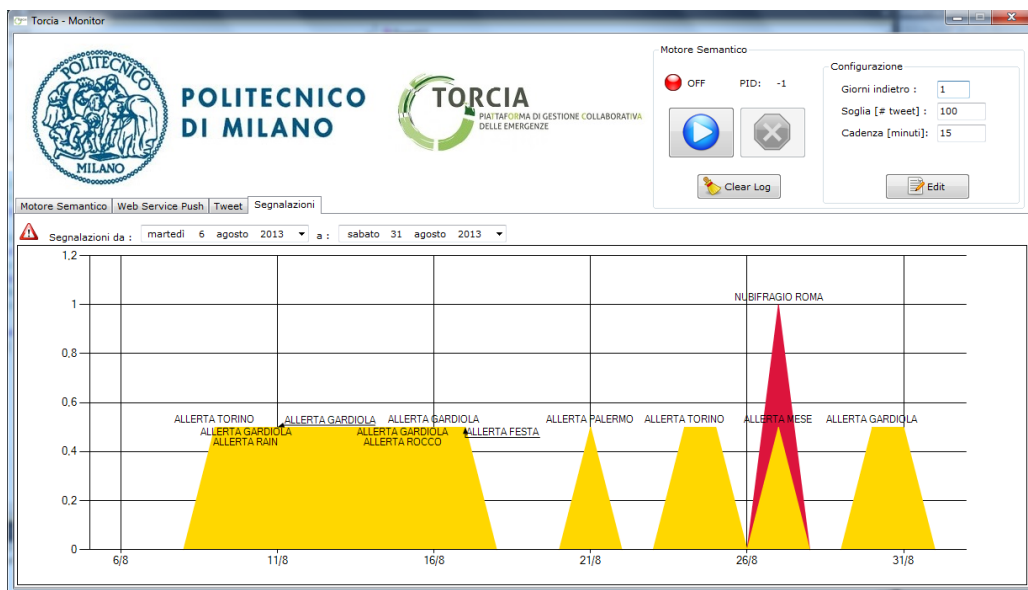


Figura 5.16: Allerte e Segnalazioni di Agosto 2013.

### Commenti ai Grafici:

#### 8-9 Agosto 2013

Maltempo diffuso su tutta la penisola, con particolare rilievo nella zona Nord-Ovest. A Torino altro nubifragio con grandinata, ma senza rischi.

#### 10 Agosto 2013 - ...

Installato a Gardiola, località in provincia di Torino, uno strumento di monitoraggio di fenomeni franosi. L'apparecchiatura è stata resa operativa dall'Istituto di Ricerca per la Protezione Idrica (IRPI), sede di Torino, col fine di studiare la situazione geologica di questa zona a rischio frane e raccogliere dati utili a fini preventivi. Parte dei dati acquisiti vengono comunicati, a fine informativo, mediante dei tweet, il cui numero varia tra i 30 e i 70 al giorno. I messaggi sono del tipo: *“Loc. Gardiola (TO), Monitoraggio Frane Val Germagnasca - Rain: 36.32 mm #iotnet”*.

Nel caso quindi che il numero di notifiche derivate dall'apparecchiatura sia più alto della norma o che il numero di tweet crawlati sia basso, tipicamente inferiore ai 600-700, queste registrazioni possono dar luogo ad una continua serie di allerte. Non abbiamo comunque ritenuto giusto blacklistare la località “Gardiola” in quanto essa non è usata impropriamente, non avendo alcuna valenza ambigua.

Blacklist: Rain( Bolzano)

#### 15 Agosto 2013

Retweet di un intervento radiofonico di Rocco Iannì.

Blacklist: Rocco (Torino)

#### 17 Agosto 2013

Allerta prodotta da una tragica alluvione nello Yemen durante una festa di nozze che ha causato la morte di 27 persone e 41 dispersi.

Blacklist: Festa (Vicenza)

#### 20-21 Agosto 2013

Nubifragi diffusi in tutto il Sud Italia, con particolare riferimento alla Sicilia e a Palermo. L'allerta è però rientrata in poche ore.

## Verifiche Empiriche e Miglioramenti

24-25 Agosto 2013

Grandine e forti piogge sul capoluogo piemontese, ma di scarsa portata.

26-28 Agosto 2013

Violento nubifragio si abbatte nuovamente sulla capitale. Sarà oggetto del nostro esame approfondito nella sezione 5.2.5.

L'allerta generata è invece dovuta alla presenza di numerosi retweet del post: “RT @guglielmoscilla: Il prossimo mese ci impegniamo a non inondare le bacheche dei nostri amici con lo status “wake me up when September ends...””.

Blacklist: Mese (Sondrio)

### 5.1.9 Settembre 2013

Grafico dei Tweet:



Figura 5.17: Numero di Tweet nelle giornate di Settembre 2013.

Grafico di Allerte e Segnalazioni:

## 5.1 Analisi mensile

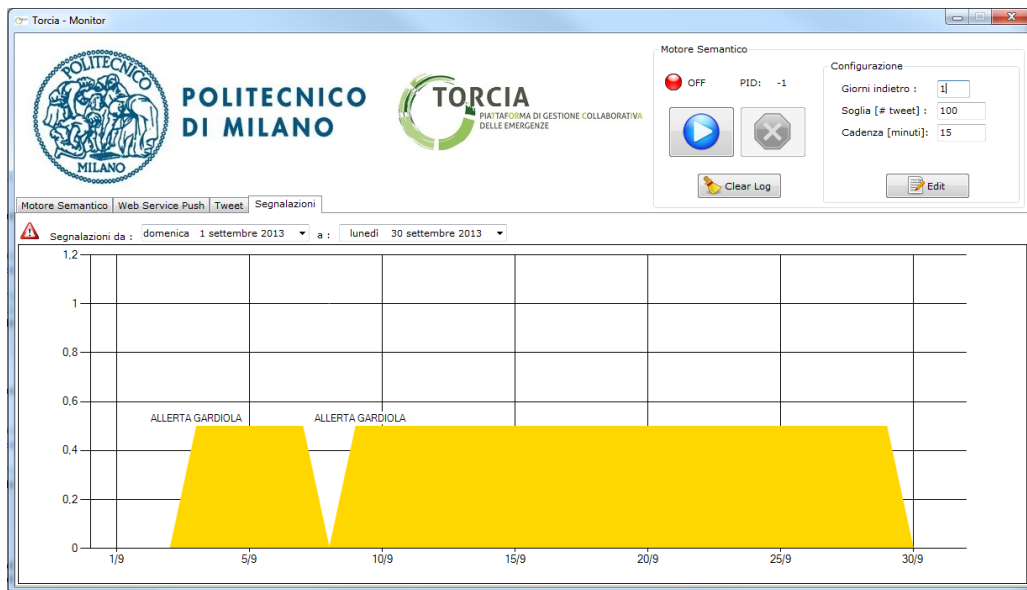


Figura 5.18: Allerte e Segnalazioni di Settembre 2013.

### Commenti ai Grafici:

#### 10 e 15 Settembre 2013

Picchi di parlato per l'arrivo di uragani in America e Messico.

#### 29-30 Settembre 2013

Maltempo in tutta la Toscana caratterizzato da una tromba d'aria, forti piogge e pericolo di allagamenti. Già dal secondo giorno però il maltempo si delocalizza, estendendosi a tutto il Centro-Sud Italia e facendo ritirare il pericolo.

### 5.1.10 Ottobre 2013

#### Grafico dei Tweet:

## Verifiche Empiriche e Miglioramenti



Figura 5.19: Numero di Tweet nelle giornate di Ottobre 2013.

## Grafico di Allerte e Segnalazioni:

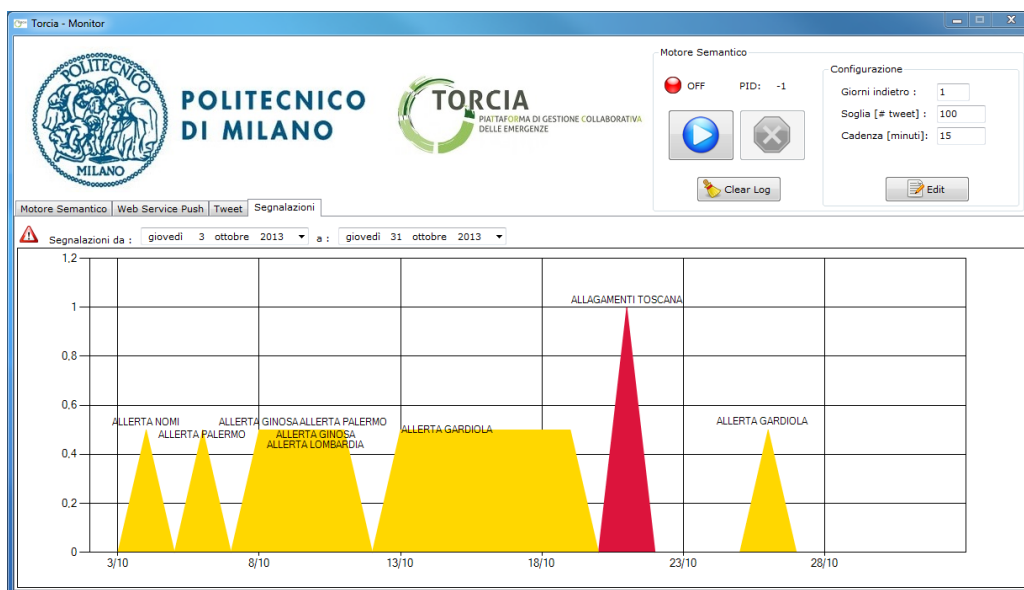


Figura 5.20: Allerte e Segnalazioni di Ottobre 2013.

## Commenti ai Grafici:

### 1 Ottobre 2013

Parlato riferito a maltempo in tutto il Centro-Sud Italia, senza localizzazioni specifiche.

### 4 Ottobre 2013

Gran numero di ricondivisioni del tweet “*RT @UomoMedievale: “Domani arriva Penelope: nubifragi e trombe aria”. Dar nomi a li temporali son le occupationi di moderni che non hanno un ...”*”.

Blacklist: Nomi (Trento)

### 5 Ottobre 2013

Diffusione di messaggi di allarme per l’arrivo del ciclone Penelope, in cui si consiglia di prestare attenzione per la prossima settimana.

### 6-12 Ottobre 2013

Il ciclone Penelope colpisce l’Italia e forti nubifragi si abbattono sulla penisola, in particolare a Palermo e in Lombardia. Nel tarantino, a Ginosa, le forti piogge hanno provocato un’alluvione che ha causato 4 morti.

### 21 Ottobre 2013

Nubifragio notturno in Toscana caratterizzato da allagamenti e frane. Sarà oggetto del nostro studio approfondito nel capitolo 5.2.6.

### 22-24 Ottobre 2013

Parlato generico sul recente maltempo e varie frasi di stampo politico.

### 28 Ottobre 2013

Picco di parlato in riferimento alla tempesta di “San Giuda” che si sta per abbattere su Francia ed Inghilterra.

## **5.1.11 Novembre 2013**

**Grafico dei Tweet:**

## Verifiche Empiriche e Miglioramenti



Figura 5.21: Numero di Tweet nelle giornate di Novembre 2013.

## Grafico di Allerte e Segnalazioni:

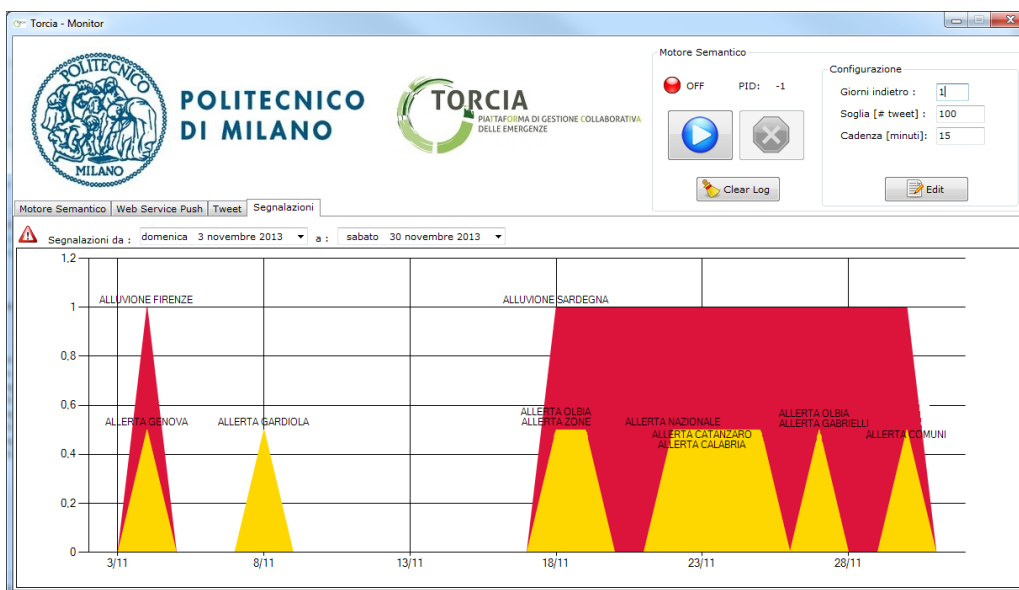


Figura 5.22: Allerte e Segnalazioni di Novembre 2013.

## Commenti ai Grafici:

### 4 Novembre 2013

Numerosi post in ricordo delle alluvioni che colpiscono Firenze e Genova rispettivamente nel 1966 e nel 2011.



### 9-13 Novembre 2013

Picchi generati da varie cause quali una tempesta nelle Filippine, post di rabbia per il maltempo diffuso su tutta Italia e retweet di massa del messaggio “*RT @FraseDiTumblr: Aveva la forza di un uragano, ma era debole come una foglia in autunno*”.

### 18-30 Novembre 2013

Maltempo in tutta la Sardegna, con forti piogge che hanno provocato un'alluvione di grosse dimensioni. In particolare ad Olbia la situazione è critica. Il 22 è stato indetto il lutto nazionale per le vittime e i disastri dell'alluvione. Questa emergenza verrà da noi analizzata nello specifico nella sezione 5.2.7.

Blacklist: Zone (Brescia), Nazionale (Bergamo)

### 23-25 Novembre 2013

Calmatasi la situazione in Sardegna, si inizia a parlare anche di altre zone colpite in maniera minore dal maltempo, in particolare la Calabria e la zona di Catanzaro.

### 27 Novembre 2013

Accuse pubbliche del prefetto Gabrielli rivolte a chi ha concesso l'edificabilità in alcune zone alluvionate che, a detta sua, sarebbero state palesemente a rischio e non adatte alla costruzione di abitazioni.

### 30 Novembre 2013

Annuncio dello stop di richiesta di versamenti al fisco per i comuni colpiti dall'alluvione.

Blacklist: Comuni (Treviso)

## **5.1.12 Dicembre 2013**

**Grafico dei Tweet:**

## Verifiche Empiriche e Miglioramenti



Figura 5.23: Numero di Tweet nelle giornate di Dicembre 2013.

## Grafico di Allerte e Segnalazioni:

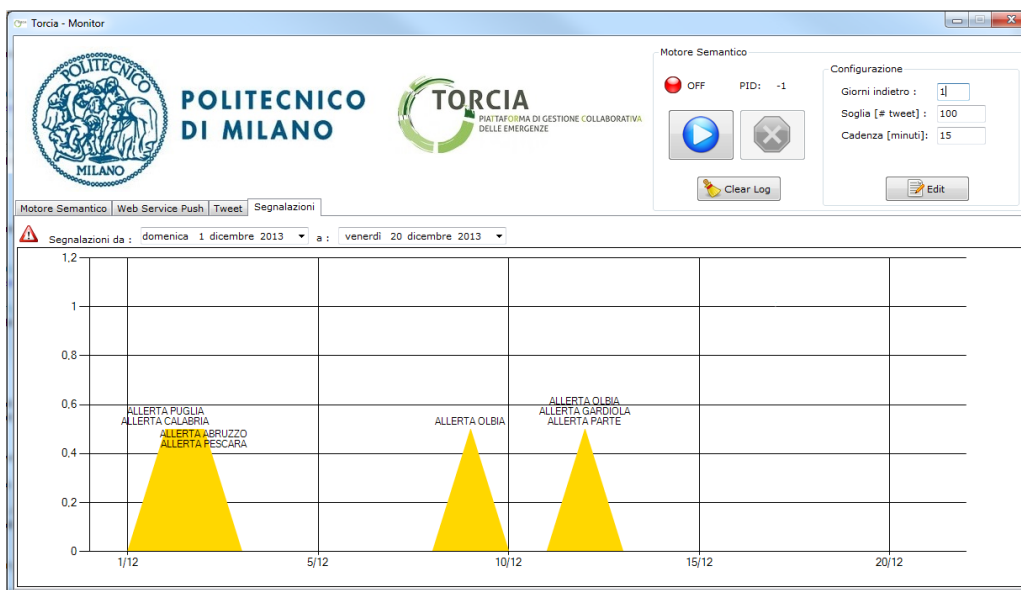


Figura 5.24: Allerte e Segnalazioni di Dicembre 2013.

## Commenti ai Grafici:

### 1-2 Dicembre 2013

Nuova ondata di maltempo sul Sud e poi Centro Italia, con località più colpite Puglia, Calabria, Abruzzo e Pescara.

## 5.2 Analisi di Allerte e Segnalazioni

---

### 3-7 Dicembre 2013

Maltempo delocalizzato su tutta la penisola, specialmente centro sud. Si discute in generale sulle alluvioni e i rischi appena passati.

### 9 Dicembre 2013

Numerosi tweet in riferimento all'uscita di un CD di beneficenza per raccolta fondi per i bambini alluvionati di Olbia.

### 10 Dicembre 2013

Vari post legati alla partita Galatasaray-Juventus rimandata per grandine.

### 12 Dicembre 2013

Il crawler è stato spento e ha registrato solo pochi tweet. Per questo sono state rilevate come allerte anche discussioni con pochi post (circa 30 per Olbia per esempio).

Blacklist: Parte (Brescia)

### 13-31 Dicembre 2013

Crawler del Politecnico spento per problemi tecnici legati al server. Non è stato possibile raccogliere dati in queste giornate.

## 5.2 Analisi di Allerte e Segnalazioni

In questa sezione verranno presentate le nostre analisi relative alle singole emergenze identificate nella sezione 5.1. Ci concentreremo più che altro sull'aspetto riguardante il Tool di Entity Discovery, ovvero la scoperta di queste emergenze, e non l'analisi della loro durata o del parlato degli ultimi giorni.

Utilizzando una metodologia di esposizione analoga a quella adottata in precedenza, le sottosezioni seguenti avranno:

- **Grafico di Tweet e Segnalazione;** contiene l'andamento del numero di tweet crawlati su cui è eseguito il Tool di Classificazione, valutati con un passo di distacco temporale breve di 15 minuti. A questo è poi associato il relativo grafico di Allerta-Segnalazione che mostra il mo-

## Verifiche Empiriche e Miglioramenti

mento in cui il nostro Motore Semantico ha identificato la presenza di una possibile situazione di allarme e quando questa si è concretizzata in una vera e propria emergenza. Abbiamo infine annotato sui grafici l'orario in cui è stato comunicato, in maniera più o meno ufficiale, uno stato di emergenza da parte delle autorità competenti, nei casi in cui ci è stato possibile reperire tali informazioni online (principalmente dal sito Emergenza24.org già citato alla sezione 2.2.4.11).

- **Commenti al Grafico;** servono a chiarificare eventuali aspetti ambigui dei risultati ottenuti, nonché per presentare le nostre riflessioni in relazione ad ogni specifico caso.

### 5.2.1 Nubifragio Catania - 21 Febbraio

Grafico di Tweet e Segnalazione:

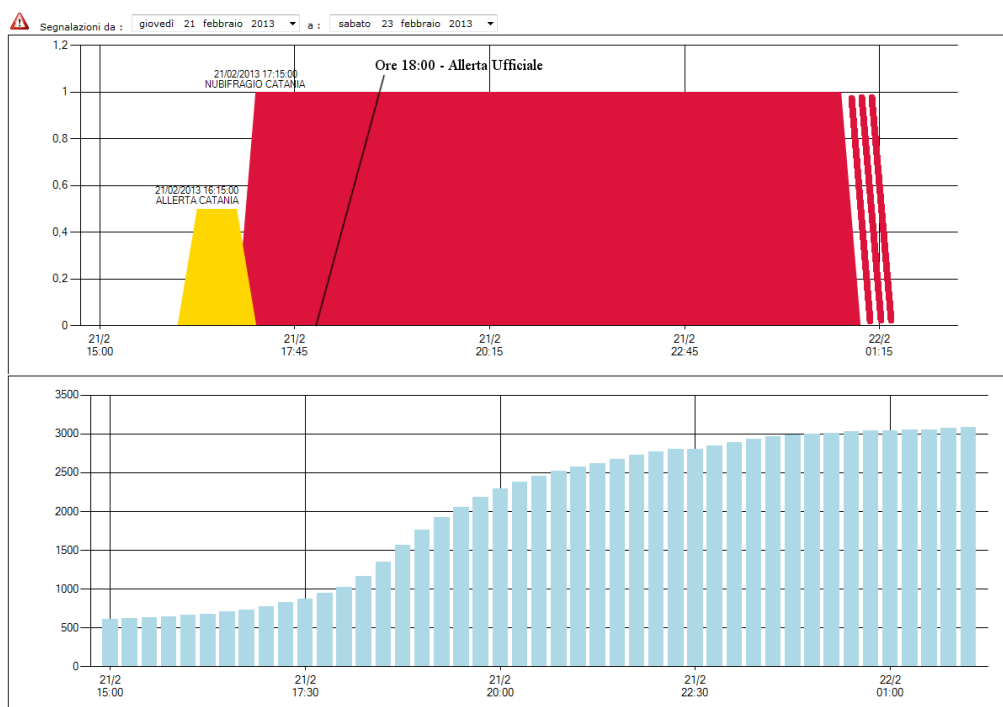


Figura 5.25: Grafico di analisi dell'Entity "Nubifragio Catania".

**Commenti al Grafico:**

La prima vera segnalazione riscontrata dall'analisi dei dati del 2013 è stata il violento nubifragio che ha colpito Catania tra il 21 ed il 22 di Febbraio.

## 5.2 Analisi di Allerte e Segnalazioni

Caratteristiche principali di un fenomeno atmosferico quale un nubifragio sono l'intensità, la breve durata e la velocità improvvisa con cui il fenomeno si manifesta. Come dimostrano i due grafici a confronto infatti, la quantità di parlato era esigua a poche ore dall'evento, ma l'incremento del numero di tweet è stato molto rapido, raggiungendo un valore quasi 6 volte più grande in sole 6 ore.

Anche la qualità del parlato è migliorata in poco tempo, poiché il Motore Semantico è riuscito ad identificare un'allerta già quando i tweet superavano di poco la soglia dei 500 e l'Entity è stata assegnata ancora prima di giungere ai 1000, anticipando quindi il vero e proprio picco di parlato.

In questo caso l'algoritmo delle allerte è riuscito ad anticipare di solo un ora quello delle segnalazioni, ma quest'ultimo ha comunque preceduto di circa un ora la comunicazione ufficiale dell'emergenza.

### 5.2.2 Alluvione Vicenza - 15 Maggio

Grafico di Tweet e Segnalazione:

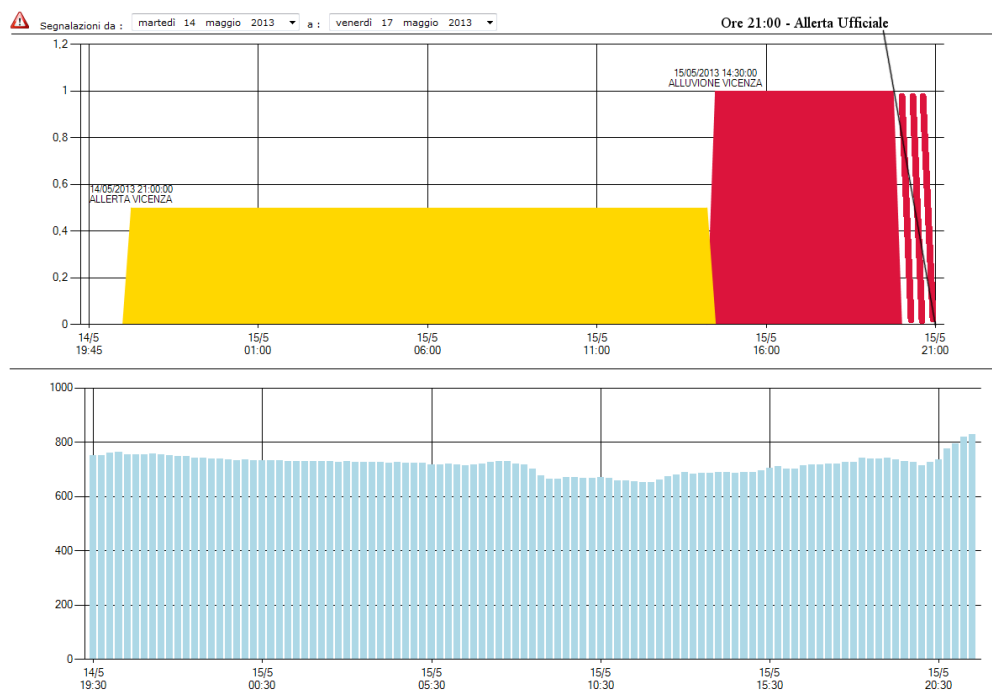


Figura 5.26: Grafico di analisi dell'Entity "Alluvione Vicenza".

Commenti al Grafico:

## Verifiche Empiriche e Miglioramenti

La seconda segnalazione da noi trovata riguarda delle forti precipitazioni che hanno interessato principalmente il Veneto tra il 14 ed il 15 Marzo e che hanno provocato a Vicenza un'alluvione che ha causato disagi e stato di allarme fino al 17 Marzo.

Il grafico dei tweet è, in questo caso, davvero molto significativo. Si può notare infatti come non esista un picco di parlato, il numero di messaggi rimane sempre intorno ai 700. A migliorare è stata quindi la qualità del parlato che si è concentrato tutto sulla situazione di pericolo che si stava generando nella città veneta, non più sul Veneto in generale. Questo ha permesso al Tool di Entity Discovery di rilevare un allarme con molto tempo di anticipo rispetto all'emergenza, ben 17 ore e 30 minuti prima. La segnalazione infine è a sua volta giunta con ulteriori 6 ore e 30 minuti circa di anticipo sulla comunicazione ufficiale, creando un gap totale di 24 ore di vantaggio.

### 5.2.3 Nubifragio Roma - 7 Luglio

Grafico di Tweet e Segnalazione:

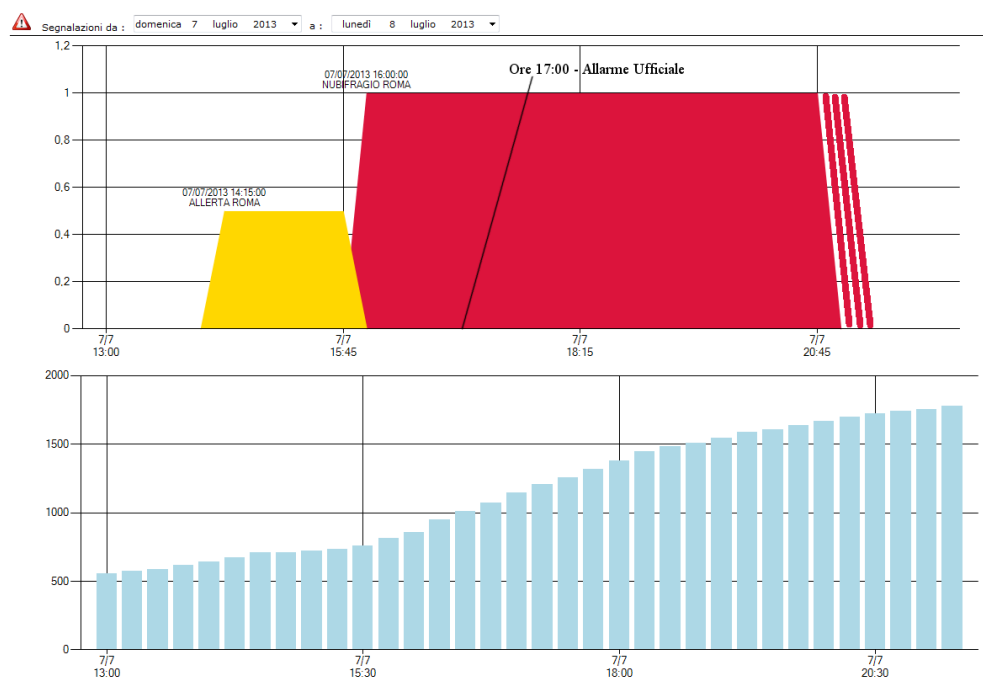


Figura 5.27: Grafico di analisi della prima Entity "Nubifragio Roma".

## 5.2 Analisi di Allerte e Segnalazioni

### Commenti al Grafico:

Come per il caso di Catania, anche il nubifragio di Roma del 7 Luglio è stato un evento caratterizzato da precipitazioni impetuose e poco preavviso.

L'andamento della curva del numero di tweet è pure simile, con pochi messaggi iniziali e un apice di parlato che si raggiunge nel giro di poche ore.

Nuovamente l'allerta viene segnalata con un margine di un ora di anticipo rispetto al rilevamento dell'emergenza e questa, a sua volta, è identificata con un ora di anticipo rispetto alle segnalazioni ufficiali disponibili in rete.

### 5.2.4 Nubifragio Roma - 21 Luglio

#### Grafico di Tweet e Segnalazione:

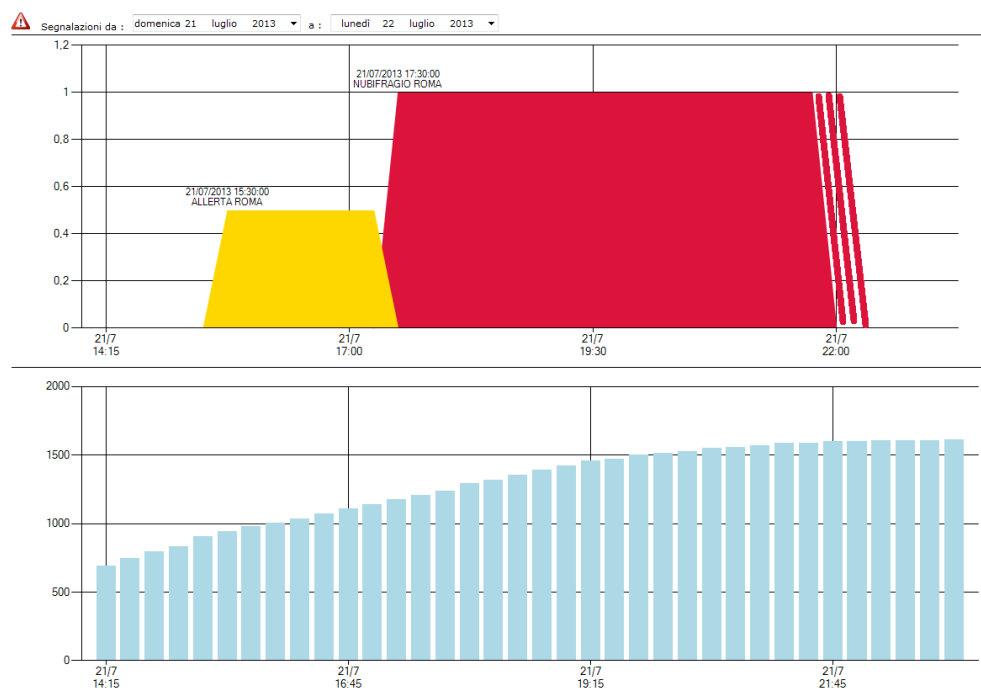


Figura 5.28: Grafico di analisi della seconda Entity "Nubifragio Roma".

### Commenti al Grafico:

Discorso del tutto analogo anche per il successivo nubifragio che ha colpito Roma ad esattamente due settimane di distanza dal primo.

## Verifiche Empiriche e Miglioramenti

Unica differenza notevole è che il parlato relativo all'allerta, e poi all'emergenza, si è intensificato oltre il normale già nelle prime fasi. Ciò è probabilmente dovuto al fatto che l'interesse della gente si è incentrato molto più velocemente sulla situazione, essendo il ricordo dei fatti di pochi giorni prima ancora fresco.

Grazie a questo, il sistema di alerting è riuscito ad identificare l'evento con due ore di anticipo rispetto alla segnalazione vera e propria. Purtroppo non siamo riusciti a trovare online riferimenti ad un orario ufficiale di comunicazione dell'emergenza.

### 5.2.5 Nubifragio Roma - 27 agosto

Grafico di Tweet e Segnalazione:

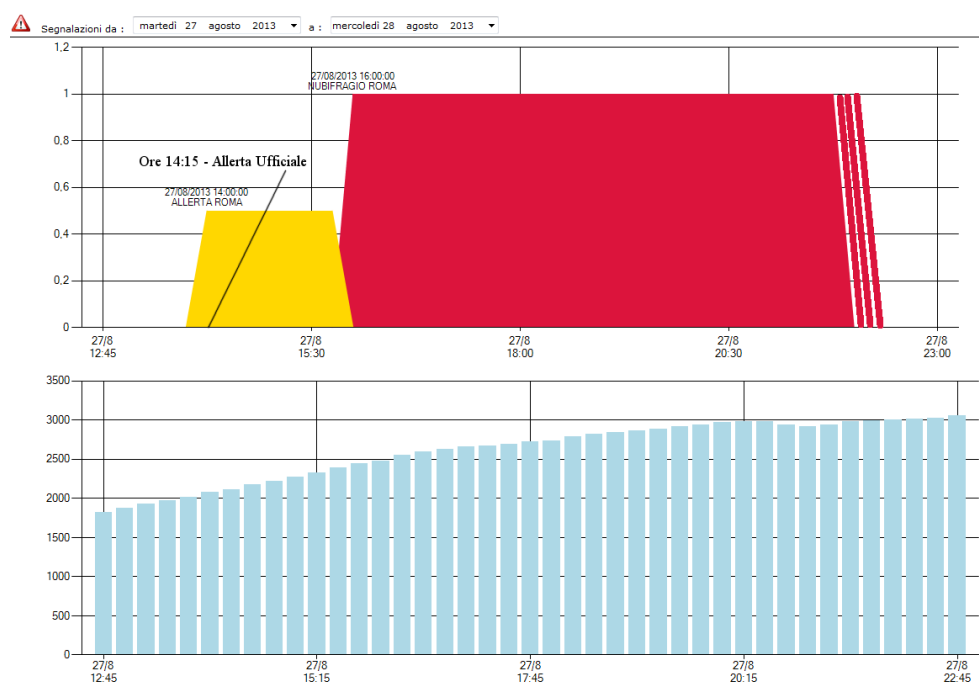


Figura 5.29: Grafico di analisi della terza Entity "Nubifragio Roma".

#### Commenti al Grafico:

Anche il discorso relativo al terzo nubifragio su Roma è analogo ai precedenti: si tratta di un evento improvviso, violento e di corta durata, ma con le potenzialità per provocare seri disastri e danni.



## 5.2 Analisi di Allerte e Segnalazioni

Questo nubifragio in particolare, segue gli ultimi due a breve distanza e, quindi, l'attenzione della gente e delle autorità era alta. Per di più il parlato riguardo al maltempo era già al centro dell'attenzione in quei giorni, in quanto si susseguivano temporali di grossa intensità in tutta la penisola.

La conseguenza è un numero di tweet a sfondo meteorologico già ampio, ma molto delocalizzato in partenza ed una focalizzazione dell'interesse delle persone e delle autorità pressoché immediata sul caso di Roma subito dopo l'inizio delle forti piogge nella capitale.

Il risultato di tutto ciò è un'allerta delle autorità che giunge praticamente contemporanea all'allerta rilevata dal nostro Tool di Entity Discovery, in anticipo di due ore sul rilevamento vero e proprio dell'emergenza.

### 5.2.6 Allagamenti Toscana - 21 Ottobre

Grafico di Tweet e Segnalazione:

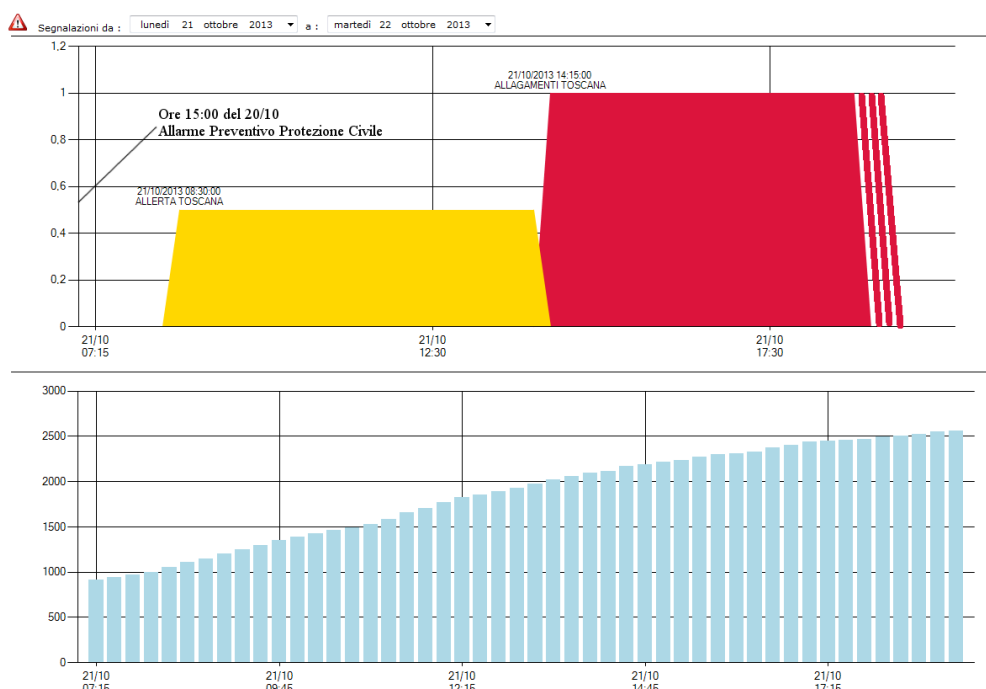


Figura 5.30: Grafico di analisi dell'Entity "Allagamenti Toscana".

**Commenti al Grafico:**

Dal grafico dell'alluvione in Toscana, avvenuta tra il 21 ed il 22 Ottobre, si

## Verifiche Empiriche e Miglioramenti

nota come il carattere di questa emergenza sia totalmente diversa da quella di un nubifragio.

La perturbazione che ha portato alla situazione di allerta era già nota e prevista alle autorità competenti che, per questo motivo, avevano già dal 19 lanciato un allarme preventivo. Inoltre si può notare come il numero di tweet aumenti in maniera praticamente lineare col passare del tempo, ma molto lentamente (ci mette più di 10 ore per raddoppiare il proprio valore da 1000 a 2000 messaggi).

Proprio per questo carattere di fenomeno lento, l'allerta riscontrata dal Tool di Entity Discovery giunge con 6 ore di anticipo rispetto alla segnalazione dell'Entità, ma, come già accennato, in questo caso l'allerta delle autorità era già stata lanciata da tempo.

### 5.2.7 Alluvione Sardegna - 18 Novembre

Grafico di Tweet e Segnalazione:

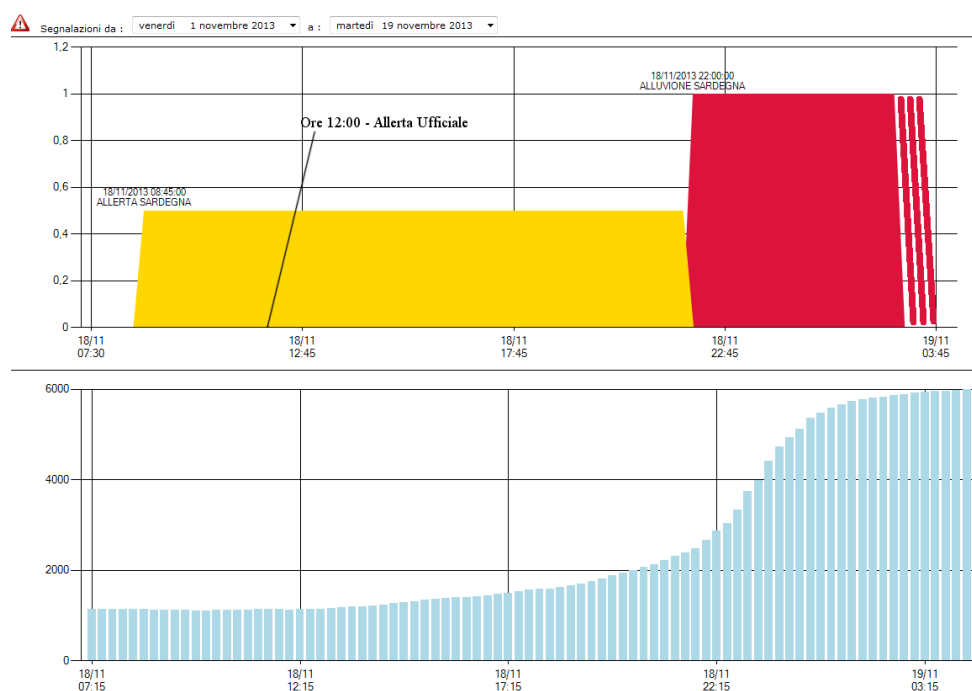


Figura 5.31: Grafico di analisi dell'Entity "Alluvione Sardegna".

Commenti al Grafico:

---

### 5.3 Riflessioni post-testing

L'ultima emergenza del 2013 da noi riscontrata è nuovamente a carattere tempestoso, come i casi di Vicenza e Toscana, ed è la più grande dell'anno, sia in termini di numero di tweet che di danni provocati. Si tratta dell'alluvione che ha colpito la Sardegna tra il 18 ed il 21 Novembre, ma argomento di discussione su Twitter fino alla fine del mese.

Ciò che spicca dal grafico è l'enorme quantità di tweet che si sono generati una volta che l'emergenza ha raggiunto la fase critica, superando addirittura la soglia dei 6000 messaggi nelle prime ore. Come le altre alluvioni, inizialmente il numero di messaggi cresce quasi linearmente e con un passo quasi lento (si aumenta di circa 1000 tweet in 5 ore), ma il vero scatto lo si ha la sera quando il maltempo investe completamente l'isola e le città principali, generando un'ondata di interesse crescente sull'argomento e un aumento impressionante di circa 4000 tweet in sole due ore.

Nuovamente si nota come il Tool di Entity Discovery sia riuscito a riscontrare la presenza di un'allerta senza bisogno di un picco iniziale, ma è palese come, in questo caso più che in altri, sia proprio l'arrivo di questo picco a far sì che quella si trasformi in un'effettiva segnalazione.

Un'ultima riflessione infine può essere fatta sugli orari. Come si vede dal grafico della segnalazione, il software di riconoscimento delle allerte rileva l'allarme con un anticipo di circa 3 ore sui comunicati ufficiali, ma questi precedono comunque di circa 10 ore la segnalazione effettiva dell'Entity. Una possibile spiegazione è imputabile nuovamente al carattere tempestoso di questa emergenza. Nel caso di un nubifragio, l'evento è qualcosa di violento ed imprevedibile, ma nel caso di una tempesta è possibile monitorare l'arrivo della perturbazione e, se lo ritengono opportuno, le autorità competenti possono lanciare degli allarmi preventivi con, eventualmente, largo anticipo. Così è successo nel caso della Toscana e, verosimilmente, anche nel caso della Sardegna.

### 5.3 Riflessioni post-testing

Con il nostro lavoro di testing siamo giunti ad una serie di conclusioni che elenchiamo di seguito.

## Verifiche Empiriche e Miglioramenti

---

In primis un appunto di carattere tecnico. Durante l'esecuzione delle prove abbiamo posto la soglia per il lancio del Tool di Entity Discovery a soli 100 tweet. Questo per essere sicuri che il programma venisse eseguito ogni giorno testato, anche in quelli con davvero poco parlato. A regime però, riteniamo che debba essere alzata a un valore compreso tra i 500 ed i 600 tweet, sia per evitare di eseguire inutilmente il Tool (un parlato sotto i 500 messaggi è davvero poco significativo) che per mantenere comunque una soglia capace di identificare anche l'inizio di un eventuale piccola allerta.

Visti i nostri risultati, riteniamo che il sistema di rilevazione di allerte da noi introdotto sia un ottimo strumento, utile per anticipare ulteriormente la segnalazione di emergenze da parte di TORCIA. Come abbiamo visto nelle varie emergenze analizzate in dettaglio, ci sono sempre diversi minuti di anticipo, se non ore, tra le nostre allerte e le segnalazioni, anche quelle ufficiali di allerta reperite online.

In particolare si può notare come l'allerta abbia un margine di anticipo ben ampio quando il tipo di emergenza che si sta verificando è a carattere alluvionare, o comunque una tipologia di emergenza che si verifica con un certo preavviso e in maniera continua nel tempo. Per quanto riguarda invece situazioni come nubifragi o forti grandinate estive, l'allerta è chiaramente vicina al momento della segnalazione, proprio per la natura improvvisa del fenomeno.

Il nostro consiglio per gli sviluppi futuri è di integrare questa funzionalità anche nel sistema TORCIA, in quanto potenzialmente utile e significativa per i fini stessi del progetto.

Il blacklistaggio delle località italiane con nomi ambigui è un compito oneroso e delicato, ma, come abbiamo visto, è necessario per evitare la generazione di troppe allerte fasulle. Nel testing effettuato, infatti, man mano che proseguivamo con l'analisi mensile inserivamo anche nel database i termini identificati come da blacklistare e, così facendo, abbiamo ridotto significativamente i falsi allarmi, arrivando praticamente a non generarne più negli ultimi mesi. Nonostante questo risultato, è necessario continuare in futuro il processo di blacklisting e integrare sempre la tabella con nuovi termini, qualora vengano riscontrati durante la normale esecuzione del software.

### 5.3 Riflessioni post-testing

---

Oltre al blacklistaggio dei luoghi inoltre, l'esecuzione del testing ci ha portato a decidere di escludere dalle Keyword la parola "Tempesta". Abbiamo infatti notato come moltissimi tweet siano stati prelevati dal Crawler proprio a causa della presenza di questo termine. Il problema è che, per la quasi totalità delle volte, si parla di tempeste che non hanno niente a che fare con un evento meteorologico. Il termine è per lo più utilizzato a sfondo politico, sentimentale, come parte di una citazione o, nella maggioranza dei casi, come termine comune del parlato. E pure nei casi in cui il termine era usato nella sua accezione utile ai nostri scopi, abbiamo rilevato come gli utenti che utilizzavano questo termine in riferimento all'emergenza fossero sempre un'esigua parte, non significativa per il rilevamento della segnalazione.



# Capitolo 6

## Conclusioni e sviluppi futuri

La metodologia e l'architettura descritte in questo lavoro di tesi hanno voluto offrire gli strumenti e un procedimento ben strutturato per sfruttare in modo completo il parlato sociale all'interno del contesto delle emergenze. La sua reale applicazione ha mostrato risultati in linea con le aspettative da noi definite nelle fasi iniziali del presente lavoro: le analisi preliminari hanno dimostrato un'alta percentuale di parlato durante particolari situazioni critiche, fornendo un significativo punto di partenza per la definizione dell'intera metodologia. Per poter acquisire informazione utile a partire da dati grezzi, si è voluto supportare il sistema di analisi sintattica con tool che potessero dare un valore semantico ad essi. In tal modo, il sistema mostra "intelligenza" nella fase di analisi dei dati, essendo in grado di poter estrapolare informazioni più precise con un alto contenuto informativo.

L'idea di sviluppare e integrare il nostro Motore Semantico all'interno del progetto TORCIA nasce proprio dal presupposto che tali informazioni dovessero essere estrapolate, manipolate e preparate per essere diffuse. Il nostro sistema quindi è un input fondamentale grazie al quale Beta80 e gli altri partner renderanno disponibili le informazioni da noi raccolte in modo immediato tramite Cruscotto e App. È soprattutto l'App il potenziale punto di forza che garantirebbe l'affermazione dell'intero progetto. L'informazione, infatti, acquisisce un valore fondamentale solo nel momento in cui è gestita in maniera real-time ed è risaputo come i dispositivi mobili siano attualmente i principali veicolatori di informazioni aggiornate; inoltre la loro ampia diffusione all'interno della popolazione porterebbe ad una rapida crescita

## Conclusioni e sviluppi futuri

---

nell'utilizzo dell'applicazione e ad una conseguente maggior sensibilizzazione dell'utente, che sarebbe sempre più incentivato ad usare la App per scopi con alta utilità sociale.

Il Motore Semantico sviluppato, e il testing su di esso effettuato, ci permettono di affermare come il suo utilizzo all'interno del contesto delle emergenze abbia un'importanza fondamentale, per avere un sistema realmente efficiente ed efficace nella gestione degli eventi critici. Inoltre, la sua flessibilità lo rende adattabile a qualunque tipo di ambito relativo alla gestione delle emergenze.

Abbiamo elencato nelle sezioni 4.7 e 5.3 dei possibili sviluppi futuri di carattere tecnico applicabili al nostro sistema. In generale TORCIA possiede, a nostro parere margini di miglioramento in termini di tempi di esecuzione, in quanto questo aspetto è stato da subito messo in secondo piano rispetto alla modularità del sistema. Inoltre, per come la sua architettura è disaccoppiata dai moduli implementati dagli altri partner, il nostro Motore Semantico è caratterizzato da una forte scalabilità per quanto riguarda l'interfacciamento. Riteniamo quindi che l'aggiunta di altri servizi web utili per la comunicazione tra Motore Semantico e il resto dei componenti del sistema sia qualcosa di facilmente integrabile, oltre che fortemente consigliato per ottenere uno scambio di feedback sempre più personalizzato e completo che possa favorire la qualità dell'output dell'intero progetto.



# Appendice A

## Categorie di classificazione

In questa appendice mostreremo i dettagli relativi al perfezionamento delle categorie, riportando nella sezione A.1 le categorie di partenza e successivamente, nella sezione A.2, il risultato delle modifiche apportate.

### A.1 Categorie di partenza

La composizione delle categorie di partenza è quella fornitaci dal Andrea Palla nel proprio lavoro di tesi [14].

**Dove:** Monitora, Monitorare, Città, Centro, Via, Viale, Piazza, Vie, Dintorni, Vicinanze, Vicino, Periferia, Periferica, Periferico, Zona, Quartiere, Regione, Limitrofo, Limitrofa, Adiacente, Adiacenze, Lontano, Distante, Posto, Località, Locale, Milano, Roma, Pescara, Genova, Lipari, Palermo, Borgogna, Liguria, Salento, Campania, Salerno, Metro, Nord, Sud, Ovest, Est, Venezia, Spagna, Lunigiana, Piazza, Tuscia, Messina, Bologna, Salernitano, Rieti, Arcore, Italia;

**Segnalazione:** Allagamenti, Allagamento, Allagata, Allagato, Allagati, Temporale, Temporali, Pioggia, Piogge, Diluvio, Alluvione, Esondazione, Esondato, Esondare, Fiume, Piena, Acqua, Fango, Tombini, Tombino, Nubifragio, Nubifragi, Tempesta, Uragani, Inondazione, Inondazioni, Grandine, Grandinata, Allagando, Uragano, Bufera, Grandinare, Inondato, Perturbazione;

## Categorie di classificazione

---

**Conseguenze:** Alluvionati, Morti, Morto, Feriti, Ferito, Vittima, Vittime, Sfollati, Sfollate, Sfollato, Evacuazione, Evacuata, Evacuato, Ucciso, Deceduto, Uccisa, Deceduta, Annegamenti, Annegamento, Annegato, Annegata, Salvata, Salvato, Salvo, Salva, Salvi, Recuperato, Recuperata, Recuperati, Danni, Crollato, Crollata, Crollati, Pericoloso, Pericolante, Pericolanti, Rischio, Rischioso, Tende, Tendopoli, Casa, Case, Frana, Franato, Franata, Crollata, Crollate, Crolli, Inghiottito, Inghiottite, Inghiottiti, Chiuse, Chiusa, Chiusi, Chiuso, Allagamento, Allagamenti, Situazione, Disagi, Calamità, Inondazioni, Ginocchio;

**Responsabilità:** Vigili, Vigile, Vigilessa, Polizia, Magistrato, Forze, Ordine, Fuoco, Pattuglia, Autopompa, Volante, Gazzella, Carabinieri, Carabiniere, Poliziotto, Poliziotta, Protezione, Civile, Capo, Comandante, Stato, Napolitano, Gabrielli, Bertolaso, Parlamento, Parlamentari, Senato, Ministro, Ministri, Piove, Gasparri, Alemanno, Gasparri, Sindaco, Comune, Ritardi, Tilt;

**Viabilità:** Percorribile, Percorribili, Libera, Libere, Liberata, Liberate, Pulita, Pulite, Ripulita, Ripulite, Sgombra, Sgombre, Chiusa, Chiusi, Chiuse, Chiuso, Via, Strada, Tangenziale, Tang, Traffico, Frana, Percorsi;

**Warning:** Intasato, Intasati, Intasata, Fogne, Fogna, Fognatura, Argini, Argini, Letto, Flusso, Attenzione, Pericolo, Pericoloso, Pericolose, Pericolosi, Inondazioni, Allagamento, Allagamenti, Avvertita, Temporali, Temporale, Diluviando, Piovendo, Esondazione, Spostamenti, Uragano, Allerta, Prevedono, Rischio, Previsti;

**Post:** Risarcimento, Risarcimenti, Ricostruzione, Lamenta, Lamentele, Lamentela, Assicurazione, Assicurazioni, Assicurati, Denaro, Soldi, Assegni, Assistenza, Alluvionati, Protesta, Protestano, Rimborso, Rimborsi, Interventi, Ripristino, Disagi, Fondi, Denunce, Sicurezza;

## A.2 Categorie migliorate

Le seguenti categorie sono il risultato definitivo dell'applicazione del processo di miglioramento descritto nella sezione 5.2. In realtà tale processo è stato applicato più di una volta fino a quando non sono stati raggiunti risultati soddisfacenti ma, per semplicità, mostreremo solo il risultato finale senza specificare tutti i passi intermedi.

**Dove:** Monitora, Monitorare, Città, Centro, Via, Viale, Piazza, Vie, Dintorni, Vicinanze, Vicino, Periferia, Periferica, Periferico, Zona, Quartiere, Regione, Limitrofo, Limitrofa, Adiacente, Adiacenze, Lontano, Distante, Posto, Località, Locale, Milano, Roma, Pescara, Genova, Lipari, Palermo, Borgogna, Liguria, Salento, Campania, Salerno, Metro, Nord, Sud, Ovest, Est, Venezia, Spagna, Lunigiana, Piazza, Tuscia, Messina, Bologna, Salernitano, Rieti, Arcore, Italia, Area, Valle, Galleria, Stazione, Ponte, Sottopasso, Sottovia, Arezzo, Valdarno, Toscana, Portici, Sanremo, Imperia, Spezia, capitale, Firenze, Napoli, Cinque Terre, Siracusa, Modena, Pescara, Inghilterra, Galles, Sarno, Piento, Badesse, Elba, Isola, Abruzzo, Bretagna, Sesto, Viareggio, Castelsardo, Sicilia, Terni, Pakistan, Canobbio, Sanremo, Friuli, Trieste, Nocera, Rossana, Mestre, Emilia, Bergamo, India, Feltrino, Umbria, Isole, Trapani, Eolie, Metaponino, NewYork, Alghero, Canarie, Azzorre, Cilento, Sommalombardo, Zurigo, Trieste;

**Segnalazione:** Allagamenti, Allagamento, Allagata, Allagato, Allagati, Temporale, Temporali, Pioggia, Piogge, Diluvio, Alluvione, Esondazione, Esondato, Esondare, Fiume, Piena, Acqua, Fango, Tombini, Tombino, Nubifragio, Nubifragi, Uragano, Uragani, Inondazione, Inondazioni, Grandine, Grandinata, Allagando, Uragano, Bufera, Grandinare, Inondato, Perturbazione, Alluvioni, Maltempo, Frane, Franato, Franta, Inondata, Grandinato, Grandinate, Inondare, Inondando, Allagare, Esondando, Franare, Franando, Alluvionato, Alluvionate, Alluvionata, Grandinando;

## Categorie di classificazione

---

**Conseguenze:** Alluvionati, Morti, Morto, Feriti, Ferito, Vittima, Vittime, Sfollati, Sfollate, Sfolcato, Evacuazione, Evacuata, Evacuato, Ucciso, Deceduto, Uccisa, Deceduta, Annegamenti, Annegamento, Annegato, Annegata, Salvata, Salvato, Salvo, Salva, Salvi, Recuperato, Recuperata, Recuperati, Danni, Crollato, Crollata, Crollati, Pericoloso, Pericolante, Pericolanti, Rischiosa, Rischioso, Tende, Tendopoli, Franato, Franata, Crollata, Crollate, Inghiottito, Inghiottite, Inghiottiti, Chiuse, Chiusa, Chiusi, Chiuso, Disagi, Calamità, Inondato, Inondata, Ginocchio, Danno, Caduto, Caduta, Caduti, Disastro, Disastri;

**Responsabilità:** Vigili, Vigile, Vigilessa, Polizia, Magistrato, Forze, Ordine, Fuoco, Pattuglia, Autopompa, Volante, Gazzella, Carabinieri, Carabiniere, Poliziotto, Poliziotta, Protezione, Civile, Capo, Comandante, Stato, Napolitano, Gabrielli, Bertolaso, Parlamento, Parlamentari, Senato, Ministro, Ministri, Gaspari, Alemanno, Gasparri, Sindaco, Comune, Ritardi, Accuse, Accusa, Squadre, Volontari, Volontariato, Interventi, Polverini, Briguglio, Fiorito, De Filippo, Moratti, Formigoni, Provincia, Comune Milano;

**Viabilità:** Percorribile, Percorribili, Libera, Libere, Liberata, Liberate, Pulita, Pulite, Ripulita, Ripulite, Sgombra, Sgombre, Chiusa, Chiusi, Chiuso, Strada, Strade, Tangenziale, Tang, Traffico, Percorsi, Statale, Ss, Autostrada, Rallentamenti, Rallentato, Rallentata, Galleria, Sottopasso, Ponte, Tilt, Circolazione, Regolare, Sottovia, Sospesa, Sospeso, A1,A3, A22, Metropolitana, Tram, Treno, Treni, Metropolitane, Ritardo, Ritardi, Piazza, Via, Vie, Duomo, Nullo, Quartiere, Viale ;

**Warning:** Intasato, Intasati, Intasata, Fogne, Fogna, Fognatura, Argini, Argini, Letto, Flusso, Attenzione, Pericolo, Pericoloso, Pericolose, Pericolosi, Avvertita, Diluviando, Piovendo, Spostamenti, Allerta, Prevedono, Rischio, Previsti, Probabili, Possibili, Possibile, Probabile, Rischia, Arrivo, Previsto, Prevista, Previste, Domani, Allarme, Attesa;

## A.2 Categorie migliorate

---

**Post:** Risarcimento, Risarcimenti, Ricostruzione, Lamenta, Lamentele, Lamentela, Assicurazione, Assicurazioni, Assicurati, Denaro, Soldi, Assegni, Assistenza, Alluvionati, Protesta, Protestano, Rimborso, Rimborsi, Interventi, Ripristino, Fondi, Denunce, Crisi, Disdette, Indagato, Indagati, Indagata, Interpellanza, Sicurezza, Piano, Lavori, Risolto, Risolti, Euro, Polemica, Polemiche;



# Elenco delle figure

2.1	Le fasi di gestione dell'emergenza. . . . .	6
3.1	Il flusso della attività svolte dal sistema TORCIA. . . . .	48
3.2	Architettura del sistema TORCIA. . . . .	54
4.1	Architettura del Motore Semantico. . . . .	64
4.2	Struttura del Database del Motore Semantico. . . . .	66
4.3	Architettura del servizio PUSH TWEET. . . . .	75
4.4	Struttura XSD per la trasmissione dei Tweet. . . . .	77
4.5	Esempio valido di un XML contenente due Tweet. . . . .	77
4.6	Confronto tra XML di un tweet e tabella tweets del Database. . . . .	79
4.7	Struttura XSD per la trasmissione dei Luoghi Notevoli. . . . .	82
4.8	Relazione tra le entità nell'algoritmo di classificazione. . . . .	83
4.9	Diagramma di Flusso per il rilevamento di un'emergenza. . . . .	94
4.10	Diagramma di Flusso per la gestione delle segnalazioni. . . . .	98
4.11	Struttura XSD per la trasmissione delle segnalazioni. . . . .	101
4.12	Esempio valido di un XML contenente una segnalazione. . . . .	101
4.13	Diagramma di Flusso per il rilevamento di un'allerta. . . . .	104
4.14	Schemi di alto livello per la Geolocalizzazione. . . . .	112
4.15	Esempio del file ms-configuration.xml. . . . .	116
4.16	Schema architetturale di Torcia Monitor. . . . .	118
4.17	Parametri di monitoraggio del Motore Semantico. . . . .	119
4.18	Monitoraggio del Web Service. . . . .	121
4.19	Monitoraggio del numero di Tweet. . . . .	122
4.20	Monitoraggio di Segnalazioni ed Allerte. . . . .	123
5.1	Numero di Tweet nelle giornate di Gennaio 2013. . . . .	129
5.2	Allerte e Segnalazioni di Gennaio 2013. . . . .	129

## ELENCO DELLE FIGURE

---

5.3	Numero di Tweet nelle giornate di Febbraio 2013. . . . .	131
5.4	Allerte e Segnalazioni di Febbraio 2013. . . . .	132
5.5	Numero di Tweet nelle giornate di Marzo 2013. . . . .	133
5.6	Allerte e Segnalazioni di Marzo 2013. . . . .	134
5.7	Numero di Tweet nelle giornate di Aprile 2013. . . . .	136
5.8	Allerte e Segnalazioni di Aprile 2013. . . . .	137
5.9	Numero di Tweet nelle giornate di Maggio 2013. . . . .	138
5.10	Allerte e Segnalazioni di Maggio 2013. . . . .	139
5.11	Numero di Tweet nelle giornate di Giugno 2013. . . . .	140
5.12	Allerte e Segnalazioni di Giugno 2013. . . . .	141
5.13	Numero di Tweet nelle giornate di Luglio 2013. . . . .	142
5.14	Allerte e Segnalazioni di Luglio 2013. . . . .	143
5.15	Numero di Tweet nelle giornate di Agosto 2013. . . . .	144
5.16	Allerte e Segnalazioni di Agosto 2013. . . . .	144
5.17	Numero di Tweet nelle giornate di Settembre 2013. . . . .	146
5.18	Allerte e Segnalazioni di Settembre 2013. . . . .	147
5.19	Numero di Tweet nelle giornate di Ottobre 2013. . . . .	148
5.20	Allerte e Segnalazioni di Ottobre 2013. . . . .	148
5.21	Numero di Tweet nelle giornate di Novembre 2013. . . . .	150
5.22	Allerte e Segnalazioni di Novembre 2013. . . . .	150
5.23	Numero di Tweet nelle giornate di Dicembre 2013. . . . .	152
5.24	Allerte e Segnalazioni di Dicembre 2013. . . . .	152
5.25	Grafico di analisi dell'Entity "Nubifragio Catania". . . . .	154
5.26	Grafico di analisi dell'Entity "Alluvione Vicenza". . . . .	155
5.27	Grafico di analisi della prima Entity "Nubifragio Roma". . . . .	156
5.28	Grafico di analisi della seconda Entity "Nubifragio Roma". . . . .	157
5.29	Grafico di analisi della terza Entity "Nubifragio Roma". . . . .	158
5.30	Grafico di analisi dell'Entity "Allagamenti Toscana". . . . .	159
5.31	Grafico di analisi dell'Entity "Alluvione Sardegna". . . . .	160



# Elenco delle tabelle

4.1	Composizione delle categorie pre modifica. . . . .	84
4.2	Tabella di contingenza. . . . .	85
4.3	Valori delle metriche prima delle modifiche. . . . .	87
4.4	Valori delle metriche dopo le modifiche. . . . .	88
4.5	Valori ottenuti per le metriche usando un diverso Test Set. . . . .	88
4.6	Soglie statiche del Tool di Entity Discovery. . . . .	108
4.7	Esempi di termini contenuti nella blacklist. . . . .	115



# Bibliografia

- [1] Il metodo Augustus. [http://www.ispro.it/wiki/images/9/95/Metodo\\_Augustus.pdf](http://www.ispro.it/wiki/images/9/95/Metodo_Augustus.pdf), 2008.
- [2] Social Media Week di Torino. 2012.
- [3] D. M. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13:11, 2007.
- [4] M. Cameron, R. Power, B. Robinson, and J. Yin. Emergency situation awareness from twitter for crisis management. In *SWDM Workshop*, pages 305–308, Lyon, France, 2012.
- [5] Claudio Carcaci. Classificazione sintattica con metodi insiemistici. Master’s thesis, Politecnico di Milano, 2012.
- [6] D. Cartwright, F. Harary, and R.Z. Norman. *Structural Models: An introduction to the Theory of Directed Graphs*. 1965.
- [7] Giulia Cesana e Alfonso D’Ettorre. Analisi, metodologia e sviluppo di un sistema integrato per la gestione delle emergenze in ambito social. Master’s thesis, Politecnico di Milano, 2013.
- [8] E. Estellés and F. González. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200, 2012.
- [9] Josh Luger. Here’s how mobile social networks have become. *Business Insider*, 2013.
- [10] Social Media Manager. I social network piú popolari in italia: Facebook, Twitter, LinkedIn. *Socia Media Manager*, Aprile 2013.

## BIBLIOGRAFIA

---

- [11] M.Cha, H.Haddadi, F.Benevenuto, and K.P.Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM 2010: Proceedings of international AAAI Conference on Weblogs and Social*, 2010.
- [12] S. Milgram. The small world problem. *Psychology today*, 2:60–67, 1967.
- [13] J. L. Moreno. *Who shall survive?: A new approach to the problem of human interrelations*. 1934.
- [14] A. Palla. Metodologia per la gestione di emergenze basata su informazione sociale. Master's thesis, Politecnico di Milano, 2012.
- [15] G. Shih. In hurricane, Twitter proves a lifeline despite pranksters. *Reuters*, 2012.
- [16] G. Simmel. *Soziologie: Untersuchungen über die Formen der Vergesellschaftung*. Ducker & Humblot, 1908.