

POLITECNICO DI MILANO
Facoltà di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria delle Telecomunicazioni
Dipartimento di Elettronica e Informazione



**METODOLOGIA PER LA GESTIONE DI
EMERGENZE BASATA SU INFORMAZIONE
SOCIALE**

Relatore: Prof. Paolo GIACOMAZZI

Correlatori: Prof.ssa Chiara FRANCALANCI
Ing. Alessandro POLI

Tesi di Laurea di:
Andrea PALLA
Matr. 682777

Anno Accademico 2011 / 2012

Sommario

Il presente lavoro di tesi intende definire una metodologia per l'estrazione di informazione rilevante dai messaggi scambiati sui *social network*, allo scopo di supportare le strategie decisionali e operative delle organizzazioni preposte alla gestione delle emergenze. Si vuole quindi ottenere un metodo efficiente con il quale, attraverso un monitoraggio e un'analisi costante delle conversazioni dei cittadini sulle reti sociali, si possa riuscire a catturare messaggi riguardanti situazioni di emergenza - come segnalazioni, allarmi, richieste di aiuto - che, una volta interpretati e organizzati, possano fornire dati utili alle autorità.

Numerosi casi recenti hanno infatti dimostrato come Internet e i *social network* si siano rivelati efficaci nel veicolare informazioni utili in contesti di crisi. Alcuni eventi verificatisi nel passato recente hanno ottenuto grande risalto su questi nuovi mezzi di comunicazione, e i messaggi a riguardo, scambiati dalle persone, hanno spesso aiutato le istituzioni nella risoluzione di vari problemi. Le informazioni ricavate dal Web 2.0, debitamente sfruttate dalle autorità, potrebbero portare a un miglioramento complessivo nel processo di risoluzione dell'evento critico, fornendo un supporto alle decisioni per tutti quegli enti impegnati a garantire la salvaguardia dei cittadini. Analogamente, esse potrebbero aiutare durante le fasi di monitoraggio del territorio, grazie a una raccolta dati distribuita basata sul paradigma *crowdsourcing*, ovvero sulla partecipazione attiva della cittadinanza e sullo sfruttamento coordinato delle conoscenze e delle capacità di un numero elevato di individui. L'utilizzo delle persone comuni alla stregua di "sensori" può infatti condurre a una diffusione più rapida delle informazioni e all'acquisizione di dettagli importanti riguardo alla situazione del territorio durante un'emergenza. Spesso questi dettagli sono però nascosti all'interno di messaggi non esplicitamente diretti alle autorità: serve perciò una strategia per ricavarli e strutturarli, in modo che risultino d'aiuto.

Poiché i messaggi scambiati sui *social network* possono assumere natura variegata, si intende qui affrontare il problema di come costruire un dominio lessicale focalizzato sul contesto delle emergenze. Tale dominio potrà poi essere sfruttato all'interno di un algoritmo sintattico, che avrà lo scopo di effettuare un'analisi automatica dei messaggi scambiati in rete, in modo da definire una categorizzazione degli stessi. Questo fornirà in uscita un contenuto informativo strutturato, dal quale poter estrarre dati utili per la rilevazione o la risoluzione di un'emergenza.

Lo studio effettuato nel presente lavoro si inserisce nell'ambito di un progetto denominato TORCIA, che vede la collaborazione del Dipartimento di Elettronica e Informazione del Politecnico di Milano e di Fondazione Politecnico con diverse imprese del settore informatico e delle telecomunicazioni. Lo scopo di tale progetto è la realizzazione di una piattaforma resiliente in grado di raccogliere, analizzare e interpretare i messaggi scambiati dai cittadini riguardo agli eventi critici che possono avvenire in ambito metropolitano, e di restituire in uscita una serie di dati aggregati fruibili dalle autorità per rendere più efficiente il processo di gestione delle emergenze. A tal proposito, la finalità primaria di questa tesi è concentrarsi sulla definizione di un'ontologia di dominio sintattico per le emergenze, ricavata a partire da un'analisi *bottom-up* sui testi scambiati nei *social network* durante una situazione di crisi. Il dominio così generato, costruito a partire da casi reali e focalizzato su una specifica emergenza, potrà poi essere esteso ad altri contesti grazie a una serie di considerazioni argomentate in chiusura. La definizione del dominio partirà da un'analisi sistematica dei messaggi scambiati su uno specifico *social network* – che nel caso del presente lavoro sarà individuato in Twitter –, così da generare un vocabolario di termini ricorrenti nell'ambito di un'emergenza predefinita, istruendo così l'algoritmo in modo che esso sia in grado di suddividere correttamente i messaggi, riconoscendone il contenuto informativo.

Quanto ricavato qui, poi, potrà essere utilizzato per l'estensione della metodologia ad altre sorgenti informative.

La tesi è strutturata come segue.

Nel primo capitolo dedicato allo “stato dell'arte”, si forniscono tutti gli strumenti per comprendere cosa sia un'emergenza e come essa venga tradizionalmente trattata. Si spiega inoltre cos'è il paradigma *crowdsourcing*, dimostrando come esso possa essere efficacemente sfruttato durante un'emergenza.

Il capitolo 2 aiuta nell'apprendere gli aspetti basilari riguardo alla classificazione delle emergenze, necessari per la definizione della metodologia.

Il capitolo 3 è dedicato all'analisi di alcuni casi d'emergenza reali in cui i *social network* e in generale gli strumenti messi a disposizione dal Web 2.0 hanno giocato un ruolo fondamentale nella gestione dell'emergenza. L'analisi preventiva, fatta sulla letteratura, è un passo necessario alla definizione del dominio.

Il capitolo 4 contestualizza quanto visto precedentemente all'interno del progetto TORCIA, fornendo anche un supporto introduttivo alla metodologia e definendo il contesto nel quale essa andrà a essere utilizzata.

Il capitolo 5 è il cuore della tesi, e definisce in maniera dettagliata la metodologia ricercata per l'estrazione di informazione sociale. La metodologia viene poi applicata a un algoritmo sintattico, al fine di ricavare risultati qualitativi in merito alla bontà dello studio effettuato.

Ringraziamenti

Desidero innanzitutto ringraziare il mio relatore Prof. Paolo Giacomazzi, e i correlatori Prof.ssa Chiara Francalanci e Ing. Alessandro Poli, per il supporto fornitomi durante la stesura della tesi, e per l'opportunità che mi hanno dato di sviluppare questo lavoro.

Un ringraziamento va a tutti i professionisti che lavorano al progetto TORCIA e con cui ho potuto collaborare in questi mesi. Molto del loro apporto e delle loro idee è confluito in questi capitoli.

Un grazie particolare alla mia famiglia e ai miei amici, quelli di sempre e quelli scovati lungo il cammino, che in questi anni mi hanno spronato affinché completassi il mio percorso di studi. E naturalmente ai miei compagni di corso, che ne hanno condiviso con me gioie e dolori.

Ma il mio grazie più grande non può che andare ai miei genitori, per l'aiuto, la comprensione, la pazienza e l'amore che mi hanno sempre dimostrato durante tutta la vita. Se questo elaborato esiste, lo devo ai loro sacrifici e alle possibilità che mi hanno sempre donato.

Questa tesi è dedicata a mamma e papà.

Elenco delle figure

Figura 1 - Il ciclo di vita delle emergenze	4
Figura 2 - Diffusione delle informazioni secondo la logica NORAD/NORTHCOM....	42
Figura 3 - L'architettura di TORCIA	59
Figura 4 - Segnalazione geolocalizzata	63
Figura 5 - Segnalazioni su una mappa.....	63
Figura 6 - Sistema UMB.....	64
Figura 7 - Relazione tra le entità usate dall'algoritmo	87

Elenco delle tabelle

Tabella 1 - Classificazione degli eventi sulla base dell'impatto	26
Tabella 2 - <i>Keyword</i> risultanti dal <i>crawling</i> nel mese di Ottobre 2012.....	77
Tabella 3 - Categorie per i messaggi	82
Tabella 4 - Tabella di contingenza	88
Tabella 5 – Tempi di esecuzione dell'algoritmo di Carcaci.....	94
Tabella 6 - Risultati qualitativi dell'algoritmo di Carcaci	95

Indice

Sommario	i
Ringraziamenti	iv
Elenco delle figure.....	v
Elenco delle tabelle.....	vi
Indice	vii
Introduzione.....	1
1 - Stato dell'arte	4
1.1 - Il ciclo di vita delle emergenze	4
1.2 - Il piano d'emergenza	6
1.3 - I social network	10
1.3.1 - Reti sociali <i>Web-based</i> e <i>social media</i>	11
1.3.2 - I principali <i>social network sites</i>	13
1.4 - Il <i>crowdsourcing</i>	18
1.4.1 Il <i>crowdsourcing</i> nelle fasi di un'emergenza	20
2 - Classificazione delle emergenze	22
2.1 - Il grado d'impatto di un evento	23
2.1.1 - Le classi d'impatto	24
2.1.2 - Definizione di disastro sulla base dell'impatto	26
2.2 - Classificazione degli eventi secondo la scala.....	27
2.3 - Classificazione degli eventi secondo il tipo	28

2.4 - Classificazione degli eventi secondo l'anticipabilità	29
2.5 - Elenco di eventi critici	30
2.5.1 - Gli eventi di origine naturale.....	30
2.5.2 - Gli eventi di origine umana/tecnologica	36
2.6 - Le emergenze in ambito cittadino	38
3 - Ruolo dei <i>social network</i> nella gestione delle emergenze.....	41
3.1 - I social come nuova strategia di comunicazione: il caso NORAD/NORTHCOM	42
3.2 - Casi emblematici durante emergenze reali.....	43
3.2.1 - Uragano Gustav in Louisiana (Agosto 2008).....	43
3.2.2 - Eruzione del vulcano islandese Eyjafjöll (Aprile 2010)	44
3.2.3 - Allagamento a Queensland, Australia (Gennaio 2011).....	45
3.2.4 - Giappone: terremoto, tsunami, e disastro nucleare (Marzo 2011)	46
3.2.5 - Terremoto dell'Emilia (Maggio 2012).....	47
3.3 - Esempi strategici di utilizzo di informazione sociale.....	48
3.3.1 - 911ICE.org («In Case of Emergency»).....	48
3.3.2 - American Red Cross Digital Operations Center	49
3.3.3 - Emergency Situation Awareness – Automated Web Text Mining (ESA-AWTM)	50
3.4 - Una piattaforma di crowdsourcing: Ushahidi	52
3.4.1 - Utilizzo di Ushahidi durante il terremoto di Haiti (Gennaio 2010)	53
4 - Il progetto TORCIA	56
4.1 - I <i>partner</i> del progetto	57
4.2 - Le attività del progetto	58
4.3 - L'architettura di TORCIA	59
4.3.1 - Input del sistema	60
4.3.2 - Il core del sistema.....	61

4.3.3 - Il motore di calcolo	61
4.3.4 - L' <i>output</i> del sistema: il supporto alle decisioni	62
4.4 - I destinatari dell' <i>output</i> TORCIA	65
5 - L'analisi dell'informazione sociale in un contesto d'emergenza specifico	66
5.1 - I passi della metodologia.....	68
5.2 - Scelta della sorgente.....	69
5.3 – Scelta di un contesto specifico di analisi.....	73
5.4 - Acquisizione di dati dalla sorgente	75
5.4.1 - Scelta delle <i>keyword</i> di ricerca: la definizione di un dominio lessicale.....	76
5.4.2 - Il crawling	77
5.4.3 - La prima analisi manuale dei dati	79
5.4 - Costruzione delle categorie	81
5.5 - Utilizzo dei dati all'interno di un algoritmo sintattico	86
5.5.1 - L'algoritmo di Carcaci	86
5.5.2 - Il sistema di test.....	90
5.5.3 - La preparazione dei dati	90
5.5.4 - Riduzione del <i>test set</i>	91
5.5.5 - Introduzione di un <i>feedback</i> nell'algoritmo	92
5.5.6 - Esecuzione dell'algoritmo con <i>feedback</i>	94
5.5.7 - Alcuni commenti relativi all'algoritmo.....	95
5.6 - I dati in uscita.....	96
5.7 - Estendibilità ad altri contesti di emergenza.....	97
5.8 - La geolocalizzazione.....	97
Conclusioni.....	99
Bibliografia.....	100

Introduzione

La diffusione di Internet e la nascita dei cosiddetti *social network*, ovvero comunità *online* in cui gli utenti possono scambiarsi messaggi e contenuti di varia natura, ha profondamente cambiato il modo in cui oggi vengono veicolate le informazioni. Quello che un tempo era dominio della stampa e dei cosiddetti *broadcaster*, che in qualche modo si occupavano di diramare le notizie secondo il proprio filtro giornalistico, è oggi invece un territorio variegato dove le informazioni divengono materiale di libera condivisione e patrimonio di tutti cittadini.

Vi sono situazioni in cui la cosiddetta “informazione sociale” assume un peso addirittura superiore a quello dell’informazione istituzionale, arrivando a raggiungere un pubblico ancora più vasto e potendo perciò risultare utile in specifici contesti, tra i quali quello delle emergenze. Diversi studi hanno infatti osservato come le persone siano naturalmente portate ad utilizzare i *social media* durante le situazioni di crisi, scambiandosi informazioni relative all’emergenza che stanno vivendo o utilizzando il Web per richiedere l’intervento delle autorità, soprattutto quando tutti gli altri mezzi si siano rivelati inefficaci. A volte le informazioni relative alle emergenze non sono nemmeno esplicitate nei messaggi scambiati attraverso la rete, ma il semplice riferimento a fatti rilevanti potrebbe costituire un dettaglio importante per meglio comprendere la situazione di crisi.

Trovare un modo per poter analizzare e interpretare questo parlato sociale significherebbe, per le autorità, avere a disposizione uno strumento utile per realizzare il cosiddetto *crowdsourcing*, ovvero sia per utilizzare un meccanismo distribuito di segnalazione e di risoluzione dei problemi, sfruttando le conoscenze di tutti quei cittadini che sono testimoni in primo piano delle emergenze.

Analogamente, l’utilizzo di Internet per realizzare campagne informative o per allertare i cittadini permetterebbe alle istituzioni di agire tempestivamente, ancor prima del veri-

ficarsi di un'emergenza, grazie alla rapida diffusione delle informazioni rilasciate su questi mezzi sociali.

La rete e il suo popolo, quindi, assumono un doppio ruolo: da un lato possono diventare contributori attivi, segnalando gli elementi critici presenti sul territorio e aiutando perciò le autorità nell'analisi dei rischi e/o dei danni connessi a un'emergenza, e dall'altro rimangono destinatari dell'informazione istituzionale, avendo però a disposizione un mezzo più flessibile per la ricezione di tale informazione.

Ricerche di sociologia e antropologia hanno fatto notare come la gente sia stimolata ad aiutare e supportare la comunità in caso di necessità. La diramazione delle proprie conoscenze è quindi un processo del tutto naturale che fa intrinsecamente parte dell'essere umano. Le emergenze, a causa della loro portata, coinvolgono emotivamente un grande numero di persone, persino coloro i quali non ne sono direttamente colpiti ma si sentono tuttavia solidali con le vittime. È facilmente osservabile come gli eventi disastrosi che accadono nella propria città, o nazione, se non addirittura nel mondo, tendano a diventare istantaneamente un argomento caldo di discussione. Oggi, grazie alle possibilità di comunicazione globale offerte dai *social media*, tale discussione viene mutuata attraverso il Web, e la diffusione delle informazioni è ormai pressoché istantanea. Saper catturare automaticamente queste conversazioni, al fine di sfruttarle per meglio gestire l'emergenza, potrebbe portare a innumerevoli vantaggi per le istituzioni. Allo stesso modo, potrebbe fornire un aiuto ai cittadini stessi, in qualche modo risollepati dalla possibilità di sfruttare un metodo facile e diffuso per offrire il proprio supporto o richiedere assistenza immediata.

Grazie all'analisi sistematica delle conversazioni scambiate sui *social network* è possibile costruire un dominio ontologico relativo alle emergenze. In altri termini, è possibile definire le parole e le frasi utilizzate dalle persone per discutere di un'emergenza. Per costruire tale dominio, naturalmente, occorre innanzitutto passare in rassegna i vari contesti reali, per comprendere di cosa parlino le persone durante una crisi, e di come lo facciano. La determinazione di questo dominio non è che il primo passo per la costruzione di sistemi automatizzati in grado di localizzare e segnalare un'emergenza, e che sappiano eseguire il calcolo di strategie risolutive che vadano a supporto delle decisioni delle autorità. Il sistema informatico non può e non deve divenire un'alternativa alle scelte umane di quei professionisti che si occupano degli eventi critici, ma può essere un valido aiuto alla pronta risposta e decisione degli stessi.

Piattaforme di questa natura potrebbero in futuro diventare parte integrante dei sistemi di monitoraggio di enti quali la Protezione Civile. Esistono già alcuni esempi nel mondo di organizzazioni che si sono dotati di strumenti di questo tipo, uno dei quali è la Croce Rossa Americana, che segue costantemente le discussioni su diversi *social media* al fine di determinare con rapidità la presenza di eventuali eventi critici.

All'interno di questo contesto, non esistono ancora in Italia sistemi di questo tipo. Il Politecnico di Milano, in collaborazione con diverse aziende del settore delle telecomunicazioni e dell'informatica, ha preso parte a un progetto finanziato da Regione Lombardia, finalizzato alla costruzione di una piattaforma resiliente in grado di monitorare e analizzare i messaggi provenienti da diverse sorgenti, tra le quali quelle del Web 2.0.

Questa piattaforma sarà in grado di realizzare le richieste prima elencate. Destinatari di tale piattaforma potrebbero essere tanto le autorità a vario livello, da quelle locali a quelle nazionali, quanto i cittadini stessi, grazie a uno strumento di aiuto per piccole emergenze di bassa portata. Un esempio per quest'ultimo punto è dato dalla città di Boston, che da qualche anno possiede un'infrastruttura informatica denominata *Citizen Connect*, grazie alla quale gli abitanti sono in grado di segnalare piccoli problemi in ambito urbano.

In questa tesi verranno forniti elementi utili per la definizione di una metodologia in grado di estrarre informazione dai messaggi scambiati sui *social network*, in particolare Twitter. Verranno presentati diversi casi reali per dimostrare che l'utilizzo della rete è veramente importante durante le emergenze.

Verrà poi presentato il progetto TORCIA, in modo da contestualizzare la metodologia in una piattaforma reale che tragga vantaggio dal suo utilizzo.

1 - Stato dell'arte

1.1 - Il ciclo di vita delle emergenze

Nella rappresentazione classica della gestione delle emergenze, ormai ampiamente riconosciuta, si è normalmente portati a considerare l'emergenza come un processo ciclico nel quale si articolano diverse fasi che, nell'insieme, costituiscono quello che viene definito "ciclo di vita delle emergenze". La suddivisione in fasi consente agli enti e ai soggetti che collaborano per la sua risoluzione di pianificare al meglio tutte quelle strategie e decisioni che permettano di minimizzare i danni e riportare quanto prima la situazione alla normalità.

Si possono dunque individuare quattro fasi principali che si susseguono l'una dopo l'altra e che, essendo l'emergenza un evento ripetibile, si chiudono in uno schema circolare. Il passaggio tra una fase e la successiva è scandito da alcuni eventi che lo determinano in modo più o meno netto.



Figura 1 - Il ciclo di vita delle emergenze

Preparedness

La *preparedness*, o fase *preventiva*, caratterizza i momenti precedenti all'evento critico che determina l'emergenza. Essa non richiede la presenza di forza in campo, ma risulta utile per la pianificazione delle strategie di intervento, oltre che per il monitoraggio costante del territorio al fine di identificare possibili scenari di pericolo. Gran parte delle conoscenze utilizzate nella *preparedness* discendono da eventi precedenti che hanno permesso l'acquisizione di un *know how* sempre più specifico sulle diverse emergenze.

Obiettivo della fase di *preparedness* è la creazione di un adeguato livello di prontezza nel rispondere alle emergenze che possono verificarsi nel territorio. Tale preparazione si deve ottenere ai diversi livelli organizzativi e decisionali (governo, regioni, comunità, individui) attraverso analisi dei rischi, circolazione delle informazioni e creazione di piani di risposta condivisi.

All'analisi rigorosa dei rischi sul territorio, eseguita dagli enti competenti, deve seguire anche una diffusione alla comunità della conoscenza acquisita, al fine di fornire i primi e più semplici strumenti di autodifesa al verificarsi dell'evento (punti di raccolta, comportamenti da seguire, comportamenti da evitare...).

Response

Il verificarsi dell'evento calamitoso determina l'inizio della fase di *Response*. Durante questa fase si mettono in atto tutte le procedure strategiche che consentano di contenere la situazione di crisi, limitare i danni, e agire prontamente dove necessario. All'interno di questa fase trovano spazio tutte le attività legate all'attivazione delle risorse disponibili per fronteggiare l'emergenza, alla ricerca di feriti e superstiti, all'eliminazione di situazioni critiche, alla creazione di rifugi e alloggi, alla creazione dei diversi centri operativi di coordinamento e controllo, al passaggio a regimi di legislazione e regolamentazione propri delle emergenze.

Durante la *response* è anche possibile valutare la bontà dei piani d'emergenza studiati precedentemente, fermo restando che in questa fase è spesso necessario rivedere rapidamente il protocollo qualora si verificano situazioni impreviste o inattese. I risultati in tal senso determineranno a seguire un'analisi delle responsabilità e un attento esame sulle capacità degli enti preposti alla risoluzione, oltre che sulle necessità future degli stessi (p.es. fondi, dotazione di nuovi strumenti e nuovi uomini, ecc.)

Recovery

La fase di *recovery* si pone come obiettivo il ritorno alla normalità nel più breve tempo possibile. Tale fase si attiva non appena si considera chiusa l'emergenza in senso stretto, ovvero nel momento in cui si è provveduto a mettere in sicurezza persone e patrimonio colpito dall'evento. Nella fase di *recovery* l'attenzione è posta tutta sul ripristino della normalità, da un punto di vista economico/produttivo, sociale e urbanistico/paesaggistico.

La "ricostruzione" va effettuata con particolare cura e attenzione, al fine di poter migliorare le strutture distrutte dal disastro e gettare le basi affinché l'evento di crisi non si verifichi nuovamente o, qualora questo fosse incontrollabile, ne vengano contenuti i danni in caso di nuova occorrenza.

Mitigation

L'ultima fase del processo è la fase di *mitigation*. Si tratta di una fase di lungo periodo durante la quale gli enti si attivano per ricercare le contromisure che evitino il ripetersi dell'evento critico o per ridurre le conseguenze, come naturale continuazione della fase di *recovery*. Tale operazione si può svolgere con diverse modalità a seconda del pericolo e del rischio presente nell'area. Si può cercare di agire sul pericolo stesso tentando di eliminarlo o ridurlo attraverso opere d'ingegneria sulle infrastrutture e sugli elementi ambientali a rischio (p.es. rinforzare gli argini di un fiume, realizzare costruzioni antisismiche), oppure si può cercare di modificare le modalità con cui l'evento interagisce con la popolazione (p.es. creando sistemi di allerta in tempo reale). O ancora, si può tentare, attraverso campagne di informazione e di comunicazione, di modificare gli stili di vita delle persone al fine di renderli edotti del rischio e di poter quindi prendere precauzioni di primo intervento (p.es. esercitazioni, creazione punti di raccolta, cartellonistica).

La fase di mitigazione sfocia poi in una nuova fase di preparazione, fornendo nuove conoscenze per il miglioramento delle strategie preventive riguardanti l'evento già fronteggiato.

1.2 - Il piano d'emergenza

Il piano d'emergenza rappresenta un documento o un insieme di documenti finalizzati alla salvaguardia dei cittadini e dei beni. Esso costituisce lo strumento unitario di rispo-

sta coordinata del sistema locale di Protezione Civile a qualsiasi tipo di situazione di crisi o di emergenza, avvalendosi delle conoscenze e delle risorse disponibili sul territorio.

Il piano di emergenza è il risultato di uno studio approfondito sulla vulnerabilità del territorio, sul rischio o possibilità che questo sia investito da un evento calamitoso, e sui punti critici che all'interno del territorio stesso possono risentire con diversa sensibilità del verificarsi dell'evento. Assume dunque la forma di un progetto che si occupa di definire i dettagli e le strategie riguardo a queste operazioni fondamentali:

- affidare responsabilità ad amministrazioni, strutture tecniche, organizzazioni e individui per l'attivazione di specifiche azioni, in tempi e spazi predeterminati, in caso di incombente pericolo o di emergenza che superi la capacità di risposta di una singola struttura operativa o ente, in via ordinaria;
- definire gli scenari di evento e di danneggiamento rispetto alle diverse tipologie di rischio presenti sul territorio;
- definire la catena di comando e le modalità del coordinamento inter-organizzativo, necessarie alla individuazione e alla attuazione degli interventi urgenti;
- individuare le risorse umane e materiali necessarie per fronteggiare e superare la situazione di emergenza.

La redazione di tale piano, affidata alla Protezione Civile, si basa sul noto **Metodo Augustus**, lo strumento di riferimento per la pianificazione nel campo delle emergenze, progettato dal geologo Elvezio Galanti e inquadrato dalla legge nazionale 225/92. Il nome del metodo si rifà a una frase dell'imperatore romano Augusto, che più di 2000 anni fa già sosteneva: «*il valore della pianificazione diminuisce con la complessità dello stato delle cose*» [Bignam]. Il metodo mette infatti in evidenza come il concetto di complessità sia intrinsecamente contrario alla pianificazione preventiva, che richiede invece strumenti e metodologie semplici e veloci.

Il metodo si basa sull'esecuzione di diverse fasi, prima su tutte la definizione dello *scenario* oggetto di pianificazione, seguita da un'analisi dei rischi peculiari dell'area considerata, per poi concludersi con l'attribuzione delle responsabilità ai diversi soggetti coinvolti nella risoluzione. Quest'ultima fase avviene attraverso l'individuazione di 14 *funzioni di supporto* che corrispondono a tutte le figure istituzionali competenti e specifiche per ogni settore. Tali figure sono direttamente coinvolte durante l'emergenza stessa.

sa, ma soprattutto nelle fasi precedenti ad essa, ovvero quelle di pianificazione e prevenzione. Le 14 funzioni sono tali in tutto il territorio nazionale e a tutti i livelli (nazionale, regionale, provinciale), tranne nel caso dei Comuni dove avviene una pianificazione che individua solo 9 funzioni di supporto. Nel dettaglio, Augustus intende essere un metodo flessibile e non legato al dominio di un particolare evento o rischio.

Quando necessario, le funzioni vengono attivate e chiamate a prendere posto presso i Centri Operativi. Questi ultimi possono essere di vario livello, a seconda del tipo di estensione geografica dell'emergenza:

- C.O.C.: Centro operativo comunale, presieduto dal sindaco o da altro soggetto da esso delegato;
- C.O.M: Centro Operativo Misto, strutture a livello provinciale o sovracomunale con lo scopo di coordinare al meglio centri comunali e sindaci afferenti al C.O.M. stesso;
- C.C.S: Centro di Coordinamento Soccorsi, struttura provinciale presieduta dal Prefetto o suo delegato che funge da ente per il coordinamento delle forze in campo per la gestione dell'emergenza. Il C.C.S. è inoltre deputato alla gestione delle richieste di soccorso raccolte attraverso la Sala Operativa Provinciale;
- C.O.R. Comitato Operativo Regionale presieduto dal presidente della regione;
- Direzione di comando e Controllo Nazionale con sede presso il Dipartimento della Protezione civile.

Le 14 funzioni che tali enti devono predisporre, individuate nel lavoro di Galanti, sono:

- **F 1 - Tecnica e di pianificazione:** Gruppi di ricerca scientifica (CNR) - Istituto Nazionale di Geofisica - Regioni - Dipartimento PC - Servizi Tecnici Nazionali
- **F 2 - Sanità, assistenza sociale e veterinaria:** Responsabile C.O. 118 - Regione / AA.SS.LL - C.R.I. - Volontariato socio-sanitario
- **F 3 - Mass-media e informazione:** RAI - Emittenti TV / radio private: nazionali e locali - Stampa
- **F 4 – Volontariato:** Dipartimento PC - Associazioni locali, provinciali, regionali, nazionali
- **F 5 - Materiali e mezzi:** C.A.P.I. - Ministero dell'Interno - Sist. Mercurio . FF.AA - C.R.I. - Aziende pubbliche e private - Volontariato
- **F 6 - Trasporti, circolazione e viabilità:** FF.SS. - Trasporto gommato, marittimo, aereo - ANAS - Soc. Autostrade - Province - Comuni - ACI

- **F 7 – Telecomunicazioni:** Ente Poste - Ministero delle Telecomunicazioni - Acconto di Telecomunicazioni
- **F 8 - Servizi essenziali:** ENEL - ENI - Gas - Acquedotto - Aziende Municipalizzate - Sistema bancario - Distribuzione carburante - Attività scolastica
- **F 9 - Censimento danni a persone e cose:** Attività produttive (ind., art., comm.) - Opere pubbliche - Beni culturali - Infrastrutture - Privati
- **F 10 - Strutture operative:** Dipartimento PC - VV.F. - Forze Armate - C.R.I. - C.C. - G.D.F. - Forestale - Capitanerie di Porto - P.S. - Volontariato - CNSA (CAI)
- **F 11 - Enti locali:** Regioni - Province - Comuni - Comunità montane
- **F 12 - Materiali pericolosi:** VV.F. - C.N.R. - Depositi e industrie a rischio
- **F 13 - Assistenza alla popolazione:** Forze Armate - Ministero dell'Interno - C.R.I. - Volontariato - Regioni - Province - Comuni
- **F 14 - Coordinamento centri operativi:** Collegamento con i centri operativi misti - Gestione delle risorse - Informatica

A capo di ogni funzione di supporto vi è un responsabile a cui spetta il compito, anche durante i periodi in cui non v'è emergenza, di mantenere, aggiornare e tenere vive le dinamiche di attivazione e gestione della singola funzione. I responsabili sono solitamente funzionari di medio-alto livello della struttura deputata all'esercizio della funzione, ed è da questa delegato a rappresentarla e a gestirne le risorse attivate in emergenza. Il punto di forza di questo metodo risiede nella flessibilità di applicazione, che permette ad esempio di non attivare una o più funzioni nel caso in cui queste non siano necessarie per lo svolgimento delle operazioni previste.

Il metodo Augustus richiede anche un aggiornamento periodico del piano redatto, in quanto lo scenario di rischio risente fortemente della dinamicità dell'assetto del territorio. Inoltre, impone la continuativa esecuzione di un'attività di informazione ai cittadini, in modo da portarli a *conoscenza* del territorio, ampliare la loro *coscienza* nei confronti dei rischi, ed educarli all'attuazione di strategie di *autodifesa*. Spesso questa sensibilizzazione passa anche attraverso *l'attuazione di esercitazioni*, specialmente in zone critiche.

Un ultimo punto fondamentale del metodo è relativo all'analisi delle risorse, ovvero di tutti quei beni e/o servizi utili o necessari alla strategia di risoluzione dell'emergenza. È evidente come una corretta gestione delle risorse sia un aspetto fondamentale nella crea-

zione di un piano di emergenza credibile e aderente alla realtà. Particolarmente delicato è il concetto di disponibilità delle risorse, ovvero la possibilità di attivarle in un dato momento per un dato compito. Come accennato più volte, molto spesso le risorse non mancano dal punto di vista “fisico”, ma mancano le procedure per attivarle o, ancor più grave, manca la conoscenza stessa della risorsa e della sua eventuale disponibilità. Accanto a un’attenta conta delle risorse disponibili, è dunque necessario effettuare una preallocazione delle stesse, per garantirne la possibilità di utilizzo durante l’emergenza.

1.3 - I social network

Le *reti sociali* (in inglese *social network*) sono strutture di individui connessi tra loro attraverso legami sociali, p.es. di tipo lavorativo, di amicizia, di interesse, oppure legami di tipo familiare. [Piselli, 2001]

Lo studio delle reti sociali è condotto da diverse discipline. La ricerca nell’ambito dei diversi approcci disciplinari ha evidenziato come le reti sociali operino a più livelli (dalle famiglie alle comunità nazionali) e svolgano un ruolo cruciale nel determinare le modalità di risoluzione di problemi e i sistemi di gestione delle organizzazioni, nonché le possibilità dei singoli individui di raggiungere i propri obiettivi.

In termini matematici, le reti sociali possono essere descritte da modelli che fanno uso della teoria dei grafi. In tali modelli, i soggetti (che possono essere individui, gruppi, comunità, ecc.) rappresentano i nodi, disegnati solitamente con pallini, mentre le relazioni che intercorrono tra essi sono rappresentate dagli archi che li collegano.

Tali relazioni possono essere:

- *univoche*, come accade p.es. in Twitter, dove è consentito seguire un utente senza necessariamente essere seguiti a propria volta da esso;
- *biunivoche*, come accade p.es. in Facebook, dove l’amicizia si instaura in entrambi i sensi.

La rete sociale è ben caratterizzata da una sua propria *densità*. La densità di una rete è un concetto basilare della teoria dei grafi, ed esprime il livello generale di coesione dei legami tra i punti di un dato grafo; vi sarà maggiore densità quanto più numerose sono le linee direttamente collegate. Definiti con N il numero di nodi del grafo e con L il numero di collegamenti, la densità si può calcolare mediante le formule:

$$\frac{2L}{N(N-1)} \quad \text{per un grafo non orientato}$$

$$\frac{L}{N(N-1)} \quad \text{per un grafo orientato}$$

La densità è dunque un parametro che può variare tra 0 e 1. Se tra tutti i nodi di rete esistessero collegamenti, la rete avrebbe densità pari a 1. Se viceversa non esistesse alcun legame tra i nodi di rete, essa avrebbe densità pari a 0. La densità, in un grafo che rappresenta una rete sociale, può essere un parametro che descrive l'efficienza nello scambio di informazioni e l'utilità per i singoli individui. Nelle opere di Mark Granovetter è evidenziato come reti piccole e dense possano tuttavia essere meno utili di reti costituite da legami deboli, poiché queste ultime potrebbero essere più flessibili, prestandosi quindi a un miglior scambio di idee e opportunità.

Il *numero di Dunbar*, conosciuto anche come “*regola dei 150*”, afferma che le dimensioni di una rete sociale in grado di sostenere relazioni stabili sono limitate a circa 150 membri. Questo numero è stato valutato attraverso diversi studi di sociologia e antropologia, e trova riscontro nell'ipotesi che esso potrebbe costituire una sorta di limite per l'abilità media degli esseri umani di riconoscere i membri di un gruppo, e di tenere traccia degli avvenimenti emotivi di tutte le persone che vi appartengono. Altre teorie si concentrano invece su motivazioni di stampo economico, evidenziando il fatto che gruppi troppo numerosi sarebbero più esposti al rischio di ingresso di elementi parassitari dannosi per l'organizzazione.

1.3.1 - Reti sociali *Web-based* e *social media*

La diffusione del Web e del termine *social network* ha creato negli ultimi anni alcune ambiguità di significato. La rete sociale è infatti storicamente, in primo luogo, una rete fisica. La versione di Internet delle reti sociali è una delle forme più evolute di comunicazione in rete, ed è anche un tentativo di violare la "regola dei 150".

Le piattaforme Web sociali offrono l'opportunità di organizzare in una mappa informatica le relazioni umane. Secondo la definizione data dagli studiosi Boyd e Ellison, si possono definire siti di reti sociali (*social network sites*) i servizi Web che permettono:

- la creazione di un profilo pubblico o semi-pubblico all'interno di un sistema vincolato;

- l'articolazione di una lista di contatti;
- la possibilità di scorrere la lista di amici dei propri contatti.

Le piattaforme sociali basate su Web permettono di aggregare amici, parenti, conoscenti, colleghi in un'unica piazza virtuale. Questi centri di aggregazione *online* hanno avuto un forte impatto, tanto da modificare le abitudini dei loro utilizzatori: l'utilizzo dei *social network* è infatti cresciuto del 93% dal 2006 al 2011.¹

I *social network sites* più diffusi a livello mondiale, e che al momento contano anche il maggior numero di utenti, sono Facebook, Twitter, LinkedIn e Myspace. La struttura di questi siti ha subito negli anni diverse modifiche, talvolta spinte dagli utenti stessi che hanno portato a semplificazioni complessive dei sistemi. La semplicità sembra infatti aver decretato il maggiore o minore successo di queste piattaforme.

Alle persone che si iscrivono a questi servizi è essenzialmente permesso, in diverse forme, di effettuare le seguenti operazioni:

- *creare un profilo utente*: ogni utente è generalmente profilato con una serie di dati, alcuni definiti dall'utente stesso a propria discrezione, altri organizzati automaticamente in base all'utilizzo della piattaforma. A seconda delle norme vigenti il profilo può essere reso più o meno fruibile alla rete in modo pubblico, semi-pubblico o privato;
- *creare gruppi*: la creazione di specifici gruppi di utenti all'interno del *network* consente una condivisione di contenuti e conoscenze più mirati;
- *ricercare*: cercare altri utenti dalle liste degli amici oppure dalla lista pubblica per ampliare le proprie conoscenze.

L'entrata in commercio degli *smartphone* ha modificato ulteriormente il panorama descritto. Parte del traffico sui *social network*, infatti, è oggi generato attraverso dispositivi mobili, e il dato è destinato a crescere: uno studio condotto da Courtland Brooks nel 2011² ha evidenziato, negli ultimi 5 anni, una crescita del 126% nell'utilizzo di applicazioni mobili di *social networking*.

Più in generale, i *social network sites* rientrano nel gruppo più ampio dei cosiddetti *social media* (in italiano *media sociali*), un termine che indica tecnologie e pratiche *online* adottate dalle persone per condividere contenuti testuali, immagini, video e audio.

¹ Fonte: GrabStats

² In "Dating and Matchmaking: benchmark report", 2011

I professori Andreas Kaplan e Michael Haenlein hanno definito i media sociali come «*un gruppo di applicazioni Internet basate sui presupposti ideologici e tecnologici del Web 2.0 che consentono la creazione e lo scambio di contenuti generati dagli utenti*».

[Kaplan et al., 2010]

I *social media* rappresentano fondamentalmente un cambiamento nel modo in cui la gente apprende, legge e condivide informazioni e contenuti. In essi si verifica una fusione tra sociologia e tecnologia che trasforma il monologo in dialogo, e ha luogo una democratizzazione dell'informazione che trasforma le persone da fruitori di contenuti ad editori degli stessi. Inoltre, i *social media* hanno anche un'influenza nelle modalità con cui le persone intessono oggi nuove relazioni, essendo divenuti un luogo di conoscenza, di approccio, di interazione, al pari e talvolta più di altri luoghi tipicamente fisici.

Alcune delle caratteristiche peculiari dei *social media* vengono evidenziate dal confronto con i media tradizionali. Per quanto entrambi offrano la possibilità di raggiungere un'*audience* molto vasta, i social media si distinguono in termini di *accessibilità*, *fruibilità*, *velocità*, e *permanenza* dell'informazione. Essi consentono infatti di generare un'informazione libera, diretta potenzialmente a tutti o a un pubblico selezionato dall'autore stesso, in maniera rapida e successivamente modificabile.

1.3.2 - I principali social network sites

In questo paragrafo si intende fornire una descrizione generale dei principali siti Web che offrono funzionalità di *social networking*, evidenziando le peculiarità e i difetti di ognuno.

Facebook

Lanciato nel 2004 e inizialmente destinato ai soli studenti dell'Università di Harvard, la piattaforma ideata da Mark Zuckerberg ha ampliato rapidamente la propria popolarità, tanto da essere oggi il *social network* più utilizzato al mondo. Il sito consente agli utenti di costruire un proprio profilo personale e di avere a disposizione una pagina su cui poter inserire contenuti meta-testuali di diversa natura, da semplici aggiornamenti di stato, a video, fotografie, note, ecc.

Tra gli utenti si possono instaurare relazioni bidirezionali che dovrebbero riflettere quelle della vita reale. Il massimo di relazioni per singolo utente è però fissato a 5000, un numero certamente più elevato rispetto alla rete di conoscenza fisica del soggetto in

questione: ciò fa pensare che il sito rifletta l'esigenza di alcuni di possedere un giro di conoscenze virtuali superiore a quello della vita reale.

I messaggi veicolati attraverso la rete possono essere di vario tipo. I messaggi personali sono scambiati tra utente e utente attraverso un metodo di messaggistica non dissimile a quello delle chat, eventualmente di gruppo, con la differenza che i messaggi non recapitati in tempo reale sono salvati all'interno di conversazioni richiamabili in qualunque momento. Il contenuto dei messaggi scambiati in questo modo resta privato ed è visibile ai soli partecipanti della conversazione. Attualmente i messaggi possono essere recapitati su altre piattaforme, come p.es. gli *smartphone* o le email, così da fornire un'integrazione del sistema con qualsiasi strumento informatico mobile.

Altri tipi di messaggio sono quelli che vengono postati sul "diario" di un utente, così chiamato in virtù del suo carattere biografico. Questi messaggi sono tipicamente visibili a una rete più ampia, definita dalle impostazioni di *privacy* proprie del singolo *post* in questione; è dunque possibile postare messaggi pubblici, aperti cioè all'intera rete degli iscritti a Facebook, oppure messaggi ristretti a una cerchia ridotta, che può essere quella dei propri contatti o di una parte di essi.

Quando un utente inserisce un contenuto sul proprio diario o su un diario altrui, a seconda delle impostazioni di *privacy* prescelte, l'operazione sarà visibile anche alla rete dei suoi contatti, grazie alla replicazione di tale operazione sulla bacheca degli stessi. La bacheca è il luogo dove tener traccia degli aggiornamenti dei propri amici, in modo da visualizzare i loro post, le loro iscrizioni, le loro attività, ecc.

Recentemente Facebook ha introdotto anche funzionalità per relazioni monodirezionali. Attraverso i cosiddetti "aggiornamenti", un utente è in grado di seguire i *post* di un altro utente senza diventarne amico, a patto che questi siano stati specificatamente impostati come visibili. Si tratta di un meccanismo che consente di tener traccia degli aggiornamenti di quelle persone che l'abbiano consentito esplicitamente, e con le quali, per un motivo o per l'altro, non è possibile o necessario stringere amicizia virtuale.

In Facebook, le conversazioni riguardo a specifici *topic* di interesse si realizzano essenzialmente attraverso le pagine e i gruppi. Le pagine sono profili dedicati ad aziende, marchi, prodotti, comunità o semplici concetti e idee. Esse sono non dissimili dai profili degli utenti, con la differenza che possono essere gestite da più amministratori e che gli aggiornamenti sono resi visibili in maniera pubblica, o comunque a chi ne faccia richiesta indicando il "mi piace" alla pagina. I gruppi richiamano invece, in larga misura, i

vecchi gruppi di discussione *usenet*. Nei gruppi, tutti gli utenti iscritti hanno facoltà di inserire materiale attinente a un *topic* specifico. Sono dunque bacheche virtuali per scambiarsi opinioni riguardo argomenti predefiniti, decisi dall'amministratore del gruppo. I gruppi possono essere *pubblici*, *privati*, o *nascosti*: nel primo caso chiunque può vederne i contenuti e iscriversi, nel secondo caso l'iscrizione è vincolata alla richiesta a un amministratore, infine nel terzo caso i gruppi sono nascosti alla ricerca e alla visibilità pubblica e l'iscrizione può avvenire solo su invito da parte di uno degli amministratori.

Facebook consente la ricerca di profili, pagine, gruppi, a patto che questa non sia stata inibita dal proprietario degli stessi. Una volta creato un profilo, questo non può essere cancellato ma solo sospeso; in caso di riattivazione, tutti i contenuti sospesi torneranno a essere visibili.

L'inserimento di contenuti è in generale stimolato da un meccanismo basato su commenti e *like* (“*mi piace*”). I commenti consentono ai contatti di far sapere la propria idea in merito al contenuto postato. I *like* esprimono un *feedback* positivo sul contenuto postato, indicando che esso è stato visto e “approvato” dal contatto che ha cliccato sul “mi piace”.

Le critiche che vengono mosse più di frequente a Facebook riguardano la generale struttura del sistema, che negli anni si è andata complicando sia da un punto di vista funzionale che da un punto di vista grafico, e varie lamentele riguardo alla *privacy*, che per molti non è adeguatamente garantita. Con un po' di disattenzione, infatti, è facile inserire contenuti dimenticandosi di impostarne le corrette restrizioni sulla visibilità: questo può spesso generare pasticci riguardo ai quali il *social network*, ovviamente, declina ogni responsabilità.

Twitter

Twitter è un servizio gratuito di *social networking* e *microblogging* che fornisce agli utenti una pagina personale aggiornabile tramite messaggi di testo con una lunghezza massima di 140 caratteri. Twitter è costruito totalmente su architettura Open Source³. Gli aggiornamenti possono essere effettuati tramite il sito stesso, o anche via SMS, con programmi di messaggistica istantanea, posta elettronica, oppure tramite varie applicazioni basate sulle API.

³ <https://dev.twitter.com/opensource>

Il servizio è diventato estremamente popolare, anche come avversario di Facebook, grazie alla sua semplicità e immediatezza di utilizzo: il 22 febbraio 2012 ha raggiunto i 500 milioni di utenti attivi che fanno accesso almeno una volta al mese. [Balestreri, 2012]

I messaggi di stato sono denominati *tweet*, e una volta inseriti divengono immediatamente disponibili sul profilo personale dell'utente, oltre che negli aggiornamenti di stato di coloro che sono diventati *follower* di quell'utente. Le relazioni sono essenzialmente di tipo monodirezionale. È anche possibile inviare messaggi privati agli utenti.

I *tweet* possono essere ricondivisi tramite una funzionalità di *retweet* (RT) ed è possibile rispondere a uno specifico *tweet* tramite la funzionalità di *mention* (@). Non c'è un'organizzazione in conversazioni come Facebook, tuttavia i *tweet* possono essere raggruppati attraverso l'utilizzo di *hashtag* (#). Gli *hashtag* consentono di definire un argomento di discussione e aiutano a richiamare tutti i *tweet* pubblici recenti che lo contengono. La ricerca può comunque essere fatta su qualunque *keyword*, e restituirà una serie di post pubblici di diversi utenti, indipendentemente dal fatto che chi ha effettuato la ricerca sia loro *follower* o meno. Ciò dimostra come Twitter sia essenzialmente un aggregatore di messaggi brevi accomunati spesso da *topic* specifici.

Recenti studi hanno dimostrato il carattere giornalistico di Twitter, sempre più destinato a diventare una piattaforma di riferimento per la diffusione di news.

Myspace

Myspace è una comunità virtuale, e più precisamente una rete sociale, creata nel 2003 da Tom Anderson e Chris DeWolfe. Offre ai suoi utenti *blog*, profili personali, gruppi, foto, musica e video. Gli iscritti sono in grado di creare un proprio profilo inserendo diversi contributi; nella sua prima versione il profilo era una pagina *html* altamente personalizzabile, successivamente fu meglio strutturata per evitare agli utenti meno esperti di manipolare erroneamente il codice, rendendo difficoltosa la navigazione ai visitatori.

L'avvento di *social network* come Facebook e Twitter ha, gradualmente, soppiantato Myspace come piattaforma utilizzata per profili e *blog* personali. Nonostante questo, per diverso tempo il sito ha continuato a essere una vetrina per i gruppi musicali, per via della possibilità di caricare file mp3 che, nei primi tempi, era preclusa ai *social network* concorrenti. Il sito ha goduto del massimo successo tra il 2005 e il 2008, divenendo popolarissimo specialmente tra gli adolescenti. Successivamente, nonostante numerose

modifiche a funzionalità e design, ha continuato a perdere iscritti e nel 2011 è stato annunciato un taglio al personale del 47%. [Marya, 2011]

Una delle *feature* più interessanti fornite da Myspace è quella di poter programmare applicazioni integrabili nel profilo, attraverso specifiche API.

LinkedIn

LinkedIn è un *social network* destinato principalmente allo sviluppo di contatti professionali. Lo scopo principale del sito è consentire agli utenti registrati di mantenere una lista di persone ritenute affidabili in ambito lavorativo. Le persone nella lista sono definite "connessioni". L'utente può incrementare il numero delle proprie connessioni invitando chi di suo gradimento. La rete di contatti a disposizione dell'utente è costituita da tutte le connessioni dell'utente, ma anche da tutte le connessioni delle sue connessioni ("connessioni di secondo grado") e da tutte le connessioni delle connessioni di secondo grado ("connessioni di terzo grado").

Gli scopi sono molteplici, essenzialmente orientati alla conoscenza di offerte lavorative o alla ricerca di possibili candidati da parte dei datori di lavoro. Si tratta dunque di un *network* professionale che raggiunge essenzialmente un'utenza *business*.

Flickr

Flickr è un sito Web che consente ai propri utenti di condividere immagini e fotografie, con una logica da Web 2.0. Esso permette di creare un proprio album, che può essere pubblico o destinato a una fetta di utenti selezionati, dove presentare le proprie foto e ottenere commenti da parte degli altri visitatori. Tramite un sistema sociale, gli utenti possono diventare amici e interagire tra loro. La discussione riguardo alle immagini stimola all'interazione e alla conoscenza tra gli utenti: non di rado, le amicizie nate in Flickr sfociano in rapporti reali.

L'utilizzo del servizio è cambiato parecchio negli anni. Inizialmente nato come strumento per ospitare le proprie immagini da pubblicare su altri siti, ha avuto grande successo grazie al fenomeno dei *blog*. In seguito si è evoluto, diventando esso stesso una comunità virtuale grazie ai gruppi tematici e ai *forum*.

A partire dal 2009, gli utenti possono anche fornire le proprie immagini per scopi commerciali, grazie a un accordo con Getty Images.

Foursquare

Foursquare è un *social network* basato sulla geolocalizzazione, ed è disponibile tramite Web e applicazioni per dispositivi mobili. Permette ai propri utenti di condividere la propria posizione attraverso il *check-in*, eseguito grazie alle applicazioni mobili su terminali dotati di GPS. La condivisione è stimolata dall'attribuzione di premi e *badges* virtuali, che rendono di fatto la piattaforma una sorta di gioco a competizione. Gli utenti possono inoltre creare una lista pubblica di cose da fare e scrivere dei brevi suggerimenti per coloro che eseguiranno *check-in* nel luogo stesso o in quelli vicini.

Le ultime modifiche alla piattaforma sono tendenzialmente orientate a rendere Foursquare un sistema a supporto delle scelte commerciali degli utenti, attraverso consigli mirati basati sui gusti degli stessi. Questi si concretizzano anche grazie a una rete pubblicitaria, che fornisce alle aziende la possibilità di offrire promozioni personalizzate.

1.4 - Il crowdsourcing

Il termine *crowdsourcing* è stato usato per la prima volta da Jeff Howe in un articolo del giugno 2006 per la rivista *Wired*, dal titolo “*The Rise of Crowdsourcing*”.

Nel corso degli anni sono state fornite differenti interpretazioni sul significato del termine, che Howe aveva solo implicitamente spiegato, tanto che nella letteratura scientifica esistono a tutt'oggi oltre quaranta differenti definizioni.

Estellés e González, nel lavoro [Estellés et al., 2012], hanno provato a raccogliere e integrare queste definizioni, al fine di proporre una universale:

« Il crowdsourcing è una tipologia di attività online partecipativa nella quale una persona, istituzione, organizzazione non a scopo di lucro o azienda propone a 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 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 benefi-

cio il contributo offerto dall'utente, la cui forma dipenderà dal tipo di attività realizzata.»

Il *crowdsourcing* può dunque essere visto essenzialmente come un modello di produzione e risoluzione dei problemi. Nell'accezione classica del termine, viene richiesta la risoluzione di un determinato problema a un gruppo non definito di persone. Gli utenti, ovvero la “*crowd*” (“folla”), solitamente si riuniscono in comunità online; tramite esse propongono soluzioni, poi vagliate dal gruppo stesso e approvate dal committente. Queste soluzioni appartengono all'istituzione o all'individuo che ha inizialmente presentato il problema, e gli utenti che hanno contribuito a trovarle in alcuni casi vengono ricompensati in denaro o con premi e riconoscimenti, oppure in altri casi con la semplice soddisfazione intellettuale. Grazie al *crowdsourcing*, le soluzioni possono tanto provenire da utenti non professionisti e/o volontari che vi lavorano nel tempo libero, quanto da esperti e piccole imprese che erano sconosciute all'istituzione committente.

Jeff Howe distingue quattro diverse strategie di *crowdsourcing*:

1. *crowdfunding* (finanziamento collettivo)
2. *crowdcreation* (creazione collettiva)
3. *crowdvoting* (votazione collettiva)
4. *wisdom of the crowd* (saggezza della folla)

I vantaggi del *crowdsourcing* sono molteplici rispetto ai tradizionali modelli aziendali di *problem solving*. Innanzitutto, le soluzioni possono essere trovate a un costo ridotto, talvolta praticamente nullo: il pagamento si basa infatti sul risultato, e a volte non è nemmeno previsto. Le organizzazioni sono poi in grado di recepire suggerimenti da un bacino molto più ampio di quello della loro rete primaria di contatti, e nel contempo hanno la possibilità di comprendere i bisogni dei propri consumatori, che a loro volta risulteranno stimolati e si fidelizzeranno al marchio.

Nel suo articolo intitolato “*The Power of Crowdsourcing*”, Matt H. Evans sostiene che il *crowdsourcing* sfrutti le idee presenti a livello globale, aiutando il lavoro delle aziende attraverso un rapido iter progettuale a costo minimo, dato che gli utenti sono sempre desiderosi di condividere le proprie idee su scala globale.⁴

La pratica del *crowdsourcing* viene però spesso accusata di non produrre sempre ottimi risultati a livello qualitativo, e di essere usata per ottenere del lavoro a costo minimo o

⁴ <http://www.exinfm.com/board/crowdsourcing.htm>

addirittura nullo. Per questo motivo il *crowdsourcing* viene sempre più spesso impiegato all'interno di un modello più ampio di "scambio di servizi creativi", dove gli utenti sono stati previamente selezionati e viene affidato a delle agenzie di mediatori il compito di agevolare lo scambio delle istruzioni in merito al progetto tra l'azienda e il pubblico. [O'Hear, 2011]

Grazie all'aumento delle potenzialità della rete Internet e agli strumenti offerti dal Web, il *crowdsourcing* è oggi esclusivamente mutuato attraverso tali canali. Un esempio importante di *crowdsourcing* sul Web, citato dallo stesso Howe nel libro [Howe, 2009], è rappresentato dal *social bookmarking*. Grazie a questi sistemi gli utenti possono assegnare dei *tag* a delle fonti condivise con altri utenti in modo da organizzare le informazioni.

1.4.1 Il *crowdsourcing* nelle fasi di un'emergenza

L'informazione sociale garantita dal paradigma del *crowdsourcing* può risultare utilissima nella gestione delle emergenze. Essa permette infatti di ricavare dati utili alle autorità, sfruttando i cittadini al pari di "sensori" localizzati sul territorio. Questo tipo di dati vanno a supporto dell'intero ciclo delle emergenze, poiché le informazioni che i cittadini possono fornire sono di varia natura.

Vediamo ora come il paradigma possa essere utilizzato nelle diverse fasi del ciclo di vita di un'emergenza.

Preparedness

Durante la *preparedness*, i cittadini possono fornire informazioni circa lo stato di rischio del territorio, informando le autorità di eventuali infrastrutture critiche, in modo che questi dati vadano a supporto dell'elaborazione del piano d'emergenza. Questa attività dovrebbe essere normalmente coordinata dalle autorità locali, come polizia, carabinieri, vigili del fuoco, che quotidianamente recepiscono lamentele da parte della gente comune. Occorre dunque sfruttare queste segnalazioni, integrandole nella conoscenza globale del territorio e nell'analisi dei rischi, effettuata da parte del personale preposto.

Inoltre, risulta utile in questa fase il cosiddetto *crowdfeeding*, ovvero la attivazione di una serie di strumenti per istruire i cittadini su cosa fare in caso di emergenza.

Response

In questa fase l'informazione sociale va a supporto delle decisioni strategiche di coloro che si occupano di risolvere l'emergenza, fornendo dati circa lo stato dei cittadini colpiti, la situazione generale del territorio (p.es. quali strade sono percorribili, quali zone risultano a rischio, ecc.), e le richieste di intervento. Questi dati, integrati e interpretati con coscienza, possono velocizzare la risposta delle autorità, facendo sì che gli interventi siano mirati e in accordo con le reali richieste delle vittime dell'emergenza.

Inoltre, già durante la *response* è possibile attivare strumenti di *crowdfunding* per raccogliere denaro.

Per quanto riguarda gli enti, il paradigma *crowdsourcing* è utile anche per eseguire un coordinamento più semplice tra i diversi gruppi operativi.

Recovery

Per comprendere al meglio quali siano gli interventi necessari per la ricostruzione, è possibile conoscere i bisogni e le richieste dei cittadini utilizzando il *crowdsourcing*. Non solo, esso può essere usato dalle organizzazioni per trovare soluzioni alle varie problematiche di ricostruzione, avvalendosi dell'aiuto di esperti esterni reclutati in maniera distribuita.

Mitigation

Durante questa lunga fase, il *crowdsourcing* continuativo può proseguire nel lavoro di comprensione delle problematiche sociali successive all'evento, ma anche raccogliere lamentele e consentire quindi di meglio definire l'operato dei diversi soggetti durante le fasi precedenti.

Inoltre, i dati raccolti durante questo periodo "di calma" andranno a integrarsi a quelli della successiva fase di preparazione.

2 - Classificazione delle emergenze

Nel momento in cui un'organizzazione decide di costruire un sistema di gestione delle emergenze, il primo aspetto che deve prendere in considerazione è quello legato alle definizioni inerenti l'emergenza stessa. Tali definizioni sono strettamente legate al settore di attività e ai servizi erogati dall'organizzazione, ma in generale la situazione di emergenza è collegata all'indisponibilità temporanea o permanente di risorse critiche, siano esse infrastrutturali o umane. [ISCOM]

Riuscire a classificare le emergenze, ovvero a indicarne parametri fondamentali come origine, evoluzione, entità coinvolte, risulta utile nella definizione delle migliori strategie di contenimento e risoluzione delle stesse.

Non tutti gli eventi critici che si verificano sul territorio richiedono piani d'intervento predefiniti. Alcuni eventi possono essere incidenti banali, risolvibili in maniera rapida e senza alcuna preparazione preventiva. In altri casi possono invece verificarsi casi di criticità più elevata, che daranno dunque l'innescò a situazioni di "crisi" o addirittura di "disastro", e che richiederanno perciò piani operativi mirati e necessariamente più elaborati, con lo sforzo da parte di più entità.

I cosiddetti "piani d'emergenza", che i governi e gli enti di protezione civile elaborano allo scopo di gestire con l'adeguata preparazione eventuali situazioni critiche, non possono nemmeno essere universali e immutabili, e richiedono invece opportune modifiche ricavabili con l'esperienza o lo studio attento di situazioni passate. Ad esempio, dopo i tragici accadimenti dell'11 Settembre 2001, il Governo Americano ha imposto una revisione delle procedure di sicurezza in tutte le infrastrutture critiche del territorio (aeroporti, luoghi pubblici, frontiere, ecc.), allo scopo di evitare la ripetizione di situazioni simili, per le quali si generò il caos a causa di una inadeguata preparazione ad accadimenti inaspettati di quel tipo.

2.1 - Il grado d'impatto di un evento

Un importante criterio di classificazione di un evento critico è quello che tiene conto dell'impatto che esso produce.

La Commissione Europea fornisce la seguente definizione di impatto:

“L'impatto è la somma totale dei diversi effetti di un incidente.”

L'impatto fornisce dunque una misura quantitativa e qualitativa della potenza intrinseca di un dato evento critico, ma anche delle misure di mitigazione dello stesso, dato che gli effetti di un incidente sono per l'appunto condizionati dalla risposta delle entità che si occupano della sua risoluzione.

I “diversi effetti” citati nella definizione richiedono di tenere in considerazione alcuni elementi fondamentali. In particolare, la Commissione Europea introduce tre dimensioni di analisi per la corretta descrizione dell'impatto che un dato incidente produce:

- *severity*, ovvero il livello di gravità;
- *scope*, ovvero l'estensione geografica;
- *effects of time*, ovvero la durata temporale.

Severity

La *severity* misura quantitativamente il numero di perdite in seguito a un dato incidente. Essa è costituita da una scala di quattro valori: *None*, *Minimal*, *Moderate* o *Major*.

Le perdite indicate non sono necessariamente perdite umane: come detto, un incidente può generare una situazione di emergenza la cui definizione dipende strettamente dalla struttura dell'organizzazione che si occupa della sua gestione. Una perdita può dunque essere di tipo infrastrutturale, se nella definizione dell'emergenza viene indicata una criticità in tal senso.

Scope

Scope indica l'estensione sociale e/o geografica affetta da un incidente. In termini d'impatto, può indicare l'estensione di un'area geografica entro la quale sono visibili effetti di criticità dovuti a perdite infrastrutturali.

Effects of time

Effects of time è indice della durata temporale di un evento critico, con particolare riferimento al tempo in cui sono visibili effetti d'impatto dell'evento stesso. Per esempio, un evento può essere improvviso, ma determinare effetti disastrosi che perdurano per diverso tempo. Occorre perciò una doppia analisi: da un lato quella che permette di definire la durata dell'evento in sé, in modo da prevedere un piano di soccorso che tenga conto della pericolosità nell'immediato, dall'altro quella che aiuta a valutare il protrarsi degli effetti al fine di determinarne il rischio futuro e definire un piano di ricostruzione una volta terminata l'emergenza.

Ulteriori criteri per la definizione degli impatti

Accanto ai tre indici fondamentali, possono essere utilizzati altri criteri per la corretta definizione degli impatti. Tali criteri tengono conto degli effetti di tipo:

- pubblico (numero di persone coinvolte, perdite umane, questioni mediche, feriti, evacuazioni, ecc.);
- economico (significative perdite di materiali e prodotti, impatto sugli indici economici, mancanza di bisogni primari, problemi energetici, ecc.);
- ambientale (effetti sulle zone limitrofe all'area colpita);
- di interdipendenza (tra diverse infrastrutture);
- politico (fiducia nelle capacità di un governo);
- psicologico (reazione delle persone nella fase successiva all'evento).

2.1.1 - Le classi d'impatto

La combinazione degli elementi descritti produce una suddivisione in tre classi principali, codificando l'impatto come *lieve*, *moderato*, e *alto*.

L'impatto è definito *lieve* se:

- (i) causa una degradazione nelle capacità di un'organizzazione o di una comunità che è di lieve durata ed estensione, e lascia pressoché inalterate le infrastrutture primarie dell'organizzazione stessa, che quindi è autonomamente in grado di reagire;
- (ii) produce danni minimi;
- (iii) produce perdite economiche minime;
- (iv) produce ferimenti lievi.

L'impatto è definito *moderato* se:

- (i) causa una degradazione nelle capacità di un'organizzazione o di una comunità che è di lieve durata ed estensione, e colpisce solo parzialmente le infrastrutture primarie dell'organizzazione stessa, ma l'efficacia delle funzioni da essa svolte risulta significativamente compromessa;
- (ii) produce danni significativi;
- (iii) produce perdite economiche significative;
- (iv) produce ferimenti in alto numero, ma non perdite di vite umane.

L'impatto è definito *alto* se:

- (i) causa una degradazione severa nelle capacità di un'organizzazione o di una comunità, tanto da impedire alla stessa di perseguire una o tutte le proprie funzioni primarie;
- (ii) produce danni gravi;
- (iii) produce perdite economiche molto elevate;
- (iv) produce ferimenti gravi e perdite di vite umane.

È possibile suddividere ulteriormente il livello di gravità dell'impatto in cinque diversi livelli, che si concentrano in particolare sui danni economici derivanti e sulle funzioni primarie della collettività che ne è colpita:

1. irrilevante
2. lieve
3. importante
4. molto grave
5. catastrofico.

La descrizione dei vari livelli è riportata in tabella 1.

Tabella 1 - Classificazione degli eventi sulla base dell'impatto

LIVELLI DI GRAVITÀ	DESCRIZIONE IMPATTO
1 IRRILEVANTE	Incidente che provoca un disturbo ma senza conseguenze nei confronti della collettività. Non ha impatti di ordine economico.
2 LIEVE	Incidente con impatto di entità minore. Può creare disagi alla collettività e comunque di impatto economico contenuto.
3 IMPORTANTE	Incidente di Impatto rilevante. Ha conseguenze sulle funzioni primarie della collettività e genera perdite economiche non trascurabili.
4 MOLTO GRAVE	Incidente con conseguenze molto rilevanti. Provoca ingenti danni economici con rilevanti conseguenze sulla capacità della collettività nel breve e medio termine.
5 CATASTROFICO	Incidente che reca un danno di elevatissima entità economica. Si accompagna a una sostanziale e completa distruzione dei mezzi della collettività. Il recupero, se possibile, è attuabile solo nel lungo termine e a fronte di ingenti investimenti.

2.1.2 - Definizione di disastro sulla base dell'impatto

Nel momento in cui un'organizzazione si appresta a definire una classificazione di un evento in relazione all'impatto, risulta spesso difficile definire i confini tra una classe e un'altra e soprattutto definire il contesto nel quale l'impatto è tale da dover parlare di disastro. Le classificazioni descritte nei paragrafi precedenti ci riportano a definire "disastro" un evento il cui l'impatto è classificato come alto o catastrofico (gli aggettivi utilizzati a livello internazionale sono: *high, severe, fatal, catastrophic*). Un impatto di questo tipo produce danni ingenti alle infrastrutture, determina perdite economiche rilevanti, e in caso di eventi che coinvolgono soggetti umani, determina perdite di vite e numerosi feriti.

2.2 - Classificazione degli eventi secondo la scala

Una diversa modalità di classificazione, proposta da Jul in [Jul, 2009], utilizza alcuni degli indici del paragrafo precedente concentrandosi maggiormente sulla dimensione sociale all'interno della quale un dato evento impatta.

Jul utilizza tre indicatori fondamentali: *magnitude*, *scope*, e *duration of the impact*.

Magnitude indica la severità della disgregazione sociale (*social disruption*) e dei danni fisici, ovvero definisce un parametro che permette di comprendere l'ampiezza del gruppo di soggetti ai quali l'evento ha imposto una modifica sostanziale delle abitudini di vita.

Scope indica i confini sociali e geografici della disgregazione sociale e dei danni fisici. *Duration*, infine, è una misura del tempo intercorso tra l'inizio del fenomeno di disgregazione sociale e dei danni fisici e il momento in cui il disastro smette di produrre tali effetti.

Grazie a questi parametri dimensionali, possiamo definire tre diverse categorie di evento secondo il parametro di scala:

- un'**emergenza** è un evento molto breve i cui effetti sono localizzati all'interno di una singola comunità. La comunità e le sue infrastrutture rimangono pienamente funzionanti, ed essa è autonomamente in grado di risolvere con successo l'evento;
- un **disastro** è un evento di durata maggiore che colpisce una singola comunità, lasciandone le infrastrutture di risoluzione sostanzialmente intatte. Tuttavia, in caso di disastro difficilmente la comunità colpita è in grado di reagire autonomamente, e necessita spesso dell'aiuto da parte di altre comunità;
- una **catastrofe** è un evento di lunga durata che colpisce più comunità, provocando danni ingenti alle infrastrutture e ai sistemi di risoluzione. Le comunità non sono in grado di reagire autonomamente e spesso competono con quelle vicine per ottenere aiuti esterni, piuttosto che aiutarsi vicendevolmente.

Per quanto riguarda le organizzazioni preposte alla risoluzione delle emergenze, Jul propone due modelli, la cui distinzione possiede una rilevanza sulle modalità operative durante l'attività di risoluzione:

- le **organizzazioni costituite** (*established organizations*) si attivano nella fase di risposta e la loro struttura organizzativa non si modifica durante le fasi di risoluzione;
- le **organizzazioni in espansione** (*expanding organizations*) si attivano anch'esse nella fase di risposta, ma necessitano di espandere la propria struttura operativa, p.es. reclutando volontari.

Queste definizioni sono di tipo generale e dipendono dal contesto e dal paese nel quale tali organizzazioni sono localizzate. Indipendentemente dalle scelte individuali, è possibile comunque dire che le organizzazioni costituite si affidano a una struttura rigida e a una serie di operatori esperti e fortemente addestrati a un lavoro che svolgono come carriera, mentre le organizzazioni in espansione necessitano di affidarsi a volontari, i quali, pur avendo ricevuto specifiche istruzioni su come comportarsi, generalmente svolgono altri tipi di lavoro nella propria quotidianità.

È ora possibile dare due ulteriori definizioni di scala agli eventi, in base al tipo di organizzazioni che partecipano alla loro risoluzione:

- le **emergenze locali** sono gestite interamente da organizzazioni costituite;
- i **disastri locali** richiedono il coinvolgimento delle organizzazioni in espansione, che dovranno a loro volta affidarsi ad alcuni volontari durante la fase di risoluzione.

2.3 - Classificazione degli eventi secondo il tipo

La seconda dimensione che definisce le caratteristiche di un disastro è il **tipo**.

Il tipo indica essenzialmente le diverse tipologie di effetto che un dato evento produce.

Un primo aspetto riguarda il *che cosa* viene colpito, domanda che origina questa classificazione:

- i **disastri comunitari** (*community disasters*) colpiscono un ampio insieme di risorse fisiche e umane;
- i **disastri settoriali** (*sector disasters*) colpiscono principalmente un segmento specifico della comunità, e possono essere trattati da professionisti.

Questo tipo di classificazione è importante perché permette anche di definire le figure di riferimento che dovranno poi occuparsi della risoluzione del disastro, indicando anche la necessaria preparazione utile allo scopo.

Un altro aspetto che riguarda il tipo di disastro è la cosiddetta **agenda sociale** (*social agenda*), che descrive il contesto sociale all'interno del quale ha luogo la risoluzione dell'evento. Questo aspetto produce una distinzione tra:

- **eventi *consensus-type***, nei quali c'è un accordo generale rispetto agli obiettivi relativi alla risoluzione;
- **eventi *conflict-type***, per i quali i diversi soggetti preposti alla risoluzione si muovono perseguendo obiettivi differenti, spesso conflittuali (p.es. riportare la situazione alla normalità piuttosto che ridefinire il concetto di normalità).

2.4 - Classificazione degli eventi secondo l'anticipabilità

La dimensione finale relativa ai disastri considerata da Jul è denominata *anticipabilità*, e risulta essere un parametro fondamentale nella fase di preparazione all'emergenza. Come osservato in diversi casi reali, gli effetti prodotti da un dato evento possono essere molto diversi a seconda di quanto quell'evento risulti o meno atteso, ovvero di quanto le organizzazioni risultino preparate e sufficientemente istruite riguardo a esso. Ritornando al già citato caso dell'11 Settembre 2011, molte delle tragiche conseguenze si sarebbero potute evitare se il folle evento scatenante (un aereo che viene dirottato e lanciato contro una torre piena di migliaia di persone) fosse stato *predicibile*.

La **predicibilità** (*predictability*) di un evento è maggiore quando esso può essere immaginato, e nel momento della sua occorrenza viene socialmente accolto come credibile. Un aereo civile usato come bomba e lanciato contro un grattacielo del centro di New York, prima del 9/11, era decisamente un'ipotesi poco credibile. La sua occorrenza, tuttavia, ha fortemente modificato le idee delle persone in tal senso, portando a una successiva ridefinizione delle misure di sicurezza americane all'interno degli aeroporti. Questo dimostra che la predicibilità non è concetto immutabile, ma si modifica con l'eventuale realizzazione di eventi fino a quel momento considerati impossibili.

Un altro aspetto riguarda l'**influenzabilità** (*influenceability*) di un evento. Essa misura quanto realistiche e realizzabili siano le strategie di riduzione dei danni causati

dall'evento stesso, date le risorse e il contesto sociopolitico nel momento e nel luogo in cui esso si verifica.

Combinando predicibilità e influenzabilità, è possibile ricavare una classificazione degli eventi secondo l'anticipabilità. Rispettivamente, abbiamo quindi eventi:

- **convenzionali** (*conventional*): facilmente influenzabili e facilmente predicibili;
- **inaspettati** (*unexpected*): facilmente influenzabili e difficilmente predicibili;
- **intrattabili** (*intractable*): difficilmente influenzabili e facilmente predicibili;
- **fondamentali** (*fundamental*): difficilmente influenzabili e difficilmente predicibili.

2.5 - Elenco di eventi critici

Anche se paesi differenti sono normalmente colpiti da diverse tipologie di eventi critici, la letteratura tende a identificare due precise categorie a seconda dell'origine degli stessi:

- **Eventi di origine naturale**: causati da fenomeni naturali, come per esempio processi meteorologici, geologici, o di altro tipo. Gli eventi naturali sono spesso suddivisi in ulteriori sottocategorie, come p.es. climatici e ambientali.
- **Eventi di origine umana/tecnologica**: sono il prodotto dell'innovazione tecnologica e dello sviluppo umano. Normalmente, per questi eventi, gli studi a riguardo sono più complessi e richiedono un costante aggiornamento che va di pari passo con l'evoluzione scientifica.

2.5.1 - Gli eventi di origine naturale

L'elenco che segue è quanto di più esaustivo è stato possibile estrarre dalla letteratura. Esso non rappresenta un elenco completo di tutti i possibili eventi di origine naturale, ma condensa gli eventi principali, fornendone una dettagliata spiegazione e cercando di evidenziare i legami e le concause presenti tra essi. È opportuno infatti notare che alcuni di questi eventi sono strettamente connessi, potendosi originare sia come conseguenza l'uno dell'altro, sia contemporaneamente.

Inondazioni e alluvioni

Un'inondazione è un fenomeno riguardante l'allagamento in tempi brevi (da ore a giorni) di un'area ben definita e abitualmente asciutta, da parte di una massa d'acqua.

L'origine dell'inondazione può essere molteplice: straripamento di corsi d'acqua, piogge insistenti e in volume maggiore rispetto alla media del luogo, scioglimenti di nevai e ghiacciai, tsunami, ecc. In taluni casi può avere origine umana. Quando l'inondazione è determinata dalla tracimazione di corsi d'acqua ingrossati per piogge elevate si parla anche di alluvione, a cui possono essere connessi anche fenomeni di erosione e variazione della morfologia delle aree interessate dal fenomeno. La parola alluvione nell'italiano arcaico veniva infatti utilizzata per indicare un evento di accumulo di materiale fluviale, anche se attualmente è talvolta sfruttata come sinonimo di inondazione in genere.

In determinate aree del mondo, a causa delle condizioni geologiche e meteorologiche della zona, le inondazioni risultano più frequenti. Tali aree possono essere mappate, allo scopo di meglio definire dei piani d'emergenza locali. A volte però la mitigazione di questi fenomeni non risulta possibile o economicamente conveniente, come nel caso del devastante uragano Katrina del 2005, che distrusse completamente la città di New Orleans, determinando un ingente numero di perdite umane. [Fischetti, 2001]

Si è osservato però come, nei paesi che applicano ridotte politiche di mitigazione del fenomeno, il numero di persone uccise o colpite da esso risulti nettamente maggiore rispetto a quello dei paesi in cui esistono piani preventivi ben strutturati.

Molti piani strategici sono stati studiati per ridurre il rischio di alluvioni nelle aree più pericolose: essi vanno dall'adozione di misure progettuali per contenere gli argini dei corsi d'acqua, alla promulgazione di leggi che impediscono la costruzione di edifici ai bordi degli stessi.

Terremoti

In geofisica i terremoti, detti anche sismi o scosse telluriche, sono vibrazioni o oscillazioni improvvise della crosta terrestre, rapide e più o meno potenti, provocate dallo spostamento di una massa rocciosa nel sottosuolo.

La fonte del terremoto è generalmente distribuita in una zona interna della crosta terrestre — nel caso dei terremoti più devastanti questa può avere un'estensione dell'ordine di un migliaio di chilometri — ma è normalmente possibile identificare un punto preciso dal quale le onde sismiche hanno avuto origine, chiamato *ipocentro*; la sua proiezione verticale sulla superficie terrestre è invece chiamata *epicentro*, il punto in cui di solito si verificano i danni maggiori.

Le onde elastiche che si propagano durante un terremoto sono di diverso tipo e in alcuni casi possono risultare in un movimento prevalentemente orizzontale o verticale del terreno (scossa ondulatoria o sussultoria). Le onde sismiche sono rilevabili e misurabili attraverso particolari strumenti detti *sismografi*; l'elaborazione incrociata dei dati di più sismografi, sparsi su un territorio a una certa distanza dal sisma, consente di stimare in maniera abbastanza accurata l'epicentro, l'ipocentro e l'intensità del fenomeno. Non è tuttavia possibile prevedere con anticipo un sisma: per questi motivi è pressoché impossibile adottare una strategia preventiva di evacuazione, mentre risulta più efficiente definire piani di reazione allo scoppio dell'evento.

I terremoti sono gli eventi naturali di gran lunga più potenti sulla terra, e possono rilasciare un'energia molto elevata in un intervallo di tempo estremamente breve. In conseguenza di ciò i terremoti possono causare gravi distruzioni e alte perdite di vite umane attraverso una serie di agenti distruttivi, il principale dei quali è il movimento violento del terreno con conseguente sollecitazione delle strutture edilizie in posa (edifici, ponti ecc.), accompagnato eventualmente anche da altri effetti secondari quali inondazioni (ad esempio cedimento di dighe), cedimenti del terreno (frane, smottamenti o liquefazione), incendi o fuoriuscite di materiali pericolosi. Se il sisma avviene sotto la superficie oceanica o marina o nei pressi della linea costiera può generare maremoti o tsunami. In ogni terremoto uno o più di questi agenti possono dunque concorrere a causare ulteriori gravi danni e vittime.

Uragani e tornado

Il termine *uragano* indica un *ciclone tropicale*, ovvero un sistema tempestoso caratterizzato da un largo centro di bassa pressione e da numerosi temporali che producono forti venti e pesanti piogge. A volte è impiegato per definire un vento di forza eccezionale, corrispondente al massimo grado della scala di Beaufort, una misura empirica dell'intensità del vento basata sullo stato del mare aperto distante dalle coste o sulle condizioni delle onde.

Gli uragani atlantici si formano nell'Oceano Atlantico, usualmente nell'emisfero nord, in estate o autunno, con venti sostenuti per al massimo un minuto a 250km/h. La maggior parte delle tempeste tropicali e degli uragani si formano tra l'1 giugno e il 30 novembre. Il *National Hurricane Center* degli Stati Uniti monitora il bacino e pubblica resoconti,

fornendo aggiornamenti riguardo ai fenomeni meteorologici tropicali che si formano nel bacino atlantico.

Le città situate nelle zone normalmente colpite da uragani possiedono piani d'evacuazione ben definiti. Tuttavia, in alcuni casi questi piani sono talmente costosi da non poter essere attuati in pratica, com'era nel caso di New Orleans prima che la tragedia dell'uragano Katrina imponesse una revisione completa dei sistemi d'emergenza. [Fischetti, 2001]

Una *tromba d'aria* o *tornado* è invece un violento vortice d'aria che si origina alla base di un cumulonembo e giunge a toccare il suolo.

Le trombe d'aria sono fenomeni meteorologici altamente distruttivi e nell'area mediterranea rappresentano il fenomeno verificabile più violento, sia pure con frequenza non elevata. Sono associati quasi sempre a temporali estremamente forti, possono percorrere centinaia di chilometri e generare venti anche di 500 km/h.

La distruttività di un tornado si calcola in base alla sua durata, velocità e intensità dei venti. Una tromba d'aria mediamente dura dai 5 ai 15 minuti, ma in alcuni casi, in relazione alla sua intensità, può arrivare a durare anche più di un'ora. Il fenomeno tornadico, per quanto possa verificarsi ovunque ad esclusione delle zone polari, si presenta in alcune zone geografiche particolari che possiedono le condizioni ottimali per il suo sviluppo. La più famosa è la cosiddetta "*Tornado Alley*" negli Stati Uniti, che risulta essere il paese maggiormente colpito da questo fenomeno per frequenza e intensità. Anche l'Italia rientra a pieno titolo nelle zone maggiormente interessate da trombe d'aria. L'intensità media dei fenomeni tornadici sulla penisola è di norma molto più bassa dei tornado americani, anche se non mancano nella storia della meteorologia italiana trombe d'aria di notevole potenza. Secondo recenti studi, l'incidenza delle trombe d'aria in Italia è destinata a salire, mantenendo stabili i ritmi di crescita che si sono verificati negli ultimi 30 anni.

Un possibile accorgimento per minimizzare i danni in caso di tornado è quello di costruire edifici che possano resistere alla forza distruttiva di questi eventi.

Eruzioni vulcaniche

Un vulcano è una spaccatura della superficie terrestre in corrispondenza della quale si ha la fuoriuscita di magma tramite un'eruzione. L'accumulo dei prodotti emessi durante

le eruzioni (lave, piroclastiti, ecc.) genera la caratteristica forma conica dell'edificio vulcanico.

Gli effetti di un'eruzione vulcanica possono causare distruzione e danni in molti modi. Essi variano a seconda della tipologia di vulcano. Il fiume di lava e l'esplosione di materiale roccioso e incandescente possono provocare ferimento o morte alle persone o danni alle costruzioni, e le ceneri possono produrre soffocamento se respirate. Alcuni effetti meno drammatici hanno comunque impatto sulla comunità: p.es., durante l'eruzione del vulcano islandese Eyjafjöll nel 2010, la compromessa visibilità impedì la circolazione aerea nelle aree limitrofe, a causa della nube di fumo che per giorni continuò a fuoriuscire dalla bocca del vulcano.

Anche le eruzioni, come i terremoti, non possono essere predette con certezza; è tuttavia possibile monitorare l'attività vulcanica in modo da favorire una celere risposta in caso si osservi un'attività anomala che faccia presagire un'eruzione imminente.

Frane

Con il termine frana si indicano tutti i fenomeni di movimento o caduta di materiale roccioso o sciolto dovuti alla rottura dell'equilibrio statico preesistente ovvero all'effetto della forza di gravità che, agendo su di esso, supera le forze opposte di coesione del terreno. Le frane possono essere più o meno pericolose a seconda dell'area in cui si verificano, dell'ampiezza del territorio franante, della densità di popolazione colpita, ecc.

In generale, le frane possono provocare perdite di vite umane, ma anche danni alla collettività qualora si verificano in prossimità di infrastrutture come strade, ponti, ferrovie. Inoltre, frane che si verificano in prossimità di corsi d'acqua possono produrre ostruzione agli stessi, portando a successive esondazioni.

Tempeste invernali di ghiaccio e neve

Durante la stagione invernale, soprattutto in zone di clima rigido, possono originarsi vere e proprie tempeste di ghiaccio e neve. Anche se fenomeni stagionali di questo tipo raramente causano morti, essi possono seriamente danneggiare le strade e impedire la circolazione, modificando dunque le abitudini delle comunità colpite.

Per quanto questi fenomeni non siano evitabili, avere degli strumenti sociali che raccolgano le segnalazioni dei cittadini potrebbe velocizzare le attività di aiuto da parte delle

autorità, fornendo alle comunità uno strumento per identificare i livelli di pericolosità delle varie zone cittadine e agire laddove ve ne fosse bisogno.

Siccità

La siccità è definita come una prolungata mancanza di acqua, dovuta primariamente a pioggia insufficiente o a temperature eccezionalmente elevate. La siccità può provocare danni all'agricoltura e ridurre i livelli delle riserve di acqua. Purtroppo non esistono conclusioni specifiche riguardo al problema, né una scala universalmente accettata per definire i rischi connessi alla siccità, che risulta quindi un fenomeno localizzato la cui risoluzione non segue schemi universali.

Temperature estreme

Tanto le alte quanto le basse temperature possono provocare danni, sia agendo sulle persone, sia portando a conseguenze sul territorio, sia provocando perdite economiche in caso perdurino per lungo tempo.

Le ondate di caldo uccidono milioni di persone ogni anno (in media 1500 persone all'anno negli Stati Uniti).

Di contro, le temperature gelide possono provocare morti per ipotermia, soprattutto per quei soggetti che vivono in condizioni precarie o sono privi di alloggio.

Le temperature estreme possono generare vere e proprie emergenze nel caso producano effetti sull'economia di un paese. Il caldo o il freddo gelido possono avere influssi sull'agricoltura, sull'allevamento, ma anche sul settore dei trasporti e su aspetti socio-economici in genere.

Incendi

Gli incendi rappresentano eventi che è possibile considerare tanto naturali quanto prodotti dall'uomo.

Gli incendi naturali possono originarsi a causa delle intense temperature, dei fulmini, dei vulcani, dei forti venti, ecc.

Gli incendi originati dall'uomo derivano invece da incidenti o possono essere volontariamente appiccati, ed avere quindi natura dolosa.

Gli incendi possono verificarsi in zone non abitate da esseri umani, p.es. foreste, e determinare quindi danni agli ecosistemi animali e vegetali. Altre volte gli incendi si denominano *strutturali*, perché colpiscono strutture di costruzione umana, determinando

così danni alle stesse e causando morti e feriti. Gli incendi strutturali sono molto frequenti nei contesti urbani e una pronta risposta da parte delle autorità può salvare molte vite e limitare il più possibile i danni.

2.5.2 - Gli eventi di origine umana/tecnologica

Come già accennato, alcuni eventi che determinano emergenze possono essere frutto dell'uomo, sia in maniera volontaria che involontaria. Nel primo caso l'esecutore compie un crimine con il chiaro scopo di generare del caos e seminare il panico, o semplicemente per colpire un determinato bersaglio con dei metodi che in seguito possono creare tragiche conseguenze per la collettività. Nel secondo caso gli incidenti possono derivare da errori, oppure essere il frutto di uno smodato avanzamento tecnologico e costruttivo che, modificando maldestramente ambienti naturali preesistenti, determina poi una risposta incontrollata da parte di essi. Spesso quindi gli eventi di origine umana hanno intrinsecamente una concausa naturale.

Dighe

Un esempio molto eloquente di quest'ultimo punto è costituito dalle dighe. Esse sono una barriera che mira alla modifica artificiale del corso dei fiumi o al contenimento di bacini idrici, con diversi scopi finali: costruzione di centri abitativi nelle zone percorse dai corsi d'acqua, accumulo di grossi volumi di acqua per la produzione di energia elettrica, ecc.

I problemi con le dighe possono verificarsi per diversi motivi. Come tutte le strutture, esse possono risentire dei segni del tempo, essere danneggiate a causa di terremoti o collisioni, o addirittura essere state progettate male. Le dighe hanno pertanto bisogno di una manutenzione continua, solitamente gestita sia da enti pubblici che privati, allo scopo di evitare catastrofi. Spesso i problemi nascono anche in conseguenza ad altri eventi naturali. Le forti piogge e le inondazioni possono originare uno straripamento oltre la diga o addirittura, qualora l'acqua acquisisse una forza dirompente, distruggerla. Si pensi alla tragedia del Vajont del 9 Ottobre 1963.

Epidemie

Le epidemie sono malattie infettive che colpiscono quasi simultaneamente una collettività di individui con una ben delimitata diffusione nello spazio e nel tempo, e causano un numero di infettati superiore rispetto ai valori attesi e con la tendenza ad aumentare.

Includono sia infezioni virali che batteriche. Per quanto virus e batteri siano essenzialmente prodotti naturali, le epidemie sorgono in relazione a specifiche condizioni igienico-sanitarie determinate dall'uomo. Pertanto, nonostante le epidemie non possano essere controllate, è possibile invece agire sulle condizioni che le originano in modo da ridurre al minimo il rischio.

Incidenti stradali

Gli incidenti stradali determinano ogni anno migliaia di morti nel mondo. Essi possono essere causati da imperizia dei guidatori, oppure nascere in conseguenza a una cattiva manutenzione stradale o della segnaletica. Gli incidenti possono avvenire anche con bassa densità di traffico, e avere conseguenze molto gravi anche qualora coinvolgessero un ridotto numero di veicoli. Per ridurre l'incidenza, è possibile tenere traccia delle condizioni stradali e intervenire tempestivamente in quei casi in cui un qualsiasi problema potrebbe determinare il rischio di incidente.

Caduta di edifici

La caduta degli edifici e di altre strutture può essere determinata tanto da una cattiva progettazione quanto da eventi naturali come terremoti, frane, inondazioni, ecc.

La caduta di edifici può determinare vittime. Per evitare che gli edifici cadano, occorre studiare con dovizia di particolari i migliori meccanismi costruttivi, stante la zona in cui le costruzioni saranno edificate.

Violenza e crimine

Violenza e crimine rappresentano un problema in molte città, e la loro incidenza è solitamente connessa al grado di benessere della città stessa, al contesto sociale di chi la abita, e all'attività del Paese nel contrastare efficacemente questi fenomeni.

La violenza e il crimine non sono emergenze di per sé, ma possono diventarlo qualora i fenomeni risultassero incontrollabili o sfociassero in manifestazioni improvvise di grossa portata.

Mancanza di corrente elettrica

In un mondo ormai dipendente dalle tecnologie, la mancanza di corrente elettrica può causare una vera e propria paralisi delle attività umane. Essa infatti è necessaria in tutti gli ambienti lavorativi, e un *blackout* prolungato potrebbe avere un'incidenza economi-

ca enorme. Inoltre, la mancanza di energia elettrica può risultare pericolosa negli ambienti ospedalieri e di soccorso.

2.6 - Le emergenze in ambito cittadino

Secondo quanto visto relativamente alla classificazione dei possibili eventi, vogliamo ora estrarre delle considerazioni riferibili alle sole emergenze che possono verificarsi in ambito cittadino.

Richiamando la definizione proposta da Jul di “emergenza”, essa è un evento di durata molto breve che colpisce una singola comunità, i cui effetti sono pertanto transitori e risolvibili da parte delle organizzazioni della comunità stessa.

Alcuni degli eventi proposti in dettaglio nella sezione precedente possono avere una dimensione variabile a seconda della loro intensità o della zona in cui essi si verificano. È difficile dunque escludere l’uno o l’altro evento sulla base delle sole caratteristiche intrinseche dello stesso. I terremoti, ad esempio, possono tanto affliggere una comunità cittadina con danni lievi – e avere dunque caratteristiche di emergenza propriamente detta – oppure possono portare a danni più gravi che colpiscono più comunità o governi, assumendo pertanto le caratteristiche di “disastro” se non addirittura di “catastrofe”.

La restrizione che opereremo sarà dunque basata su un criterio logico che si fonda sui seguenti punti:

- (i) l’evento considerato può concretamente verificarsi in un ambito cittadino;
- (ii) l’evento considerato deve avere un grado d’impatto basso, ovvero coinvolgere una comunità ristretta, determinare poche o nessuna perdita umana, avere effetti limitati nel tempo;
- (iii) l’evento considerato ha le caratteristiche di emergenza locale;
- (iv) l’evento considerato è predicibile ed esistono piani definiti per la sua risoluzione.

In accordo con queste considerazioni, passiamo ora in revisione gli eventi catalogati precedentemente, partendo da quelli di origine naturale.

Fenomeni come siccità, temperature estreme, uragani, hanno normalmente una dimensione che supera i confini cittadini, assumendo spesso quelli nazionali. La siccità produce danni economici con effetti normalmente durevoli nel tempo, determinando un grado di impatto troppo consistente. Per quanto riguarda gli uragani, invece, essi non rappre-

sentano un problema di rilievo italiano, non essendo l'Italia un paese interessato dal fenomeno.

Fenomeni come terremoti o eruzioni vulcaniche, al contrario, rivestono un importante punto critico del nostro paese. Come detto, questi eventi possono essere considerati solo parzialmente cittadini, ma val la pena di tenerli in conto nella nostra analisi.

Inondazioni, alluvioni, frane e tornado colpiscono spesso singole aree metropolitane o collettività ridotte, e rientrano quindi a pieno titolo nelle possibili emergenze cittadine, per quanto esse possano verificarsi con diverse probabilità.

Le tempeste invernali sono sicuramente un problema rilevante. Ghiaccio e neve possono causare rischi che i Comuni devono saper prevenire o risolvere.

Gli incendi, sia di origine naturale che umana, rappresentano alcune delle emergenze più diffuse. In particolare, risultano rilevanti gli incendi strutturali, che per le caratteristiche di ridotta espansione soddisfano certamente i quattro punti indicati.

Per quanto riguarda gli eventi di origine umana/tecnologica, escludiamo immediatamente le epidemie. Esse richiedono innanzitutto una conoscenza molto settoriale del problema, e sono eventi assai improbabili e certamente di rilevanza non cittadina.

Gli incidenti stradali, la caduta di edifici, la mancanza di corrente elettrica sono sicuramente emergenze interessanti in ambito cittadino quando i loro effetti investano più persone e richiedano risorse comunitarie per la loro risoluzione. La violenza e il crimine, parimenti, rivestono un importante interesse comunitario e sono dunque da prendere in considerazione.

Le dighe, infine, sono normalmente costruite in territori non metropolitani, e quindi risultano avere uno scarso interesse in tale ambito. Le possibili emergenze derivanti da esse, poi, si verificano in caso di imperizia nel monitoraggio della struttura o piuttosto per eventi naturali anomali come quelli che abbiamo già considerato e incluso in elenco.

Possiamo poi integrare altri fenomeni. L'interruzione di infrastrutture primarie può generare situazioni di emergenza: si pensi alla mancanza di acqua in determinate zone, o alla forzata interruzione dei trasporti cittadini, che genera un impatto sulla qualità complessiva della vita. Anche gli incidenti sulle stesse linee di trasporto possono risultare materia da trattare.

Stante queste considerazioni, possiamo allora stilare un elenco delle principali emergenze in ambito cittadino:

- Inondazioni
- Alluvioni
- Tornado
- Ghiaccio e neve
- Incidenti stradali
- Incendi strutturali
- Caduta di edifici
- Violenza e crimine
- Interruzione di servizi di sostentamento (p.es. mancanza d'acqua o di energia elettrica)
- Problemi connessi ai servizi di trasporto pubblico

A questo elenco possono essere aggiunti, con le opportune considerazioni già messe in luce, anche:

- Frane
- Terremoti
- Eruzioni vulcaniche

3 - Ruolo dei *social network* nella gestione delle emergenze

In questo capitolo si passeranno in rassegna una serie di esempi che possano aiutare a comprendere l'utilizzo dei *social network* all'interno di reali scenari di emergenza, allo scopo di offrire una testimonianza concreta di come essi abbiano giocato un ruolo importante se non addirittura fondamentale per la loro gestione o risoluzione.

Innanzitutto, si spiegherà come le reti sociali possano costituire una nuova strategia di comunicazione per i cittadini e il motivo per il quale essi offrono una più efficiente distribuzione delle informazioni rilevanti rispetto a quei sistemi tradizionali che si affidano semplicemente al mezzo stampa.

Si passerà poi all'analisi di alcuni eventi del passato nei quali i *social network* hanno costituito delle piattaforme utili alla diffusione di informazioni da parte delle autorità, e hanno permesso alle persone colpite dai disastri di avere un sistema tramite il quale richiedere assistenza, ricevere istruzioni, porre lamentele o semplicemente rimanere informate sull'evoluzione del quadro d'emergenza.

Si proseguirà con la descrizione di alcuni sistemi attualmente adottati da enti e organizzazioni per estrarre informazione sociale e sfruttare i messaggi dei vari *social network* per costituire dati aggregati utili alla prevenzione e alla miglior risoluzione di eventi disastrosi.

Infine, ci si concentrerà sulla descrizione della piattaforma **Ushahidi**, il miglior esempio di paradigma *crowdsourcing* applicato a contesti d'emergenza, che aiuterà a meglio definire il concetto di segnalazione e come questo possa servire per la costruzione di *database* informativi a supporto delle decisioni in contesti di pericolosità e rilevanza civica.

3.1 - I social come nuova strategia di comunicazione: il caso NORAD/NORTHCOM

NORAD (*North American Aerospace Defense Command*) e NORTHCOM (*U.S. Northern Command*) rappresentano le principali agenzie militari americane per la difesa, e costituiscono un ottimo esempio di enti che si sono dotati di strategie intelligenti per lo sfruttamento dei *social media*.

Il nuovo piano di comunicazione, introdotto nel 2010, prevede di associare ai canali tradizionali (le classiche agenzie di stampa), alcuni *account* propri sui principali *social network* (*Twitter, Facebook, Flickr, YouTube*) unitamente a un *blog* gestibile direttamente dal comandante per le comunicazioni più importanti e tempestive.

Il personale complessivo dei due enti consta di circa 1200 lavoratori, tra militari e civili, appartenenti a diverse categorie e con differenti mansioni. Il nuovo meccanismo di diffusione delle informazioni prevede una modalità esponenziale che sfrutta le ricondivisioni. In una prima fase, il comando invia messaggi a una rete che comprende i 1200 dipendenti interni, ai quali è fatta richiesta di ricondividere tale messaggio con la propria rete privata di contatti, p.es. amici, familiari, ecc. È plausibile ritenere che a loro volta anche questi ultimi eseguano lo stesso tipo di operazione, con il risultato che l'informazione, partita da un'entità singola e certificata (il comando), tenderà a espandersi autonomamente secondo lo schema indicato in figura, e a raggiungere dunque un'alta percentuale di cittadini. [Lesperance et al., 2010]

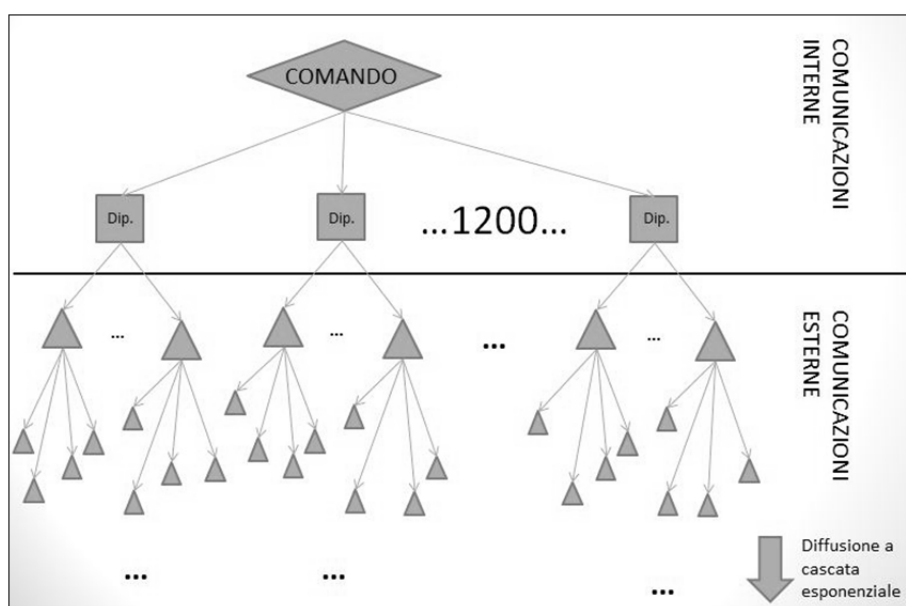


Figura 2 - Diffusione delle informazioni secondo la logica NORAD/NORTHCOM

Da un'analisi, è emerso che le informazioni con maggiore risonanza sono:

- Dati sui dipendenti interni
- Missioni ed eventi in corso (p.es. Terremoto di Haiti)
- Eventi connessi al 9/11
- Eventi di rilevanza internazionale (p.es. *Superbowl*)

È inoltre possibile pensare che le informazioni più importanti, nella cui categoria rientrano quelle relative a situazioni di emergenza, tendano spontaneamente a essere ricondivise e ad acquistare maggior interesse sociale, anche all'esterno della rete principale dei dipendenti.

Per evitare *overflow* informativo in caso di emergenza, i capi generalmente si incontrano con gli staff di comunicazione della *Federal Emergency Management Agency*, della Casa Bianca, e della USNORTHCOM, e indicano *conference call* con le organizzazioni operative. Durante la conversazione, ogni persona presente aggiorna i propri *social network* in tempo reale, inserendo però messaggi e stati preventivamente accordati con tutti i partecipanti della riunione, in modo che non sorgano conflitti né possano diffondersi informazioni false o poco accurate. La catena informativa, dunque, risulta separata da quella di comando: se da un lato i mezzi tradizionali consentono un accurato e costante controllo delle informazioni fornite, dall'altro risultano più lenti rispetto meccanismo di diffusione sociale ora descritto.

3.2 - Casi emblematici durante emergenze reali

3.2.1 - Uragano Gustav in Louisiana (Agosto 2008)

L'uragano Gustav fu il settimo ciclone tropicale e il secondo uragano maggiore della stagione degli uragani atlantici 2008. Si formò la mattina del 25 agosto 2008 a 420 chilometri da Port au Prince, Haiti; passò a tempesta tropicale già nel pomeriggio e divenne uragano il giorno successivo. L'1 Settembre 2008, Gustav raggiunse le coste della Louisiana e si abbatté sulla capitale New Orleans, già devastata nel 2005 dall'uragano Katrina.

Memore della tragedia avvenuta tre anni prima, il Governo impose l'evacuazione obbligatoria per evitare vittime. Per meglio coordinare gli spostamenti e permettere alle persone coinvolte di rimanere in contatto, il Dipartimento della Sicurezza Americano chie-

se aiuto al social network MySpace, che ai tempi godeva ancora di grande popolarità ed era considerato un leader nelle piattaforme sociali, complice il fatto che Facebook era ancora ritenuto un *network* destinato al solo ambito universitario. Su MySpace gli utenti avevano la possibilità di possedere un proprio profilo personale, che poteva essere arricchito attraverso la manipolazione di semplice codice html e l'aggiunta di applicazioni predefinite denominate *widget*.

Il Dipartimento della Sicurezza chiese pertanto a MySpace di sviluppare una *widget* che permettesse di:

- aiutare nella ricerca delle persone scomparse grazie a un collegamento tra i profili degli utenti e il *National Emergency Family Registry and Locator System*, un sistema di localizzazione che viene attivato in caso di disastri di grande portata;
- donare fondi;
- seguire l'avanzamento in tempo reale dell'uragano verso la costa.

Lo sviluppo richiese un paio d'ore e fu molto rapido perché Myspace forniva già per sua stessa natura i servizi *social* richiesti. L'utilizzo della *widget* aiutò nella gestione della fase di evacuazione e permise una riduzione del traffico telefonico d'emergenza sulle reti tradizionali.

I *social network* risultarono però utili anche per quegli Stati che erano stati chiamati ad accogliere gli sfollati della Louisiana. Un gruppo CERT (*Community Emergency Response Team*) dell'Arkansas, per esempio, vide fallire i metodi classici di ricerca di volontari. Il gruppo si affidava a *call list* e *email*, ma poiché l'emergenza scoppiò durante l'orario notturno, molti volontari non risultarono pronti a rispondere alla chiamata. Il CERT si affidò allora a messaggi di stato Facebook, visto che la gente tendeva in media a consultare spesso il *social network* e a diffondere le informazioni attraverso una ricondivisione più efficiente di quella dei canali tradizionali.

3.2.2 - Eruzione del vulcano islandese Eyjafjöll (Aprile 2010)

A partire dal 20 Marzo 2010, in conseguenza ad un'attività sismica continuativa che perdurava dall'anno precedente, si ebbero una serie di eventi di natura vulcanica sul vulcano Eyjafjöll, nella parte meridionale dell'Islanda. L'eruzione del 14 aprile 2010, a causa della densa nube di fumo che impediva la visibilità in una vasta area nei dintorni del vulcano, creò seri problemi alla navigazione aerea in Europa, paralizzando completamente il traffico aereo fino al 23 aprile e continuando a causare chiusure a intermitten-

za degli aeroporti del Nord Europa fino al 9 maggio. L'attività vulcanica cominciò poi a diminuire, finché nell'ottobre del 2010 non venne dichiarata conclusa, benché la zona continui ad essere a tutt'oggi geotermicamente attiva.

Lo stop del traffico aereo richiese l'implementazione di una strategia di pronta comunicazione nei confronti dei turisti e dei lavoratori che necessitavano di raggiungere l'Islanda, o che erano rimasti bloccati a terra e dovevano quindi contattare le strutture alberghiere locali. I siti Web delle compagnie aeree risultarono sovraccarichi di richieste, e il sistema tradizionale di gestione di queste ultime si rivelò farraginoso. Una dopo l'altra, le principali compagnie - KLM, Lufthansa, Airbaltic, British airways – si unirono alle conversazioni dei principali *social networks* per meglio comunicare con i propri clienti e fornire ad essi un servizio d'assistenza rapido e funzionale. Su Twitter si diffusero gli *hashtag* #ashcloud e #ashtag, di cui si ebbero oltre 55'000 *mention* in soli 7 giorni. Inoltre, gli *hashtag* #getmehome, #roadsharing, #stranded vennero usati dai locali per offrire passaggi, alloggi e pasti ai turisti bisognosi.

L'organizzazione per la sicurezza dei voli, EuroControl, utilizzò i propri *account* sociali per postare informazioni utili per i viaggiatori, usando persino i gruppi d'aviazione più rilevanti in *LinkedIn*, *network* normalmente destinato all'offerta e alla ricerca di lavoro.

La piattaforma Flickr, che giornalmente raccoglie migliaia di fotografie da parte dei suoi numerosi utenti e che costituisce un immenso database iconografico, custodisce diverse immagini scattate nei giorni dell'emergenza, con alcune riprese aeree della bocca del vulcano in attività.

Si sono registrati danni per 1.7 Miliardi di Dollari alle compagnie aeree nei primi 7 giorni dall'inizio dei problemi causati dall'eruzione. Tale cifra sarebbe potuta essere superiore in assenza di una strategia comunicativa adeguata, il che testimonia il ruolo dei *social networks* nell'efficiente diffusione di informazioni utili ai cittadini e nella gestione della comunicazione con essi.

3.2.3 - Allagamento a Queensland, Australia (Gennaio 2011)

Tra l'11 e il 15 Gennaio 2011 si registrarono oltre 56'000 *tweet* relativi all'allagamento nella città australiana di Queensland. Molti di questi *tweet* riportavano falsità e leggende metropolitane create ad arte: una di quelle più curiose raccontava di un coccodrillo che, sfruttando le strade allagate dal fiume in piena, si aggirava indisturbato per la città.

La gente ricorse all'utilizzo di due *hashtag* fondamentali: *#qldfloods*, per parlare dell'evento, e *#Mythbuster*, per indicare le informazioni false ed evitarne la propagazione.

L'auto-propagazione di *hashtag* antifrode indica come le persone siano naturalmente portate a identificare e correggere le informazioni false, senza l'intervento di autorità.

3.2.4 - Giappone: terremoto, tsunami, e disastro nucleare (Marzo 2011)

L'11 Marzo 2011 si verificò un sisma al largo della costa della regione di Tōhoku, nel Giappone settentrionale, alle ore 14:46 locali. Il sisma, di magnitudo 9 e con epicentro in mare alla profondità di 30 km, è attualmente catalogato come il più grande terremoto mai accaduto in Giappone e come il settimo a livello mondiale⁵. In seguito alla scossa si generò uno tsunami con onde alte oltre 10 metri che raggiunsero una velocità di circa 750 km/h. Le coste giapponesi più colpite dalle onde anomale furono quella della prefettura di Iwate, dove si registrò l'onda più alta di ben 40.5 metri, e quella della prefettura di Miyagi, che subì i maggiori danni. [BBC, 2011]

Il sisma provocò lo spegnimento automatico di undici centrali nucleari, grazie ad un sistema d'emergenza di *early warning*. Anche per la centrale nucleare Fukushima I il sistema d'emergenza spense i tre reattori funzionanti, ma poiché i meccanismi di raffreddamento risultarono danneggiati dalle scosse, la temperatura all'interno salì a livelli incontrollati e alle 15:40 (6:40 UTC) dell'11 marzo il reattore n.1 subì la fusione delle barre di combustibile, con un'esplosione visibile anche dall'esterno, che provocò il crollo di parte delle strutture della centrale. Lo stesso problema si sarebbe potuto verificare il giorno seguente con il reattore n.3, e il 14 Marzo con il reattore n.2, che però furono subito irrorati con acqua marina e acido borico, capace di assorbire neutroni e contenere la reazione. Il 15 Marzo il reattore n.4, seppur spento, esplose: il malfunzionamento agli impianti di raffreddamento portò a un surriscaldamento e alla vaporizzazione dell'acqua delle piscine all'interno delle quali esso era posto, e quindi alla successiva reazione tra il vapore bollente e lo zirconio che riveste le barre di combustibile.

Gli avvenimenti sono stati classificati dall'Agenzia per la sicurezza nucleare e industriale del Giappone al grado massimo della scala INES (*International Nuclear and radiological Event Scale*), a pari livello con il Disastro di Černobyl'. Si stima che a seguito della sola prima esplosione siano state rilasciate più radiazioni che nell'arco di un anno.

⁵ Dati dello United States Geological Survey
<http://earthquake.usgs.gov/earthquakes/centeqsww/Quakes/usc0001xgp.php>

A causa di questi avvenimenti, fu imposta l'evacuazione di oltre 110'000 persone nel raggio di 30 chilometri dall'impianto di Fukushima. A questo numero si deve aggiungere quello delle persone costrette a lasciare le proprie case in seguito al sisma e al conseguente maremoto.

Nei giorni del disastro, risultò praticamente impossibile comunicare attraverso le reti telefoniche tradizionali, velocemente saturate e dunque inaccessibili. La rete Internet, tuttavia, risultò funzionante e le persone si affidarono ai *social network* per comunicare. Secondo stime ufficiali diffuse da Twitter, nei cinque giorni successivi alla prima scossa, il traffico sul *social* risultò più che triplicato rispetto alla media, raggiungendo i 7.5 milioni di utenti attivi⁶.

Molte città colpite dal disastro aprirono un account per meglio diffondere le informazioni ai propri cittadini, e la rete venne sfruttata anche per la raccolta fondi e le iniziative benefiche.

Il *Web developer* Yusuke Wada creò *Anpi Report*, un database per organizzare le informazioni presenti su Twitter riguardo agli scomparsi. Sempre usando Twitter e Facebook, riuscì a reclutare 200 volontari che lo aiutarono nella gestione del database. Un'iniziativa simile e meglio strutturata fu successivamente presa anche da Google.

Numerose piattaforme di *videostreaming* sociale, come *Ustream* o *NicoNicoDouga*, vennero utilizzate per fornire una copertura *live* degli eventi, sfruttando reperti video forniti dai cittadini stessi.

3.2.5 - Terremoto dell'Emilia (Maggio 2012)

Durante la sequenza di terremoti che sconvolse l'Emilia Romagna nel Maggio 2012, si osservò il confluire su Twitter di milioni di messaggi con *hashtag* spontanei. Alcuni *blogger* si unirono per dare vita a reti civiche di soccorso, con lo scopo di diffondere le informazioni laddove vi fosse carenza informativa da parte degli organi ufficiali. Per esempio, il sito Internet www.terremotosanfelice.org fu ideato e coordinato da persone comuni, per poi divenire, dato il suo successo, il canale ufficiale di contatto fra istituzioni e cittadini. Il sito, a dispetto della propria natura informale, riuscì a raccogliere oltre 610'000 € di donazioni, quando CariParma, una delle banche principali della regione, si fermò a quota 370'000 €.

⁶ Fonte: *Nielsen NetRatings Japan*

L'utilizzo di canali informativi non ufficiali, tuttavia, apriva la strada alla possibilità che le notizie diffuse fossero troppo generiche, incomplete, o addirittura errate o fittizie. Per questo motivo il sindaco di San Felice sul Panaro (MO) richiese una certificazione istituzionale dei *post* al fine di garantirne la veridicità e sottolineare l'importanza del servizio. Tale certificazione fu realizzata attraverso una cooperazione tra gli autori del sito e il Comune stesso, a sua volta in contatto con gli enti regionali e statali, in modo che le notizie pubblicate fossero il più possibile corrette, e dunque utili alla popolazione.

Un altro esempio di sito utile alla causa fu quello di APT Servizi Emilia Romagna. Si tratta del portale regionale dedicato al turismo, uno dei settori che più rischiava di essere danneggiato dagli eventi. Attraverso un attento lavoro di comunicazione, il portale riuscì a informare efficacemente le persone in modo da contenere le perdite economiche derivanti dalla paura dei turisti di recarsi, nell'immediatezza o in futuro, nelle aree limitrofe a quelle a rischio per trascorrere le vacanze.

Facebook risultò utile invece per richiedere aiuti e supporti economici. Il Comune di Mirandola (MO) ottenne numerose risposte ai propri appelli, senza la necessità di passare attraverso lunghe pratiche burocratiche. Nacquero inoltre alcuni gruppi Facebook di aiuto, come p.es. "Un tetto per i cittadini della bassa" (creato in accordo con la protezione civile di Modena), il cui scopo fu raccogliere i riferimenti di persone o famiglie del modenese bisognose di un posto letto.

3.3 - Esempi strategici di utilizzo di informazione sociale

3.3.1 - 911ICE.org («In Case of Emergency»)

911ICE.org costituisce un esempio di implementazione di un *social network* "ex novo" per la gestione delle emergenze del singolo, piuttosto che della collettività, per quanto le sue funzionalità fossero solo parzialmente *social* e maggiormente orientate alla costituzione di gruppi sociali ridotti per il breve periodo di risoluzione dell'emergenza.

Il sistema fu ideato dal paramedico Bob Brotchie nel 2005, ma purtroppo fallì dopo la fase di *beta testing* per mancanza di un modello di *business* adeguato.

L'idea era equipaggiare tutti i telefonini con un contatto speciale chiamato 911ICE, attivabile dai primi soccorritori in caso di emergenza per il proprietario dell'apparecchio.

Una volta attivato, il sistema si sarebbe occupato di chiamare una rete di contatti selezionati per avvisarli dell'emergenza (amici, parenti, ma anche medici) e, grazie a una

pagina dedicata sul sito 911ICE.org, tali contatti avrebbero potuto collegarsi tra loro per gestire rapidamente la situazione e visualizzare le informazioni mediche del paziente o il suo stato di salute.

Il risultato sarebbe stato una risposta più tempestiva all'emergenza del singolo, una rapidità nell'avvisare le persone interessate e il personale medico, e una maggiore facilità nel garantire la comunicazione tra le parti coinvolte.

3.3.2 - American Red Cross Digital Operations Center

La Croce Rossa Americana si è dotata, nel Marzo 2012, del primo centro operativo digitale per la gestione di emergenze sulla base di un monitoraggio costante e intelligente dei *social network*. Il centro è stato sviluppato da DELL basandosi sul proprio sistema di *feedback* degli utenti (*The Social Media Listening Command Center*) e inserito in un ufficio del *Disaster Operations Center* a Washington D.C.

All'interno del centro, sei monitor forniscono costantemente dati aggregati ricavati dai principali *social network*, che possano avere una rilevanza nel monitoraggio delle situazioni d'emergenza e possano dunque aiutare a prevenirle o a meglio gestirle durante tutte le fasi del ciclo delle emergenze. Il controllo delle reti sociali consente inoltre ai cittadini di fornire attivamente informazioni alla Croce Rossa Americana, per meglio integrare i servizi di aiuto durante l'emergenza tramite una raccolta aggregata e razionale di dati.

Solitamente nel centro è sempre fisicamente presente almeno un operatore, ma i dati sono comunque accessibili attraverso una piattaforma via Web. Inoltre, esiste una rete di volontari digitali pronta ad avvisare nel caso di eventi particolarmente importanti o di tenere traccia della loro evoluzione quando chiamati ad agire personalmente. I volontari possono accedere a un *front-end* telematico attraverso il quale avvisare in caso di situazioni di pericolo o ricevere istruzioni precise su come muoversi, certificando la propria identità e la propria posizione. L'utilizzo di volontari sul campo, virtualmente disponibili anche fuori dall'orario di servizio ufficiale, fa sì che il sistema di segnalazione e comunicazione risulti scalabile e tempestivo.

I dati registrati e forniti dai monitor non costituiscono ovviamente, di per se stessi, motivo per agire. Essi rappresentano esclusivamente un supporto informativo alle decisioni umane prese dai cosiddetti decisori operativi. Per i disastri maggiori, i decisori valutano le informazioni fornite dal *Disaster Operations Center* e collaborano con esso per risol-

vere nel migliore dei modi l'emergenza, sia estraendo informazione utile dai dati, sia sfruttando le tecnologie del centro per inviare informazioni *push* a quei cittadini che fanno uso delle applicazioni per *smartphone* sviluppate dalla Croce Rossa o sono iscritti ai suoi account nei diversi *social network*.

Durante i disastri, il *Disaster Operations Center* invia *report* continui alle centrali operative ed è in grado di comunicare con i singoli volontari digitali attraverso specifiche *console*. I *report* sono anche condivisi con vari partner utili alla causa (p.es. associazioni umanitarie o servizi di assistenza alla collettività).

I dati aggregati sono forniti agli operatori attraverso differenti formati di visualizzazione:

- **heat map**, che illustra il volume delle conversazioni su una mappa, permettendo di individuare i punti caldi;
- **community**, che illustra i profili *social* delle persone che parlano di un certo *topic*, individuando anche gli utenti con più *follower*;
- **universe**, che illustra i volumi di conversazione attuali riguardo a specifici *topic*, permettendo di compararli tra loro e di evidenziare il sentimento delle persone a riguardo;
- **conversation dashboard**, che illustra i volumi di traffico lungo un arco temporale, evidenziando quando specifici *topic* smettono di generare traffico.

Vengono seguite conversazioni con specifiche *keyword* sui principali *social network*, ma anche su *forum*, *newsgroup*, *blog*, siti di notizie o di immagini/video. Questo consente di costruire dei *topic profile*, ovverosia dei profili di conversazione relativi a uno specifico avvenimento che, analisi dopo analisi, può essere collocato nel tempo e nello spazio e quindi consentire di agire sulla base di un'informazione mirata.

La *privacy policy* impone che possano venire aggregati solo ed esclusivamente post pubblici, in modo che non ci sia alcuna invasione nella vita privata degli utenti né alcuna questione di tipo legale.

3.3.3 - Emergency Situation Awareness – Automated Web Text Mining (ESA-AWTM)

ESA-AWTM è un sistema di *data mining* a supporto delle decisioni prese dai *watch officer* nei centri di coordinamento crisi presenti in ogni centro urbano in Australia, corrispondenti ai nostri uffici di protezione civile.

Il *data mining* risulta utile in uno scenario di *situation awareness*, ovvero di monitoraggio quotidiano della situazione locale al fine di identificare per tempo eventi di interesse per la collettività. Esso consente un'accurata analisi di dati alla ricerca di informazioni utili per la gestione e la prevenzione di una crisi. [Salerno et al, 2004]

Specificatamente, nel caso in oggetto, i dati analizzati vengono originati dal *social network* Twitter, attraverso la collezione, revisione, aggregazione e organizzazione di intensi volumi di *tweet*, alla ricerca di dati specifici riguardo ad eventi che possano impattare sulla sicurezza dei cittadini.

Il problema primario del sistema è la necessità di convertire il grandissimo flusso di messaggi in informazioni utili per la gestione dell'emergenza, in un tempo molto breve, scartando i dati non necessari o falsi. Velocità, praticità, ed efficienza sono dunque i requisiti primari degli algoritmi utilizzati.

Il sistema è stato sviluppato dal *CSIRO ICT Centre* sulla base delle specifiche richieste dei *watch officer*. Le funzionalità che il software deve implementare includono la ricerca di incidenti, la sintesi e l'indicizzazione dei messaggi, la classificazione e la revisione dei *tweet* di alto valore, l'identificazione e il tracciamento dei problemi, e l'analisi forense per utilizzi futuri. [Cameron et al., 2012]

Esse vengono fornite come segue:

- **ricerca di incidenti:** il sistema esamina le parole contenute nei *tweet* e usa dati storici per costruire un modello statistico di ripetizione di termini;
- **sintesi e indicizzazione dei messaggi:** tramite due motori esterni, *Carrot* e *Apache Solr*, vengono creati cluster di *tweet* che parlano di un medesimo argomento;
- **classificazione e revisione dei messaggi di alto valore:** è stato sviluppato un sistema statistico che permette di classificare i *tweet* in base all'interesse, istruito per identificare messaggi riguardanti "danni alle infrastrutture" grazie a vecchi *tweet* del terremoto 2011, e poi riadattato per incidenti di diversa natura;
- **identificazione, tracciamento, e gestione dei problemi:** un algoritmo si occupa di unire/dividere/cancellare *cluster* con contenuto simile lungo *burst* temporali;
- **analisi forense:** gli allarmi vengono memorizzati in un *database* storico, successivamente consultabile.

Per quanto riguarda il terzo punto, l'interesse intrinseco dei *tweet* è determinato da fattori quali la ricondivisione degli stessi o la presenza in essi di *keyword* dall'alto valore semantico.

I *tweet* di interesse sono ottenuti usando la *Twitter search API*, definendo una zona e un raggio di ricerca per coprire integralmente Australia e Nuova Zelanda.

In uscita, una mappa permette di posizionare i *tweet* in base all'origine, grazie a informazioni intrinseche di geolocalizzazione, o piuttosto a stime effettuate sui *tweet* stessi. Alcuni *marker* sulla mappa indicano le zone dei possibili incidenti, fornendo anche un raggio di interesse degli stessi. Inoltre, le *timeline* riportano il volume dei *tweet* catturati, suddivisi in finestre temporali di 5 minuti ciascuna. L'analisi dei volumi di traffico consente di comprendere l'evoluzione degli stessi, e la scelta di finestre temporali così brevi tiene conto della necessità di identificare eventi dirompenti o istantanei.

Tramite questo sistema, sono stati analizzati 600 milioni di *tweet* tra Giugno 2010 e Gennaio 2012, che hanno permesso di gestire almeno 12 importanti situazioni di emergenza in diverse città australiane. [Cameron, 2012]

3.4 - Una piattaforma di crowdsourcing: Ushahidi

Ushahidi, Inc. è una azienda *no-profit* con sede in Kenya che sviluppa software *open source* gratuito per la raccolta, la visualizzazione e la geolocalizzazione interattiva di informazioni. È stata fondata nel 2008 da Erik Hersman, Ory Okolloh, Juliana Rotich, David Kobia. Il motto adottato dai fondatori è “*Crowdsourcing Crisis Information*”: attraverso i propri progetti, l'azienda mira a sfruttare il paradigma del *crowdsourcing* rivolto all'attivismo sociale e alla responsabilità pubblica, servendo come modello iniziale per quello che è stato coniato come “*activist mapping*”, ovvero la combinazione di attivismo sociale, giornalismo cittadino e informazione geospaziale. Ushahidi offre pertanto degli strumenti che permettono ad osservatori locali – cittadini comuni – di inviare informazioni utili per la collettività usando i propri telefoni cellulari o la rete Internet, allo scopo di creare archivi temporali e geospaziali di eventi di interesse pubblico.

Ushahidi iniziò la propria missione ancor prima della fondazione dell'azienda, aprendo il sito Web <http://legacy.ushahidi.com/> nel periodo immediatamente successivo alle contestate elezioni presidenziali del 2007 in Kenya. Il sito raccolse le testimonianze oculari (“ushahidi” in lingua swahili significa proprio “testimonianza”) riguardo gli episodi di violenza scoppiati in seguito alle contestazioni, inviate dai cittadini via email o sms e geolocalizzate su mappe estratte da *Google Maps*. [Baree, 2008]

Nonostante ai tempi l'interattività con il sito fosse pressoché nulla, la raccolta di dati fu preziosa e costituisce ancora oggi un importante database storico riguardo a quei tragici avvenimenti.

Sulla base di quell'idea venne dunque costituita l'organizzazione, che iniziò a sviluppare una piattaforma più flessibile per la segnalazione di eventi, indipendentemente dal tipo di evento o dalla sua localizzazione.

Il concetto oggi alla base del sito www.usahidi.com è quello di rendere possibile agli utenti la generazione di mappe di pubblico dominio, con possibilità di segnalazione da parte di chiunque, allo scopo di raccogliere dati puntuali in riferimento a uno specifico evento, scelto da colui il quale genera la mappa. Tali eventi possono avere una marcata rilevanza sociale e inserirsi all'interno di contesti di emergenza, arrivando a fornire un quadro spaziotemporale dell'evoluzione degli stessi, grazie alle testimonianze di chi si trova a doverli affrontare. I dati raccolti possono quindi essere utilizzati sia dai cittadini, per meglio identificare le zone di pericolosità di un dato territorio, sia dalle istituzioni preposte alla risoluzione dell'emergenza, come supporto alle decisioni critiche.

Le mappe possono essere installate su domini esterni tramite l'utilizzo di un codice *php open source* scaricabile dal sito ufficiale, oppure possono essere generate direttamente su di esso attraverso un *tool* molto semplice, che fornirà poi un indirizzo Web al quale raggiungere la mappa così creata.

Nel 2011, a Ushahidi si è affiancata SwiftRiver, una suite di prodotti accessibili attraverso specifiche API, che va a supporto della raccolta e dell'analisi in tempo reale di grandi volumi di dati. SwiftRiver ha tre funzioni primarie: strutturare dati non strutturati, classificare e assegnare priorità alle informazioni in tempo reale, e infine contestualizzarle in termini spaziotemporali.

3.4.1 - Utilizzo di Ushahidi durante il terremoto di Haiti (Gennaio 2010)

In seguito al gravissimo terremoto di Haiti del 2010, Patrick Meier unì in uno sforzo comune le organizzazioni Ushahidi, *The Fletcher School of Law & Diplomacy* e UNOCHA (*United Nations Office for the Coordination of Humanitarian Affairs*), avviando il sito <http://haiti.usahidi.com/>. In pochissimo tempo, altre organizzazioni si unirono alla causa, e numerosi volontari presero parte all'iniziativa che fu denominata *Ushahidi Haiti Project (UHP)*.

UHP permise di raccogliere e diffondere informazioni critiche sui luoghi colpiti, aiutando sia le organizzazioni umanitarie in loco sia quelle esterne a meglio indirizzare gli aiuti. Inoltre facilitò gli spostamenti delle persone presenti nel territorio colpito, anche grazie alla generazione di mappe *OpenStreet* sulle quali furono tracciati i percorsi sicuri e i luoghi praticabili.

In accordo con il report [Morrow et al.,2011], è possibile a posteriori verificare come la piattaforma sia stata concretamente usata sia dal Dipartimento Statale Americano degli Analisti, sia dall'esercito, per conoscere la situazione locale e per meglio organizzare le operazioni. Il motivo per cui Ushahidi ricoprì un ruolo tanto fondamentale nasce dal fatto che fu il più completo, se non l'unico, aggregatore di informazioni sulle aree colpite, in un territorio abbastanza carente di supporti informatici efficienti.

Diverse interviste successive a responsabili ufficiali hanno permesso di definire il reale apporto di Ushahidi al salvataggio di vite umane, per quanto non sia stato possibile effettuare una stima numerica sul dato.

Si è invece stimato che ci siano state tra le 40'000 e le 60'000 segnalazioni e che siano stati mappati 3'584 eventi ad Haiti, di cui l'80% nel primo mese successivo all'evento sismico.

Le difficoltà maggiori incontrate nell'utilizzo delle informazioni fornite dai volontari furono innanzitutto di natura tecnologica. Considerato il territorio, le tecnologie locali risultarono obsolete, con una connessione alla rete Internet lenta o in alcune zone addirittura assente; inoltre numerose persone colpite erano del tutto prive di un supporto fisico tramite il quale accedervi. Questo rese complicato effettuare segnalazioni per le mappe di crisi.

Ma oltre a questa limitazione, ve ne furono altre più formali. In primo luogo la struttura dei dati aggregati raccolta da Ushahidi non sempre corrispondeva a quella rigidamente richiesta dalle autorità per poter prendere decisioni concrete. Poiché il progetto era ancora agli albori e la sua implementazione fu molto rapida, non ci fu il tempo sufficiente per organizzare un *output* del sistema efficiente in accordo con i Governi; per questo motivo, molte delle informazioni fornite da UHP non poterono essere sfruttate per tempo, a causa di un'organizzazione di informazioni non rispondente a certi requisiti.

Anche una generale carenza di conoscenza riguardo al *crowdsourcing* determinò una riduzione di efficienza delle informazioni fornite. È stato provato come UHP risultò molto utile alle squadre dotate di supporti informatici come la cosiddetta *Mission 4636*,

3.4 - Una piattaforma di crowdsourcing: Ushahidi

mentre fu del tutto ignorato da altri gruppi di potenziali utilizzatori che avevano meno familiarità con Ushahidi. Alcune interviste rivelarono un “generale sospetto nei confronti della massa”, e critiche varie sull’utilizzo di informazioni non certificate.

4 - Il progetto TORCIA

Tutto il lavoro presentato nella seguente tesi è stato svolto presso il Dipartimento di Elettronica e Informazione del Politecnico di Milano nell'ambito della ricerca per un progetto denominato **TORCIA**.

Il progetto TORCIA – *PiaTtafORma di gestione CollaboratIvA delle emergenze* – ha l'obiettivo di utilizzare le tecnologie del Web 2.0 disponibili per la gestione delle emergenze nei processi di pianificazione e controllo delle situazioni critiche – incidenti o calamità naturali – garantendo la resilienza delle infrastrutture di trasporto.

Il progetto si inserisce all'interno di un bando promosso dalla regione Lombardia, finalizzato alla presentazione di lavori di ricerca industriale e sviluppo sperimentale nei settori strategici della Regione e del ministero dell'istruzione, dell'università e della ricerca.

Attraverso la cooperazione di diversi partner, il team di sviluppo mira a definire un sistema integrato che sia in grado di utilizzare l'informazione sociale proveniente dai cittadini, al fine di organizzare e interpretare tempestivamente i dati raccolti tramite *crowdsourcing* e di fornirli a supporto di enti e organizzazioni che si occupano di gestione delle emergenze. Inoltre, la piattaforma che si intende realizzare è di tipo resiliente, garantendo così la piena operatività in contesti critici.

TORCIA si concentra su eventi di emergenza che avvengono nelle aree metropolitane. Durante la fase di sviluppo, i casi di studio saranno estrapolati dall'area milanese, con la collaborazione attiva dei cittadini. L'informazione da analizzare sarà estratta da fonti di varia natura; la novità del progetto, da un punto di vista tecnologico, è l'utilizzo dei *social network* in modo completo e integrato, sia per interpretare le informazioni in fase di emergenza, sia per fornirle alle autorità. Per far fronte a queste richieste si è pensato a

una infrastruttura di elaborazione basata su piattaforme *cloud* geograficamente distribuite e collegate tra loro da una rete ottica ad alta capacità.

TORCIA si propone di facilitare le operazioni di raccolta dei dati sul territorio per consentire un'immediata comprensione delle situazioni critiche e delle iniziative da compiere. Gli utenti del Web 2.0 diventano protagonisti nel fornire, tramite i *social network*, informazioni georeferenziate e dati utili per una migliore gestione delle emergenze. Il progetto intende infatti sfruttare al meglio queste tecnologie realizzando una piattaforma per raccogliere, razionalizzare e interpretare le informazioni che i cittadini rendono disponibili sul Web, in modo da condividerle con le autorità competenti alla sicurezza.

4.1 - I partner del progetto

Alcatel-Lucent

Alcatel-Lucent è *leader* di progetto, e mette a disposizione di TORCIA la rete ottica di *backbone* che collega i diversi *server*. Fornisce inoltre un sistema di messaggistica innovativo, che può essere utilizzato per informare istantaneamente i cittadini e le autorità del pericolo imminente o in corso.

ACT Solutions

ACT Solutions è una società specializzata in sistemi avanzati di supporto alle decisioni, basati su competenze di previsione, simulazione, ottimizzazione e controllo di processo e costruiti sull'esperienza e sull'ampiezza delle competenze ingegneristiche del team. L'azienda ACT Solutions, esperta nella ricerca operativa e *business intelligence*, lavora agli algoritmi per la *detection* dell'emergenza sulla base del *monitoring* dei dati raccolti.

Beta 80 Group

Grazie alle soluzioni proprietarie di *Emergency Management*, Beta 80 Group è *leader* in Italia nella realizzazione delle Centrali Operative negli ambiti di Sanità, Protezione Civile, Vigili del Fuoco, Assistenza Domiciliare e Grandi Strutture.

Beta 80 fornisce il cruscotto di monitoraggio del progetto.

Vidiemme Services

Oltre a supportare i clienti nello sviluppo di applicazioni software complesse, Vidiemme Services è in grado di offrire soluzioni nell'ambito dello sviluppo di applicazioni che integrano le tecnologie mobili con il Gis (Cartografia con contenuti georeferenziate e

realtà aumentata). Con lo sviluppo del progetto TORCIA l'azienda intende accrescere in modo significativo le proprie competenze nel settore del Gis e del *Mobile Solutions*, che rappresentano uno degli aspetti importanti nello sviluppo del progetto.

L'azienda Vidiemme esegue il *monitoring* del traffico della zona osservata e fornisce informazioni sulle vie preferenziali per le evacuazioni e i soccorsi.

Dipartimento di Elettronica e Informazione del Politecnico di Milano

Il Dipartimento di Elettronica e Informazione del Politecnico di Milano, comunemente chiamato DEI, è tra i maggiori dipartimenti universitari in Europa per dimensioni e qualità della ricerca. La sua missione è quella di produrre un costante miglioramento della conoscenza in campo scientifico e tecnologico nei settori dell'informatica, delle telecomunicazioni, dell'automatica e dell'elettronica.

I ricercatori del Politecnico di Milano forniscono il motore semantico Web 2.0, realizzando una struttura Web che può essere facilmente letta dal sistema del progetto.

Fondazione Politecnico di Milano

La Fondazione Politecnico di Milano, nata nel 2003, è stata voluta dal Politecnico, da importanti aziende e da enti locali per favorire e rendere sistematiche le collaborazioni fra i dipartimenti dell'università, le imprese e la pubblica amministrazione, per valorizzare i risultati della ricerca e per contribuire alla creazione di realtà imprenditoriali innovative.

La Fondazione cura le attività di diffusione e affianca Alcatel-Lucent nel coordinamento manageriale del progetto.

4.2 - Le attività del progetto

Le attività di TORCIA si concentrano sullo sviluppo di un sistema in grado di monitorare, interpretare e comunicare le informazioni inviate dagli utenti sui *social network*.

- **Monitorare.** Il monitoraggio avverrà in tempo reale tramite la combinazione di analisi tradizionali (ad esempio informazioni inviate da navigatori, isoradio) e destrutturati, raccolti in tempo reale tramite *crawling* georeferenziato sui *social network*, garantendo la presenza di un cruscotto cartografico e operativo.

- **Interpretare.** Interpretazione immediata di quanto sta accadendo con indicazioni dell'impatto che questo avrà con l'infrastruttura di trasporto locale e l'indicazione delle azioni da intraprendere, sia a livello di istituzioni sia a livello di cittadinanza.
- **Comunicare.** Fornire informazioni, diffondendo dati in ottica multicanale verso la cittadinanza per poter rendere immediate le iniziative da intraprendere.

Si vuole creare una soluzione unica e integrata che sia in grado di fornire uno strumento di supporto per gli addetti ai lavori (Protezione Civile, Polizia Locale...). Perché questa sia efficace tuttavia deve essere utilizzabile e funzionante anche in condizioni di maxi-emergenza ed eventi catastrofici. Per questo motivo la struttura del sistema deve essere resiliente. Pertanto nell'ambito del progetto si studieranno anche metodi innovativi per rendere resilienti delle infrastrutture preesistenti utilizzando il *cloud computing* distribuito.

4.3 - L'architettura di TORCIA

L'architettura di riferimento del sistema è indicata in figura.

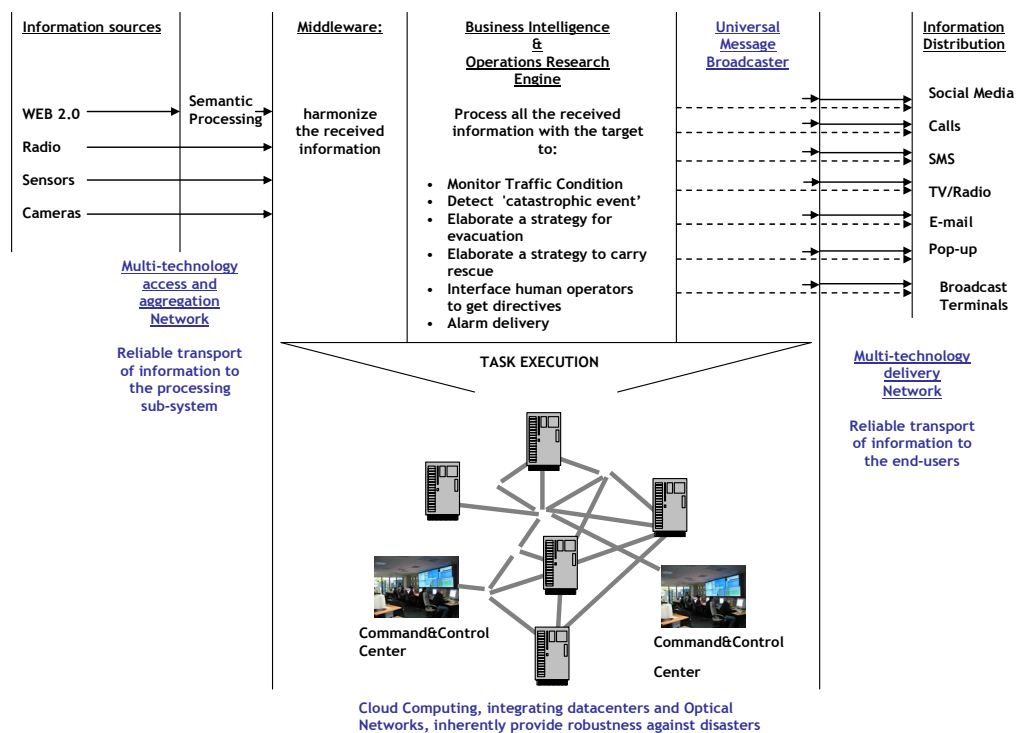


Figura 3 - L'architettura di TORCIA

4.3.1 - Input del sistema

Il sistema TORCIA può sostanzialmente attingere informazione da diversi tipi di sorgente, che trasportano due tipologie di messaggi:

- **Messaggi indiretti:** si intende il flusso informativo non diretto alla centrale operativa, ma estratto da altre fonti che normalmente mutuano informazioni generiche e non strettamente correlate alle emergenze. Questi messaggi richiedono un'analisi approfondita al fine di estrarre da essi informazione strutturata, che dovrà poi essere validata dall'operatore e certificata dalle autorità competenti.
- **Messaggi diretti:** si intende tutto il flusso informativo indirizzato esplicitamente alla centrale operativa, tramite differenti tipi di canale, da una tipologia variegata di fonti. Queste ultime possono essere suddivise in due categorie: istituzionali (Comuni, enti locali ecc.) e persone private (singoli cittadini). Le informazioni derivanti da fonti istituzionali sarebbero automaticamente validate dal sistema, mentre le altre richiederebbero una validazione manuale da parte dell'operatore di centrale.

Tra i messaggi indiretti troviamo quelli estratti dai *social network Web-based*, che sono oggetto della presente tesi, e saranno dunque meglio dettagliati in seguito.

Tra i messaggi diretti, invece, troviamo le chiamate agli operatori di emergenza, gli sms, le email, e in generale messaggi inviabili tramite applicazioni mobili apposite.

Ove possibile, i messaggi diretti dovrebbero contribuire alla geolocalizzazione della segnalazione. Ad esempio nel caso della telefonia mobile, l'informazione sulla localizzazione potrebbe essere ricavabile dalla geolocalizzazione tramite GPS/AGPS o, nel caso in cui l'ente utilizzatore del progetto TORCIA sia di tipo istituzionale (ad esempio protezione civile) e previo accordo con i *carrier* nazionali, anche tramite triangolazione delle celle. Per le altre fonti invece la localizzazione sarebbe affidata esclusivamente al contenuto informativo del messaggio.

Un ulteriore sorgente può essere costituita da telecamere e sensori già in uso alle istituzioni preposte al controllo del territorio. Effettuare un monitoraggio anche su queste fonti ed estrarre da esse eventuali dati può ampliare l'efficienza del sistema, soprattutto per quando riguarda la pronta localizzazione di eventi critici.

4.3.2 - Il core del sistema

L'analisi e la comprensione del contenuto informativo trasportato dai messaggi in *input* al sistema TORCIA rappresentano le operazioni basilari che costituiscono il *core* dell'architettura. È riguardo a queste operazioni che si muove la ricerca del Politecnico di Milano, nell'ambito della quale è stato sviluppato questo lavoro di tesi.

In particolare, il Politecnico si occuperà di realizzare un motore di analisi, strutturato per rispondere alle seguenti esigenze:

1. **Analisi semantica** – L'analisi permetterà la categorizzazione delle conversazioni sui *social media* in base a un'ontologia progettata specificamente per la gestione di emergenze e basata su 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.
2. **Business intelligence semantica** – La *business intelligence* permetterà di individuare tutte le informazioni utili alla riduzione del rischio di future emergenze. Sarà basata su tecniche di *data mining* semantico, che consentono l'individuazione di eventi (p.es. picchi conversazionali su una specifica criticità) legati a un'inadeguatezza (non voluta e non prevedibile) delle procedure di gestione dell'emergenza.
3. **Crowdsourcing e analisi dei piani di emergenza** – Si proporranno soluzioni che integrano architetture a servizi utili alla comunicazione tra sistemi informativi eterogenei (p. es. VVFF, Croce Rossa, Enti Locali), e GIS per la visualizzazione e l'aggiornamento collaborativo (*crowdsourcing*) dei piani di emergenza.

Questa tesi, in particolare, si concentra sul punto 1, provando a definire una metodologia per la costruzione dell'ontologia di dominio.

4.3.3 - Il motore di calcolo

ACT Consulting si occuperà di sviluppare un motore di calcolo per la definizione dei percorsi e delle vie di fuga in scenari di crisi.

A fronte dell'identificazione di una segnalazione di emergenza il motore di ottimizzazione sarà in grado di calcolare :

- vie di accesso alla zona dell'emergenza;
- vie di fuga dalla zona dell'emergenza;
- percorsi alternativi alla zona dell'emergenza.

Il calcolo di queste tre tipologie di soluzioni viene effettuato tenendo in considerazione:

- la mappa stradale;
- le informazioni raccolte dai vari canali di *input* alla piattaforma e i relativi livelli di affidabilità;
- la situazione del traffico;
- le informazioni aggiuntive sul territorio (p.es. siti a rischio rilevante).

L'obiettivo che si vuole raggiungere è l'implementazione di uno strumento che consenta di valutare con rapidità la situazione del territorio, al fine di comprendere come raggiungere o lasciare i luoghi colpiti dalla crisi agendo su una o più delle seguenti funzioni obiettivo:

- Minimizzazione dei tempi di percorrenza (tempo di viaggio)
- Minimizzazione delle percorrenze (distanza totale viaggio)
- Minimizzazione della pericolosità percorso (punti "pericolosi" lungo il percorso)

Per ogni percorso valutato, si terrà ovviamente conto dei mezzi di trasporto a disposizione, in modo da valutare correttamente i tempi di percorrenza o le necessità proprie del mezzo per il transito in quella zona.

4.3.4 - L'*output* del sistema: il supporto alle decisioni

L'*output* ricavato da TORCIA potrà essere utilizzato dai diversi enti preposti alla salvaguardia del cittadino come supporto alle decisioni in condizioni critiche. Le informazioni estratte dalle sorgenti e successivamente aggregate costituiranno un aiuto per la comprensione maggiormente dettagliata dello scenario d'emergenza, allo scopo di valutare dove è necessario agire e come farlo al meglio. L'analisi continuativa del parlato sociale potrà aiutare nella gestione dell'emergenza in ognuna delle sue quattro fasi, potendo fornire dati utili per qualsiasi situazione, sia a emergenza in corso che a emergenza finita.

Grazie all'integrazione con la cartografia, il sistema sarà in grado di geolocalizzare le richieste d'intervento e quindi di contestualizzarle all'interno del territorio metropolitano. Il cruscotto consentirà, attraverso un'interfaccia accessibile agli operatori, di visualizzare le segnalazioni sulla mappa, con la possibilità di leggere le informazioni associate alla segnalazione, come il tipo, la gravità, la descrizione, la provenienza, la data e l'ora, ecc. Inoltre sarà possibile anche visualizzare eventuali immagini/fotografie associate alla segnalazione (ad esempio inviate tramite cellulare).

Lo sviluppo di apposite applicazioni per *smartphone* consentirà di avere un supporto per l'invio di messaggi diretti, e al tempo stesso offrirà alle autorità la possibilità di raggiungere la popolazione con apposite segnalazioni utili alla loro salvaguardia. Il sistema provvederà infatti a supportare l'invio di messaggi in *broadcast* attraverso diverse piattaforme, dai *social network* alle email, dagli sms agli allarmi su apparecchiature specifiche in dotazione agli abitanti delle zone più critiche.

Le funzionalità di *output* del sistema TORCIA faranno leva sull'UMB (*Universal Message Broadcaster*), che è in grado di:

- diffondere ogni tipo di messaggio usando numerosi mezzi di comunicazione;
- gestire tutte le applicazioni che richiedono la diffusione di un messaggio;
- essere facilmente usato in qualunque area geografica, poiché utilizza infrastrutture (reti telefoniche, Internet) disponibili ovunque.

L'UMB, che è basato su piattaforma proprietaria Alcatel-Lucent, viene sviluppato con un approccio che consente in ogni fase di aggiungere le funzionalità richieste da un particolare progetto/applicazione. In questo modo l'UMB cresce continuamente in termini di funzionalità e interfacce disponibili.

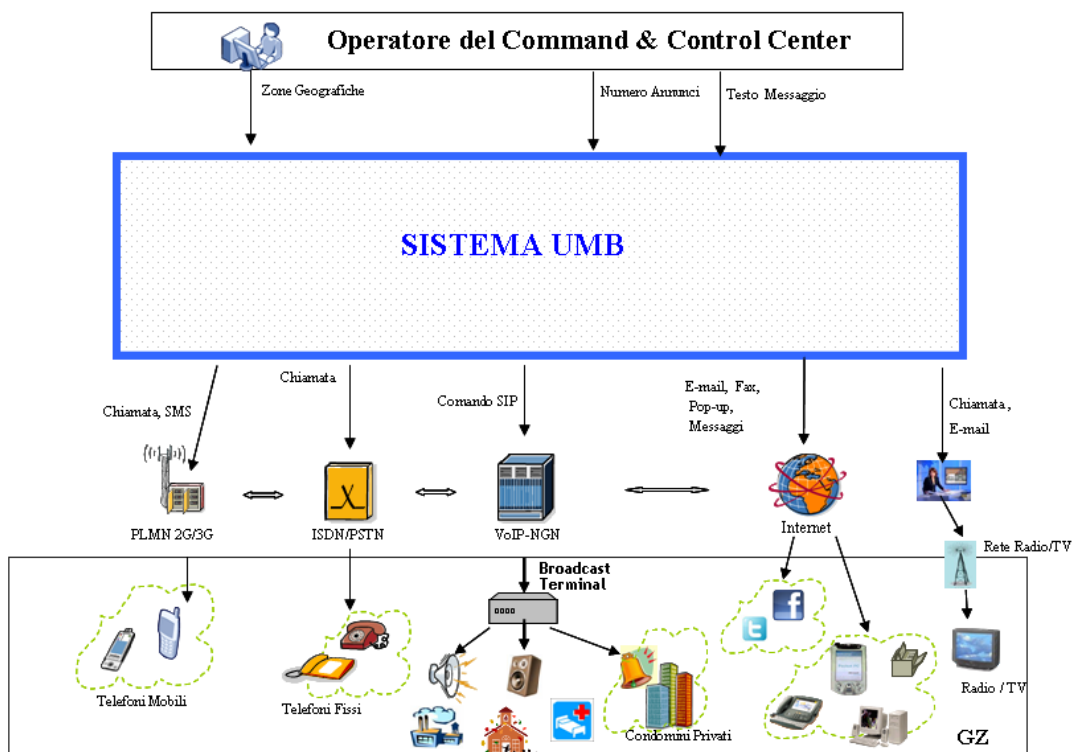


Figura 6 - Sistema UMB

In particolare in TORCIA si prevede di utilizzare le seguenti reti di comunicazione:

- Le reti di telefonia PSTN e VoIP/NGN (per chiamate a telefoni fissi)
- Le reti di telefonia mobile (per sms e chiamate a telefoni mobili)
- Internet (per email, fax, pop-up su PC e su Smartphone, interfaccia con i *social media*, attivazione dei *Broadcast Terminal*)
- Reti radio e TV (per diffusione delle informazioni tramite annunciatori radio/TV)

L'operatore avrà la possibilità di selezionare i mezzi di comunicazione attraverso i quali il messaggio sarà diffuso.

Il *Broadcast Terminal* (BT) è un dispositivo innovativo che può essere facilmente installato (tramite un normale accesso Internet) negli uffici, scuole, ospedali, case, ecc.

Come rappresentato nella precedente figura il BT, comandato dall'UMB, può diffondere nella sua area annunci registrati (tramite altoparlante interno o diffusori esterni), visualizzare alcune righe di testo, e pilotare campane, sirene, lampeggiatori e attuatori disponibili localmente.

4.4 - I destinatari dell'output TORCIA

I destinatari dei messaggi di *output* del sistema TORCIA verranno raggruppati in categorie predefinite. Le categorie sono caratterizzate dal ruolo delle persone che vi appartengono, e determinano il tipo di messaggio e il relativo contenuto che gli appartenenti alle categorie devono ricevere.

Le categorie previste attualmente sono:

- autorità (sindaco, presidente della Provincia, prefetto, ecc.);
- squadre di emergenza;
- cittadini;
- persone diversamente abili.

Altre categorie potranno essere aggiunte nel corso del progetto, in funzione delle necessità. Per esempio la categoria "squadre di emergenza" potrà comprendere Vigili del Fuoco, Pronto Soccorso, Protezione Civile, i cui operatori registrati all'UMB riceveranno messaggi con istruzioni operative, mentre la categoria "Cittadini" riceverà messaggi solo informativi relativi all'evento in atto.

5 - L'analisi dell'informazione sociale in un contesto d'emergenza specifico

L'obiettivo primario di questa tesi è la progettazione di una metodologia che consenta di effettuare un'analisi sulle informazioni fornite dagli utenti attraverso vari canali di comunicazione, primariamente quelli di tipo sociale, allo scopo di utilizzare questa informazione a supporto delle decisioni in contesti di emergenza.

Il primo passo per giungere a questo risultato si realizza attraverso un'analisi contestuale che permetta innanzitutto di definire che tipo di informazione utilizzare e da quali sorgenti di *input* attingerla. Esistono diversi aspetti collegati a questa questione. Alcuni sono di tipo tecnologico, e passano attraverso una rassegna delle possibili fonti che possono essere realmente sfruttate, perché dotate di interfacce specifiche che sono state progettate allo scopo. Gli stessi *social network*, già indicati nel paragrafo 1.3, offrono funzionalità assai differenti e veicolano intrinsecamente diversi tipi di informazione, più o meno sfruttabile e più o meno utile. Inoltre, alcune di queste piattaforme possiedono impostazioni di *privacy* molto restrittive che non consentono un accesso pubblico ai messaggi inseriti dagli utenti, rendendo di fatto inutile se non addirittura illegale un eventuale lavoro di estrazione da queste origini. Ma al di là degli aspetti tecnologici, anche le considerazioni puramente ontologiche sull'informazione da utilizzare risultano di estrema rilevanza. Non tutti i dati che vengono inseriti dagli utenti possiedono un significato intrinseco utile alla nostra metodologia. Nell'attività di raccolta è fisiologico imbattersi in una grande quantità di informazione che a conti fatti risulta fuorviante. Un attento controllo preventivo sulle fonti, allo scopo di indicare quelle che normalmente veicolano un'informazione più attinente al contesto di studio, può servire a risparmiare

un successivo lavoro di scarto da parte di eventuali algoritmi che si occuperanno dell'aggregazione automatica dei dati.

La nostra metodologia deve pertanto necessariamente partire da un'analisi rigorosa e "manuale" su insiemi di dati generici, ovvero con un controllo umano sul tipo di informazione normalmente veicolata dalle fonti in esame. Quest'analisi può, attraverso una strategia di tipo *bottom-up*, aiutare anche nella creazione di un'ontologia specifica utile per istruire successivi algoritmi di elaborazione che sfruttino il paradigma del *crowdsourcing*. L'ontologia specifica permette essa stessa una categorizzazione dell'informazione. La categorizzazione è importante in quanto genera una miglior strutturazione dei dati acquisiti, e aiuta in una risposta più tempestiva da parte di chi sfrutterà questi dati. L'*output* del sistema deve infatti fornire un'informazione aggregata e strutturata, nella quale siano evidenziati tutti i punti critici e utili alla causa, che invece risultano ambigui e nascosti nell'informazione in ingresso.

La categorizzazione che proponiamo viene definita sulla base di quelle che sembrano essere le informazioni di primaria importanza nella fase di risposta e risoluzione di un'emergenza. Tali necessità passano attraverso una rapida comprensione della situazione di pericolo in termini di caratteristiche, localizzazione spazio-temporale, effetti prodotti. Talvolta lo stesso *input* può veicolare un'informazione che ricade in categorie differenti. Occorre pertanto definire una strategia intelligente per evitare ridondanza nell'*output* senza produrre uno scarto eccessivo di dati.

Una volta eseguita l'analisi manuale e definita l'ontologia di dominio, le conoscenze acquisite sono state utilizzate all'interno di un algoritmo sintattico progettato da Claudio Carcaci nell'ambito del suo lavoro di tesi, svolto presso il dipartimento di Elettronica e Informazione del Politecnico di Milano. I test condotti su questo algoritmo non solo possono servire al miglioramento dello stesso, nell'ottica della progettazione futura di un robusto motore di analisi sintattica, ma consentono anche di rielaborare le considerazioni iniziali in modo da ridefinire un dominio più complesso. L'algoritmo sintattico può infatti operare associando nuovi termini alle classi e, in parallelo con una nuova analisi manuale, servire a migliorare le strategie di sfruttamento dell'informazione precedentemente definite. Uno degli obiettivi è dunque elaborare un metodo logico per realizzare un *feedback* nell'algoritmo proposto da Carcaci, lavorando sui risultati dei test condotti con esso.

È importante osservare che la metodologia proposta si concentra esclusivamente su un ambito specifico e si basa sull'utilizzo di alcuni strumenti primari, che però non sono i soli possibili. Il nostro lavoro è essenzialmente correlato alle emergenze, inserendosi all'interno del progetto TORCIA, ma nulla vieta di riadattare la stessa metodologia ad altri contesti di elaborazione dell'informazione sociale. Il paradigma proposto, così come l'algoritmo utilizzato, può infatti trovare spazio in lavori differenti, realizzando di fatto un esempio concreto e argomentato di *crowdsourcing*, riadattabile all'occorrenza secondo le proprie esigenze.

5.1 - I passi della metodologia

Riepilogando le considerazioni di inizio capitolo, vogliamo definire in modo rigoroso i passi condotti dalla nostra metodologia, per poi sviscerarli uno a uno all'interno dei paragrafi successivi.

Le attività che andremo a descrivere sono state eseguite seguendo questo schema ordinato:

- 1) analisi dello stato dell'arte e di casi reali;
- 2) scelta della sorgente;
- 3) scelta di un contesto specifico di analisi;
- 4) acquisizione di dati dalla sorgente in un periodo che includa:
 - i. situazioni senza crisi;
 - ii. situazioni di crisi;
- 5) analisi manuale dei dati;
- 6) costruzione delle categorie e dominio lessicale;
- 7) utilizzo dei dati all'interno dell'algoritmo sintattico;
- 8) elaborazione di un feedback di miglioramento dei dati e delle categorie;
- 9) analisi conclusiva per raffinamento dell'algoritmo.

Il passo numero 1 si è basato sullo studio delle informazioni già descritte nei primi capitoli. L'analisi di casi reali ha meglio consentito di approcciare a questo lavoro, con una certa consapevolezza riguardo all'orientamento della gente nell'utilizzo dell'uno o dell'altro strumento tecnologico in caso di emergenza.

5.2 - Scelta della sorgente

Come predetto, la scelta della sorgente per la nostra metodologia nasce sulla base di considerazioni tecnologiche, pratiche, di *privacy* e di effettiva realizzabilità di strumenti utili alla cattura di messaggi veicolati su detta sorgente.

Abbiamo osservato come l'architettura di TORCIA permetta, nelle proprie linee guida, di interfacciarsi con diversi tipi di sorgente, che abbiamo suddiviso in sorgenti *dirette* e sorgenti *indirette*. (cfr. par. 4.3.1)

Le sorgenti dirette possono essere evidentemente fornite solo da partner che già si occupano della gestione delle emergenze, i quali possiedono piattaforme di monitoraggio e raccolta dati con cui ricavare messaggi specifici, in qualche modo preventivamente strutturati e filtrati. Infatti, il sistema stesso di raccolta è fornito agli utenti per la segnalazione dei soli eventi critici, e consente quindi una partecipazione attiva dei cittadini con l'invio di informazione mirata, già riferibile al contesto.

Queste sorgenti, chiaramente, non possono essere utilizzate in questa tesi, in quanto richiederebbero una collaborazione con gli enti che si occupano della risoluzione delle emergenze. Una possibile estensione del nostro lavoro potrebbe essere fatta generando opportunamente messaggi diretti fittizi, ma la genuinità dei risultati ne risentirebbe, a causa del fatto che verrebbero meno le caratteristiche di generalità dell'informazione che desideriamo utilizzare per implementare e istruire il motore di analisi. Uno dei nostri scopi è infatti testare questo motore, e verificare che sia in grado di scartare l'informazione non necessaria, ovvero di ampliare il dominio ontologico preventivamente costruito con l'analisi umana.

Ci concentriamo pertanto sulle sole sorgenti di tipo indiretto, e cioè su quelle che veicolano messaggi di natura generale, non già riferibili al contesto. Tra queste, selezioniamo quelle provenienti da *social network* preesistenti.

Un tentativo di generazione di un *social network ex novo* per la gestione delle emergenze è stato fatto da Iacopo Pace in un'altra tesi al Politecnico di Milano. Basandosi su Ushaihidi, Pace ha creato il network **Crosafe**, coinvolgendo amici e colleghi in una fase di *testing* che ha consentito la raccolta di diversi messaggi di segnalazione per eventi di criticità minore sul territorio milanese. Un'attenta analisi sui messaggi ha però rivelato una scarsa partecipazione degli utenti quando chiamati a utilizzare una rete sociale creata *ad hoc* per la gestione di emergenze. Per questo motivo, risulta più utile ai fini della

nostra trattazione ricercare la sorgente in una rete già ampiamente utilizzata dai cittadini, che sono quindi più invogliati a veicolare attraverso di essa eventuali messaggi che si riferiscano a un'emergenza, qualora questa si verificasse.

L'utilizzo di una sorgente non specifica, ovviamente, solleva il problema di avere a che fare con una qualità informativa abbastanza imprevedibile, e richiede quindi strumenti più completi per la costruzione di un motore d'analisi efficiente. I messaggi indiretti, in generale, possono infatti contenere informazione utile in varia misura, o non contenerne affatto.

Ancor più in dettaglio, in caso di emergenza, l'informazione fornita da questi messaggi, data la loro natura generale, potrebbe essere di due tipi: *estrinseca* o *intrinseca*. L'informazione estrinseca è un'informazione chiara e riferita al contesto, che contiene dati precisi e puntuali e dunque ha la sola necessità di essere registrata. Messaggi indiretti con informazione estrinseca potrebbero in taluni casi essere assimilabili a messaggi diretti, poiché contengono segnalazioni precise, riferibili a un determinato evento critico. L'informazione intrinseca, di contro, necessita di essere estratta e interpretata, grazie ad alcune deduzioni che si possono fare sulla sintassi del messaggio. Per esempio, un messaggio potrebbe non contenere dati specifici sul luogo in cui si verifica un dato evento, ma fornire una serie di indizi che aiuterebbero a desumerlo. Com'è intuibile, però, non sempre l'informazione intrinseca può essere dedotta e registrata, soprattutto da algoritmi informatici.

L'informazione poi può essere *completa* o *incompleta*. Un'informazione completa permette la registrazione di più dati specificatamente riferibili a un evento, e consente dunque una risposta più efficace da coloro i quali se ne serviranno. L'informazione incompleta di un messaggio, tuttavia, è comunque interessante, soprattutto se integrata e confrontata con quella proveniente dalle altre segnalazioni. Un numero di segnalazioni elevato, quindi, aumenta la probabilità di ottenere un'informazione valida e utile, riducendo di contro la possibilità che l'informazione risulti falsa.

La sorgente che ricerchiamo deve possedere, dunque, le seguenti caratteristiche:

- (i) deve essere utilizzata da un grande numero di persone che possiedono un *know how* specifico sul mezzo che stanno utilizzando;
- (ii) deve possedere ridotte o nulle impostazioni di *privacy* per consentire l'acquisizione del maggior numero possibile di messaggi;

- (iii) deve possedere un'interfaccia attraverso la quale poter eseguire l'acquisizione;
- (iv) deve consentire l'inserimento e la lettura di messaggi istantanei, in altri termini non deve essere moderata *a priori*;
- (v) deve poter veicolare un'informazione concisa piuttosto che prolissa e/o vaga.

Ulteriori caratteristiche utili nella scelta fanno riferimento alle tipologie di messaggi inseribili sul *social network* in esame. Poiché alcuni dei dati fondamentali relativi a un'emergenza si riferiscono al luogo e al momento in cui essa si verifica, possiamo dire che le seguenti sono caratteristiche desiderabili per la nostra sorgente, per quanto non fondamentali:

- (vi) deve permettere il *geotagging*, ovvero la geolocalizzazione del messaggio per avere informazione estrinseca sulla sua origine;
- (vii) deve indicare in maniera inequivocabile la data e l'ora in cui un messaggio viene inserito;
- (viii) deve fornire un meccanismo di ricondivisione.

Sulla base di queste e delle successive considerazioni, e in relazione all'elenco dei *social network* redatto nel par. 1.3.2, la nostra scelta è ricaduta su Twitter.

Twitter è un *social network* asimmetrico testuale basato su aggiornamenti di status di 140 caratteri [Bennato, 2010]. Esso soddisfa pienamente il punto (i), grazie a una media giornaliera di 3,64 milioni di accessi, valutata nel Settembre 2012.⁷ [Rossitto, 2012]

Date le sue caratteristiche primarie (messaggi trasparenti, brevi, ricondivisibili, marcati temporalmente e organizzabili secondo *keyword*), inoltre, riesce a soddisfare anche tutti gli altri punti. L'unica eccezione è il punto (vi), dato che il *geotagging* non è una funzionalità obbligatoria per tutti i messaggi, ma comunque opzionale.

Twitter rappresenta un ottimo strumento di diffusione delle informazioni, grazie a una struttura basata su *follower* che stimola alla condivisione e genera il fenomeno sociale della *influence*. La teoria tradizionale sulla comunicazione afferma infatti che una minoranza di utenti, denominati *influencers*, eccellono nella persuasione di tutti gli altri [Rogers, 1962]. È stato dimostrato che questo meccanismo di persuasione non dipende es-

⁷ Dati Audiweb/Nielsen

senzialmente dal numero di *follower* di un dato utente. La diffusione delle informazioni avviene infatti attraverso meccanismi di *retweet* (ricondivisione di un *tweet*) e *mention* (risposta/citazione a un *tweet*), ma i volumi dell'uno e dell'altro non dipendono dal grado di popolarità dell'utente che ha inserito il *tweet* originale: i *retweet* sono infatti stimolati dal valore intrinseco del messaggio, mentre le *mention* sono stimolate dalla stima percepita verso il suo autore. Questo porta alla definizione di una teoria fondamentale, denominata “*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. [Cha et al., 2010]

Da un'osservazione riguardo all'*influence* dei *top user*, si è compreso come le associazioni giornalistiche inneschino un altissimo numero di *retweet*, a causa del fatto che pubblicano messaggi dall'elevato valore semantico. Altri *top user*, come le celebrità ad esempio, stimolano molte *mention*, grazie alla stima che i diversi *fan* hanno di loro.

[Kwak et al., 2010] hanno indagato come Twitter possa essere considerato un sistema per una “lenta ma costante diffusione di notizie” [Lerman et al., 2010]. Le caratteristiche giornalistiche del network rimarcano la sua efficacia in un contesto di emergenza.

[Bennato et al., 2010] hanno eseguito una ricerca riguardo le dinamiche della circolazione delle notizie in Twitter. Essi hanno identificato l'esistenza di due modelli chiave:

- le notizie **esogene** – ovvero che i *tweeters* ricevono da altri media e poi condividono su Twitter – crescono in maniera velocissima per poi stabilizzarsi con un basso contributo dei *retweet* alla diffusione;
- le notizie **endogene** – che nascono dentro Twitter e poi in alcuni casi trascinano su altri media – hanno una curva di crescita relativamente lenta, ma in cui il ruolo dei *power user* è basso mentre molto più importante è il ruolo degli utenti “normali” e dei *retweet* dei loro contatti.

In termini di emergenza, possiamo considerare i messaggi di nostro interesse tanto di un tipo quanto dell'altro. È tuttavia ragionevole ritenere che i *media* giochino un ruolo fondamentale nell'inserimento di notizie in caso di emergenza, producendo quindi una crescita estremamente rapida di messaggi descrittivi di essa, i quali tenderanno a diffondersi rapidamente non tanto con un meccanismo di *retweet*, ma con uno spontaneo passaparola.

5.3 – Scelta di un contesto specifico di analisi

Nel paragrafo 2.6 abbiamo definito una serie di possibili casi di emergenza che si possono verificare in ambito cittadino. Vogliamo ora isolarne uno, adducendo le dovute motivazioni, al fine di procedere con la nostra metodologia.

L'evento che andiamo ricercando deve innanzitutto permettere un'analisi su un numero sufficientemente elevato di *tweet*: deve perciò produrre un certo volume di traffico, il che significa che deve potersi verificare con elevata probabilità e/o frequenza.

Poiché desideriamo concentrarci su più fasi di un'emergenza, ricerchiamo anche un evento che stimoli informazione riferibile a ognuna di queste fasi. In altre parole, il nostro evento deve tanto produrre messaggi di segnalazione, quanto messaggi di allarme pre-incidente e messaggi relativi alle conseguenze dello stesso, in modo da coprire interamente il ciclo delle emergenze.

Meglio sarebbe poi selezionare un evento che avviene tendenzialmente su piccola scala, che affligge cioè un contesto limitato nello spazio e nel tempo, facendo perciò in modo che le diverse fasi del ciclo delle emergenze si susseguano con rapidità e consentano di monitorare le eventuali attività risolutive e di ricostruzione, oltre che quelle di risposta immediata. Le conseguenze previste dall'evento dovrebbero essere di molteplice natura e coinvolgere diversi enti, nonché portare le persone a uno spontaneo “chiacchiericcio” su Twitter riguardo sia le caratteristiche dell'emergenza in corso, sia la località, sia i danni immediati che essi sono in grado di riferire.

A seguito di queste considerazioni, la scelta per questo lavoro di tesi è ricaduta su **alluvioni e allagamenti**. Senza riportare le caratteristiche peculiari dei due fenomeni, già ampiamente discusse nel par. 2.5.1, ci limitiamo a indicare e riassumere le motivazioni fondamentali che hanno spinto a questa scelta:

- sono fenomeni che si possono verificare in ambito metropolitano, e rimanerne entro i confini;
- sono fenomeni frequenti sull'intero territorio italiano, paragonati ad altri fenomeni naturali;
- sono argomento di discussione frequente sui *social media*;
- giornali e organi di informazione se ne occupano ampiamente, anche in previsione di un'emergenza;

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

Poiché gran parte di questo lavoro nasce all'interno del progetto TORCIA, finanziato da Regione Lombardia, un altro motivo di scelta va ricercato nell'osservazione di alcune delle tematiche calde in fatto di emergenza nel territorio lombardo, e, in vista di un'estensione del progetto, in quello italiano. Alcuni dei danni più rilevanti degli ultimi anni sono stati generati proprio a partire da alluvioni e allagamenti, che hanno prodotto l'esondazione dei diversi corsi d'acqua presenti sul territorio.

La Lombardia per esempio visse l'ultima delle più grandi alluvioni del secolo scorso, che avvenne tra il 12 e il 14 settembre 1995. Furono duramente colpiti alcuni centri del milanese e del varesotto, tra cui Gallarate e Legnano. La tragedia fu causata dall'esondazione dei torrenti Rile, Tenore, Arnetta e dal fiume Olona, che si ingrossarono a causa delle abbondanti precipitazioni (376 mm in 30 ore), sebbene la causa principale sia stata la cementificazione selvaggia lungo le sponde dei corsi d'acqua. L'alluvione provocò danni quantificabili in circa 500 miliardi di lire (circa 258 milioni di euro).

A titolo esemplificativo, il bilancio di quell'evento fu il seguente:

- centinaia di sfollati;
- 94 aziende colpite per complessivi 86 miliardi di lire di danni;
- altre centinaia di miliardi di danni alle abitazioni e alle opere pubbliche;
- migliaia di operai in cassa integrazione;
- 18 strade provinciali bloccate o fuori uso;
- 72 comuni alluvionati, tra le province di Varese, Milano e Como;
- diversi paesi rimasti senz'acqua potabile a causa della contaminazione dei pozzi.

L'Italia è il paese Europeo con il più alto numero di morti e di danni causati dalle alluvioni dal 1980 al 2002 [Asproni, 2007]. Tra il 2000 e il 2012, si sono contate in Italia 23 alluvioni. La più grave fu quella in Piemonte dal 13 al 16 Ottobre 2000, in cui si contarono 23 morti, 11 dispersi, e più di 40'000 sfollati.

Tutti questi numeri supportano perciò la nostra scelta, evidenziando nelle alluvioni, e nei fenomeni ad esse connessi, un ottimo evento su cui lavorare con la nostra metodologia.

È comunque opportuno osservare che i passi seguenti possono essere ripetuti anche effettuando una scelta differente: basterà modificare opportunamente il dominio di riferimento, lasciando inalterate le modalità di esecuzione dei passi.

5.4 - Acquisizione di dati dalla sorgente

L'acquisizione di dati da Twitter può effettuarsi attraverso una ricerca mirata di *tweet* contenenti specifiche *keyword*, nel pieno rispetto della struttura e delle funzionalità proprie di tale *social network*. In generale, una *keyword* o un insieme di *keyword* definiscono in Twitter un *topic* di interesse, ovvero restituiscono un gruppo di discussioni da parte di diversi utenti focalizzate su un dato argomento. Nella nostra trattazione, potremmo definire *topic* “le alluvioni e gli allagamenti”, e ricercare pertanto le *keyword* che si riferiscono a tale argomento, in modo da costruire un insieme di messaggi che ne parlano.

La ricerca può essere effettuata manualmente, usando la barra di ricerca fornita dal sito, oppure automaticamente grazie alla comunicazione di software terzi con la *Twitter Search API*, che è un'API dedicata per l'esecuzione di ricerche di *tweet* recenti sulla base di *query* di interesse. Un software che esegue questa operazione è denominato **crawler**, e l'attività di ricerca dei contenuti in modo metodico e automatizzato, con il successivo salvataggio di una copia testuale dei dati trovati, viene denominata **crawling**.

Non esiste un intervallo temporale predefinito di ricerca, in quanto Twitter permette la restituzione dei *tweet* sulla base della loro rilevanza, ovvero tende a escludere quelli meno “chiacchierati” e a mantenere quelli che si riferiscono ai cosiddetti *trending topic*.

Per questa tesi è stato utilizzato un *crawler* sviluppato dal Dipartimento di Elettronica e Informazione del Politecnico di Milano, i cui dettagli tecnici sono riservati e comunque del tutto irrilevanti ai fini del proseguimento del nostro lavoro. Ci limitiamo tuttavia a dire che questo software produce in *output* dei file in formato .csv (file con valori separati da virgola) contenenti due valori, il primo che riporta il *timestamp* temporale del *tweet* (ovvero data e ora di inserimento dello stesso), e il secondo che riporta il contenuto del *tweet* racchiuso tra virgolette. La struttura del singolo campo del .csv è dunque:

publication_date,"text"

con publication_date espresso come

AAAA-MM-GG hh:mm:ss

dove AAAA rappresenta l'anno, MM il mese, GG il giorno, hh l'ora in formato 24h, mm il minuto, ss il secondo.

Tale riferimento temporale è registrato dal server centrale di Twitter e risulta quindi un valore certificato e non ambiguo.

Un esempio è il seguente:

2012-11-20 11:23:41, "Questo è un tweet di esempio catturato dal crawler"

5.4.1 - Scelta delle *keyword* di ricerca: la definizione di un dominio lessicale

Focalizzandosi sul *topic* selezionato, la scelta delle *keyword* da ricercare tramite il *crawler* riveste una grande importanza, poiché andrà a definire lo specifico *dominio lessicale* su cui opera la nostra metodologia. Il *lessico* o *vocabolario* è il complesso delle parole e delle locuzioni di una lingua, oppure anche solo una parte di tale complesso. La scelta di un numero finito di termini afferenti al solo argomento di cui ci stiamo occupando, perciò, definisce in maniera univoca il lessico di base della nostra metodologia.

La scelta del dominio non è banale, poiché deve raggiungere un *trade-off* tra due opposte richieste: non essere troppo ridotto, tanto da escludere *tweet* importanti, né troppo ampio, tanto da includere troppi messaggi "spazzatura".

Le *keyword* principali possono essere ricavate a partire dal titolo del *topic*, "alluvioni" e "allagamenti", e procedendo con l'aggiunta di sinonimi e/o termini similari estratti dal medesimo contesto. Un problema da affrontare è quello relativo al genere e al numero dei termini ricavati: occorre infatti considerare tutte le possibili combinazioni di singolare/plurale e maschile/femminile, qualora esistenti. Oltre ai sostantivi, risulta utile ricercare anche i verbi in diverso tempo/modo, ma per restringere il campo è meglio con-

siderare solo participi passati (assimilabili ad aggettivi) e gerundi, molto ripetuti in forme verbali composte.

Riassumendo, possiamo quindi costruire il seguente insieme di termini con i quali effettuare attività di *crawling*:

Allagamenti, Allagamento, Allagando, Allagare, Allagata, Allagato, Alluvionata, Alluvionati, Alluvione, Esondare, Esondato, Esondazione, Frana, Franare, Franata, Franato, Frane, Grandinando, Grandinare, Grandinata, Grandinate, Grandinato, Grandine, Inondando, Inondare, Inondata, Inondato, Inondazione, Inondazioni, Nubifragi, Nubifragio, Tempesta, Uragano.

L’inserimento di termini come “Frana”, “Franare”, “Franata”, “Franato”, “Frane” risulta interessante per il fatto che tali eventi possono nascere in conseguenza ad alluvioni, o esserne l’origine.

Si osservi che le *keyword* di ricerca, da sole, non costituiscono ancora il nostro dominio lessicale completo. Il nostro scopo è infatti quello di usarle come punto di partenza per restringere la raccolta dei messaggi al solo ambito delle alluvioni. Successivamente ci occuperemo di estendere il nostro vocabolario per generare un dominio completo e strutturato dell’informazione.

5.4.2 - Il crawling

L’attività di *crawling* in Dipartimento di Elettronica e Informazione del Politecnico di Milano è iniziata il 24 Settembre 2012, senza mai interrompersi.

Nel periodo intercorso tra l’1 e il 27 Ottobre 2012, sono state osservate le seguenti ricorrenze di *keyword* del dominio:

Tabella 2 - *Keyword* risultanti dal *crawling* nel mese di Ottobre 2012

KEYWORD	Totale 1-27 Ottobre 2012
ALLAGAMENTO	328
ALLAGAMENTI	582
GRANDINE	453
GRANDINATA	62
GRANDINATE	15
TEMPESTA	5698
ALLUVIONE	2890

INONDAZIONE	190
INONDAZIONI	237
ESONDAZIONE	136
FRANA	2157
FRANE	264
URAGANO	1804
NUBIFRAGI	1552
NUBIFRAGIO	1166
INONDARE	96
INONDATA	284
INONDATA	231
INONDANDO	77
ALLAGARE	306
ALLAGATO	681
ALLAGATA	1427
ALLAGANDO	219
ESONDARE	18
ESONDATO	31
ESONDANDO	0
FRANARE	21
FRANATO	17
FRANATA	21
FRANANDO	0
ALLUVIONATO	0
ALLUVIONATE	0
ALLUVIONATI	92
ALLUVIONATA	28
GRANDINARE	49
GRANDINATO	14
GRANDINANDO	0

Non sono presenti in dipartimento strumenti automatici di monitoraggio costante dell'attività intorno alla singola *keyword*: non è perciò possibile fornire in questo lavoro dei grafici corretti che dimostrino l'andamento relativo ai *tweet* contenenti la singola *keyword* in uno specifico intervallo di tempo ridotto. L'implementazione di uno strumento di questo tipo consentirebbe un'analisi più approfondita per determinare volumi di traffico precisi e meglio contestualizzati. La sua mancanza non consente – a causa del meccanismo di Twitter che non rende possibile l'analisi a posteriori di vecchi *tweet* – di rimediare. Si consiglia perciò di servirsene in previsione di estensioni future di questa metodologia.

Un'osservazione importante si è però potuta fare anche senza l'ausilio di strumenti informatici. In data 13 Ottobre 2011 è scattata infatti un'allerta alluvione a Roma, e in data 15 Ottobre è stata perciò eseguita un'analisi sui messaggi raccolti dal *crawler*. Questi dati sono stati poi utilizzati per l'esecuzione dei passi successivi della metodologia. Ciò ha permesso di soddisfare le richieste del punto 4) indicato nel par. 5.1, ovvero l'utilizzo e l'analisi di *tweet* sia in situazioni senza crisi, sia in situazioni di crisi.

In generale, osservazioni qualitative hanno indicato un aumento nell'utilizzo delle *keyword* di interesse, in associazione con lo specifico evento in corso a Roma. Ciò non deve stupire ed è in perfetto accordo con le osservazioni riguardo alla diffusione dell'informazione fatte nel par. 5.2.

5.4.3 - La prima analisi manuale dei dati

Come già osservato, l'analisi manuale dei dati forniti dal *crawler* costituisce un passo necessario per poter meglio approcciare al problema di definizione del dominio.

Quest'analisi ha lo scopo di definire elementi qualitativi e quantitativi sull'informazione diffusa attraverso Twitter in uno specifico intervallo di tempo e relativamente a un dato *topic* di interesse. Le conclusioni a cui si vuole giungere portano alla comprensione più specifica e dettagliata di *che cosa* le persone parlano sul social network, di *quando* lo fanno, e di *come* lo fanno, ovviamente in riferimento alla nostra emergenza di interesse, ovvero le alluvioni. Questi elementi – che cosa, quando, dove – possono aiutare a meglio contestualizzare l'informazione raccolta, in modo da strutturarla secondo categorie predefinite associabili a una specifica sintassi e a uno specifico vocabolario di termini ricorrenti. L'analisi consente inoltre di restringere il campo di ricerca, evidenziando *keyword* inutili o dannose, che rischiano di produrre una serie di risultati fuorvianti per l'identificazione dell'emergenza.

L'analisi manuale è stata effettuata sulla base dei primi elementi raccolti dal *crawler* in data 27 Settembre 2012. Essi si riferiscono a una serie di *tweet* raccolti tra il 24 Settembre 2012, data in cui il *crawler* è entrato in funzione, e il 27 Settembre 2012, data in cui sono stati prodotti i primi file .csv.

Riportiamo ora alcune delle deduzioni ricavate da questa prima analisi.

Le voci più chiacchierate

I termini più diffusi sono risultati essere: allagamenti (60 occorrenze), allagata (90), allagato (42), frana (783), frane (56), grandine (77), inondata (42), inondato (23), inondazioni (34), nubifragio (53), tempesta (370), uragano (73).

In particolare, il termine “frana” ha ottenuto un numero così elevato di *tweet* a causa della frana lungo la “Via dell’Amore”, una strada a picco sul mare presso le Cinque Terre, in Liguria, avvenuta proprio il 24 Settembre 2012. Anche se non connesso a un’alluvione, l’evento è comunque interessante perché permette di dimostrare come una notizia apparsa su diversi quotidiani abbia ricevuto grande rilevanza su Twitter, grazie a ben 294 *retweet*.

Gli altri termini, non presenti in elenco, hanno prodotto meno di 15 messaggi nell’intervallo considerato. Le motivazioni vanno ricercate sia in una minore rilevanza sintattica del termine (che probabilmente risulta meno utilizzato nella lingua italiana), sia nel fatto che alcuni degli eventi richiamati dal termine stesso non si sono verificati in quei giorni.

I termini poco pertinenti o fuorvianti

Alcune *keyword* hanno generato un buon traffico, ma nell’analisi sono risultate poco pertinenti, ovverosia hanno prodotto un’informazione non rilevante per la definizione dello scenario di emergenza.

Il termine “allagando”, per esempio, è risultato molto fuorviante perché ha generato un elevato numero di messaggi che si riferiscono ad allagamenti casalinghi. Lo stesso per il termine “allagata”, che a causa della desinenza femminile si è spesso accompagnato alla parola “casa”, generando ben il 30% di *tweet* di scarto.

Altre *keyword*, come “grandinata” (13), sono risultate pertinenti all’evento meteorologico ma poco utili per quanto riguarda l’evento alluvionale.

Il termine “inondata” (42) nella quasi totalità dei messaggi si riferiva a una città “inondata di manifesti abusivi”.

I termini associabili e/o ripetuti

Molti messaggi possedevano più *keyword*, oppure contenevano una *keyword* strettamente associata a un’altra parola utile nel contesto delle emergenze.

Per esempio, in diversi messaggi “inondazione” (4) e “inondazioni” (34) erano spesso legate a “morti”, “danni”, “evacuazioni”, anche perché a volte si riferivano ad eventi passati, fornendo dunque informazione sulle conseguenze di un’emergenza.

Parole come “maltempo” o “pioggia”, essendo eventi meteorologici e concause, sono state evidenziate molte volte nei messaggi analizzati.

Le locuzioni, i detti popolari, le citazioni

Alcuni dei termini di ricerca nel linguaggio comune possono riferirsi a precise locuzioni o richiamare detti popolari, libri, canzoni. Questo produce messaggi che contengono tali termini in un contesto che non è di nostro interesse.

Per esempio la parola “frana” (783 occorrenze), ha generato 48 messaggi (6%) che contenevano la locuzione “sono una frana”, mentre “tempesta” (370) si è riferito a volte a “mare in tempesta”(19) o alla poesia di Giacomo Leopardi dal titolo “la quiete dopo la tempesta” (14).

Gli eventi passati

Alcune *keyword*, più di altre, sembrano generare un’informazione riferibile al passato. Il termine “alluvionati”, per esempio, ha generato messaggi relativi a eventi alluvionali dei mesi o degli anni scorsi. Questi messaggi sono comunque utili in quanto si riferiscono a una diversa fase del ciclo delle emergenze, ma richiedono lo studio di una regola sintattica che consenta di categorizzarli con giudizio.

5.4 - Costruzione delle categorie

Come indicato nel paragrafo 5.2, i messaggi estratti da una specifica sorgente, soprattutto di tipo indiretto, possono contenere al proprio interno un’informazione tanto estrinseca quanto intrinseca. L’informazione fornita non riveste sempre, però, lo stesso interesse pratico. Riferendoci al ciclo delle emergenze, infatti, osserviamo come nelle sue diverse fasi si tenda a necessitare di dati differenti. Per esempio, nella fase di *preparedness* risulta utile eseguire *crowdsourcing* per raccogliere segnalazioni e allarmi da parte dei cittadini al fine di evitare il verificarsi di incidenti evitabili, mentre nella fase di risposta sembra essere quanto mai necessaria la conoscenza della zona nella quale intervenire. Allo stesso modo, nella fase di *recovery* occorre conoscere i bisogni primari della gente

per agire laddove vi sia più necessità, e nella fase di *mitigation* può servire fare il punto sulle responsabilità di modo che l'evento di crisi non si verifichi nuovamente in futuro.

Per questi motivi, appare quanto mai evidente la necessità di strutturare l'informazione raccolta, categorizzandola opportunamente sulla base dei dettagli che essa veicola. La categorizzazione dei messaggi è importante perché consente di meglio definire il contenuto informativo, e di meglio strutturare l'*output* dell'algoritmo di analisi, fornendo così, a chi ne necessita, indicazioni più dettagliate a supporto delle decisioni. Inoltre, lo studio delle categorie permette di ampliare il dominio lessicale dell'evento, andando a estrarre dai messaggi i termini di uso più corrente, utili per lo smistamento nell'una o nell'altra categoria.

La nostra proposta di categorizzazione si basa perciò su considerazioni di tipo lessicale e sintattico, dovendo poi servire come base per l'esecuzione di un algoritmo di questa natura. Costruiremo quindi ogni categoria evidenziandone un certo dominio lessicale, ovvero definendo un'associazione tra la categoria e le sue *keyword* specifiche, ricavabili dai *tweet* e dall'analisi supportata dalla letteratura di settore.

Definiamo quindi sette categorie principali, fornendo una breve descrizione dei messaggi che ricadono in ognuna di esse:

Tabella 3 - Categorie per i messaggi

NOME CATEGORIA	DESCRIZIONE DEI MESSAGGI	FASE DI RIFERIMENTO
Segnalazione	Messaggi generici che indicano un evento critico	<i>Response</i>
Conseguenze safety/security	Messaggi che contengono informazioni relative ai danni a strutture e/o persone e/o meccanismi sociali	<i>Response e Restoration</i>
Viabilità	Messaggi relativi alla viabilità, al traffico, all'utilizzo di specifiche vie	<i>Response, Restoration e Mitigation</i>
Luogo	Messaggi che contengono un'indicazione sul luogo in cui si verifica/si è verificato un evento e consentono la geolocalizzazione	<i>Response e Restoration</i>
Responsabilità	Messaggi che contengono l'indicazione delle responsabilità di un dato evento e/o identificano in specifici soggetti o enti le figure preposte alla risoluzione.	<i>Response, Restoration e Mitigation</i>
Warning pre-evento	Messaggi che contengono allarmi e segnalazioni utili a evitare che si verifichi un'emergenza.	<i>Preparedness</i>
Post evento	Messaggi che si riferiscono a situazioni che avvengono dopo che l'emergenza è terminata, p.es. lamentele dei cittadini, richieste di risarcimento, informazioni sulla ricostruzione, domande alle autorità.	<i>Restoration e Mitigation</i>

Dobbiamo ora costruire l'associazione tra messaggi, *keyword* e categorie. Vogliamo in altri termini ricavare delle *keyword* specifiche che permettano di inserire i messaggi che le contengono all'interno della categoria prescelta.

Un primo problema da affrontare è relativo alla **dimensione** degli insiemi di *keyword*. Il dominio, infatti, non deve essere troppo ristretto da trascurare messaggi importanti, ma nemmeno troppo ampio da includere spazzatura.

Un secondo problema è invece causato dalla **non-univocità** della funzione che associa *keyword* e categorie: esistono infatti termini che possono richiamare più categorie.

Infine, il genere e il numero dei sostantivi può generare anche in questo caso un aumento nella complessità del dominio: è necessario tenere presente, per una data parola, tutte le sue possibili **desinenze**.

Ciò predetto, e in riferimento all'evento di nostro interesse, seguono ora le *keyword* selezionate per ciascuna categoria. Esse non sono le uniche possibili, anche in previsione dell'esecuzione dell'algorithmo sintattico, con il quale si vuole verificare l'ampliamento automatico del dominio.

Sono inclusi nel successivo elenco anche una serie di esempi tratti dai *tweet* realmente registrati durante il *crawling*, oltre ad alcune osservazioni specifiche su ciascuna categoria elencata.

SEGNALAZIONE

- **Set termini:** Allagamenti, Allagamento, Allagata, Allagato, Temporale, Temporali, Pioggia, Piogge, Alluvione, Alluvionati, Esondazione, Esondato, Fiume, Piena, Acqua, Fango, Tombini, Tombino, Nubifragio, Nubifragi, Tempesta
- **Esempi:**
 - «Temporale fortissimo a Bergamo. Se avete segnalazioni di allagamenti o problemi scriveteci. Dalle vostre parti tutto bene?»
 - «Forti piogge hanno provocato l'esonazione di diversi corsi d'acqua nel sud del Texas e negli stati della Louisiana e del Mississippi.»
- In questa categoria rientrano moltissimi messaggi, dato il carattere di generalità dei termini e il fatto che si riferiscano al dominio di base. Potremmo considerare praticamente tutti i *tweet* relativi al *topic* come *tweet* di segnalazione.

CONSEGUENZE SAFETY/SECURITY

- **Set termini:** 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, Tendo-poli, Casa, Case, Frana, Franato, Franata, Crollata, Crollate, Crolli, Inghiottito, Inghiottite, Inghiottiti
- **Esempi:**
«domani per rischio allagamento le scuole rimangono chiuse.»
«Molti disagi e allagamenti nel maceratese. La pioggia incessante ha provocato frane e smottamenti»

VIABILITÀ

- **Set termini:** Percorribile, Percorribili, Libera, Libere, Liberata, Liberate, Pulita, Pulite, Ripulita, Ripulite, Sgombra, Sgombre, Circolazione, Circolare, Chiusa, Chiuso, Via, Viale, Strada, Piazza, Ponte, Circonvallazione
- **Esempi:**
«giovedì_18.10 codice giallo, inizio allagamento Piazza S.Marco, nessun disagio alla circolazione in città.»
- Serve lavorare sui termini specifici che si riferiscono alla buona e alla cattiva viabilità.
- Ci sono dei termini che potrebbero sovrapporsi alla categoria “Luogo”.

LUOGO

- **Set termini:** Città, Centro, Via, Viale, Piazza, Vie, Dintorni, Vicinanze, Vicino, Periferia, Periferica, Periferico, Zona, Quartiere, Regione, Sindaco, Prefetto, Consiglio, Comune, Limitrofo, Limitrofa, Adiacente, Adiacenze, Lontano, Distanto, Posto, Località, Locale, Paese
- **Esempi:**
«via del Quartaccio allagamenti fra Quartaccio e via Boccea, macchine ferme, passano autobus?»
«RT @Genialloyd: Oggi rischio grandine in prov. SIRACUSA Ritirate le auto!»

- Categoria importante perché può fornire informazione di localizzazione in assenza di *geotagging* esplicito.
- Sarebbe utile includere nei termini nomi di vie e località prese dallo stradario, in modo da rendere la categorizzazione più specifica.

RESPONSABILITÀ

- **Set termini:** 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
- **Esempi:**
«Il temporale fa esondare il fiume Seveso. Allagamenti nella zona Niguarda: Le accuse dell'assessore»
«@elaya Ostia sotto controllo, alcuni interventi per risolvere allagamenti. Sala operativa mobile monitora situazione #meteoroma»
- Occorre ampliare il set di termini inserendo sia nomi di personalità politiche, sia di tutti gli enti principali che si occupano di emergenze.

WARNING PRE-EMERGENZA

- **Set termini:** Intasato, Intasati, Intasata, Fogne, Fogna, Fognatura, Argini, Argini, Letto, Flusso, Attenzione, Pericolo, Pericoloso, Pericolose, Pericolosi, Rischio, Pericolante
- **Esempi:**
«permane il rischio allagamenti a infernetto, palocco e tiburtina. interventi strutturali ormai ineludibili»
«Ci sono azioni contro il rischio allagamenti pubblicati sul sito del comune. Se peggiora ci sarà avviso.»

POST EMERGENZA

- **Set termini:** Risarcimento, Risarcimenti, Ricostruzione, Lamenta, Lamentele, Lamentela, Assicurazione, Assicurazioni, Assicurati, Denaro, Soldi, Assegni, Assistenza, Risolto, Risolti, Risolta

- **Esempi:**
 - «Risolti piccoli allagamenti in metro Porta Furba e Numidio Quadrato»
 - «La protesta degli alluvionati: A votare non ci andiamo più»
- Categoria che si riferisce a tutte le attività dopo che l'emergenza è passata.
- Utile per comprendere lo stato d'animo dei cittadini.

5.5 - Utilizzo dei dati all'interno di un algoritmo sintattico

Il passo successivo della metodologia ha previsto l'utilizzo dei dati precedentemente definiti all'interno dell'algoritmo sintattico ideato da Claudio Carcaci in [Carcaci, 2012].

Il motivo per cui si è scelto l'utilizzo di un analizzatore sintattico piuttosto che semantico deriva essenzialmente dalla velocità di esecuzione dello stesso, chiamato a operare in contesti *real time*.

Prima di procedere, richiamiamo gli elementi fondamentali di tale algoritmo, nonché i parametri che risultano utili per la nostra trattazione.

5.5.1 - L'algoritmo di Carcaci

Secondo la definizione proposta dall'autore, «*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.*»

In informatica, il *parsing* o *analisi sintattica* è il processo atto ad analizzare uno *stream* continuo in *input* (letto per esempio da un *file* o una tastiera) in modo da determinare la sua struttura grammaticale grazie a una data grammatica formale. Tipicamente, il termine italiano viene utilizzato per riferirsi al riconoscimento di una grammatica e alla conseguente costruzione di un albero sintattico. In questo contesto, l'algoritmo vuole operare una classificazione sintattica con metodi insiemistici.

Le entità di riferimento su cui l'algoritmo opera sono dunque testi, termini e classi. La definizione degli stessi è stata ampiamente trattata nei paragrafi precedenti. In termini algoritmici, la relazione tra queste entità è descritta dalla figura 7.

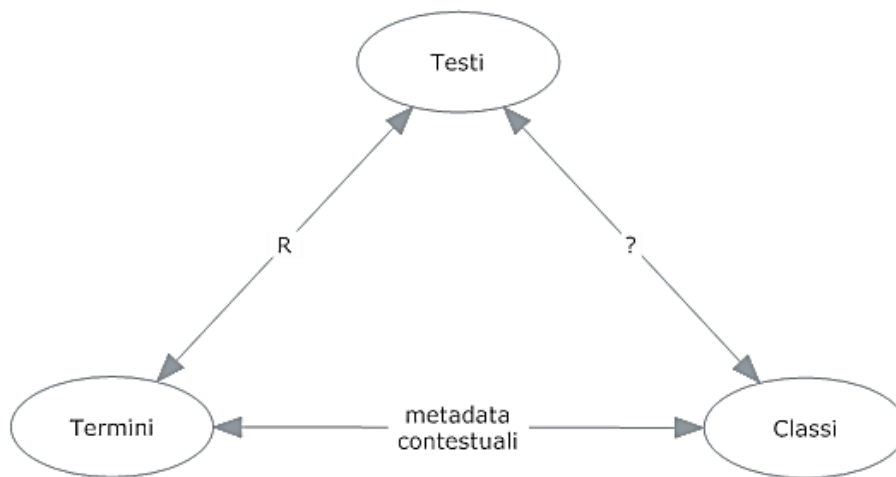


Figura 7 - Relazione tra le entità usate dall'algoritmo

1. **Relazione tra testi e termini:** essa viene espressa da una matrice, detta di rilevanza, dove vengono calcolati dei valori che rappresentano l'importanza di un termine all'interno del testo. Tale calcolo si basa sul numero di comparse di un termine nei testi del *dataset* e sul numero di comparse nel singolo testo.
2. **Relazione tra testi e classi:** la costruzione di questa relazione è quanto gli algoritmi di classificazione si propongono di fare. A livello teorico la relazione preferita come riferimento è quella funzionale: a ogni testo sia associata al massimo una classe e possibilmente sia di tipo suriettivo, ossia non vi siano classi a cui non sono assegnati testi. Nel nostro caso, tuttavia, è necessario modificare questa relazione, ammettendo che un testo possa venire associato a più classi.
3. **Relazione tra termini e classi:** questa relazione è definibile in due modi, tramite *metadata* preventivi circa i termini che identificano una classe, oppure dopo la classificazione tramite i termini che appartengono a testi classificati.

La prima versione dell'algoritmo proposto da Carcaci prevede l'inserimento in *input* di una serie di messaggi sui quali il motore esegue un'analisi sintattica: lo scopo è effettuare una suddivisione dei termini trovati e una loro classificazione all'interno di classi generate durante l'esecuzione, calcolando al contempo dei valori qualitativi riguardo alla classificazione effettuata.

Un algoritmo funzionante in questo modo è di scarsa utilità ai fini del nostro lavoro. Il nostro scopo è infatti quello di riuscire a ottenere una categorizzazione all'interno di categorie predefinite, per le quali sia già nota una sintassi di base.

Per prima cosa si è perciò preceduto a modificare opportunamente il codice in collaborazione con Carcaci. La nuova versione è dunque in grado di operare su tre elementi in ingresso:

- un insieme di messaggi, specificatamente di *tweet*;
- un insieme di categorie;
- un insieme di termini riferibili a una data categoria.

Con questa nuova versione, l'algoritmo è capace di suddividere banalmente i messaggi nelle diverse categorie, sulla base dei termini presenti in essi.

Per valutare la bontà dell'algoritmo, occorre misurare quanto gli assegnamenti testo-classe siano corretti. Bisogna anche tenere conto del contesto in cui opera l'algoritmo, ossia di quali siano gli effetti indiretti di una buona o cattiva classificazione sul sistema esterno che fa uso di tali dati.

In [Lewis] è ampiamente descritta una modalità per effettuare questa valutazione. Essa si basa sulla definizione di una tabella, denominata “di contingenza”, che associa le decisioni prese dall'algoritmo alle risposte corrette che avrebbe dovuto fornire:

Tabella 4 - Tabella di contingenza

	Risposta corretta: appartiene	Risposta corretta: non appartiene
Decisione presa: appartiene	<i>a</i>	<i>b</i> (errore seconda specie)
Decisione presa: non appartiene	<i>c</i> (errore prima specie)	<i>d</i>

Data una specifica classe e un *test set* relativo è possibile contare il numero di testi classificati correttamente o non correttamente conoscendo l'appartenenza o meno a una classe. In questo senso è possibile ottenere quattro valori numerici:

$$a, b, c, d \in \mathbb{N}^+$$

In particolare il valore b rappresenta un falso positivo che tenderà a peggiorare una metrica denominata *precision*, mentre il valore c rappresenterà un falso negativo che tenderà a peggiorare una metrica denominata *recall*.

Un'ulteriore metrica, denomina *accuracy*, si occupa invece di indicare le decisioni corrette prese, concentrandosi su una classe specifica.

In formule, le tre metriche si esprimono come segue:

1. **Precision:** è il numero di testi classificati correttamente tra i soli identificati come appartenenti

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

2. **Recall:** è il numero di testi classificati correttamente tra quelli che effettivamente appartengono alla classe

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

3. **Accuracy:** rappresenta le decisioni corrette prese per una data classe

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

Appare evidente la necessità di un'attenta e rigorosa ricerca manuale e preventiva sui *tweet* al fine di identificare i termini più comuni afferenti a ciascuna categoria, costruendo un dominio lessicale il più possibile determinato. La ricerca sintattica del motore di analisi si fa dunque tanto più raffinata quanto più complete sono le informazioni di associazione tra vocabolario e categorie indicate in ingresso. Ciò determina risultati migliori in termini di *precision* e *recall*, per il semplice fatto che queste metriche dipendono esclusivamente da come vengono scelti gli insiemi di termini, ovvero da quanto rigorose sono le associazioni tra termini e categorie. Queste due metriche da sole, pertanto, costituiscono un interesse solo parziale per i nostri test. Valutare l'*accuracy* permette invece di ottenere informazioni più dettagliate circa la bontà dell'algoritmo, perché consente di valutare in maniera più fine la sua percentuale di precisione nel prendere decisioni corrette.

Non bisogna inoltre dimenticare che l'algoritmo deve funzionare in modalità *real time*. La **velocità di esecuzione**, dunque, rappresenta un ulteriore parametro da valutare per definire la bontà dell'algoritmo di classificazione.

5.5.2 - Il sistema di test

Il sistema di test utilizzato è un comune calcolatore, di seguito le caratteristiche:

CPU: Intel® Pentium® Dual CPU T2370 @1.73GHz

Memoria principale, RAM: 2 slot di memoria DDR2 1024MB PC2-5300 @333MHz.

Memoria di massa, hard disk: 320GB 2,5'' cache 16MB SATA @7200Rpm.

Sistema Operativo: Microsoft Windows 7 Professional® Service Pack 1 64bit.

Java Virtual Machine: JRE 1.7 update 5 64bit.

Processi concorrenti in esecuzione: kernel del sistema operativo e Avast!Antivirus®.

5.5.3 - La preparazione dei dati

La prima operazione da effettuare, in previsione dei test con l'algoritmo, consiste nella preparazione dei dati che dovranno essere forniti in *input*. La preparazione consta di una prima fase di selezione dei dati, in modo da generare un *test set* di dimensione adeguata, seguita da una fase in cui quest'ultimo deve essere organizzato in un formato di *input* corretto, definito dal particolare algoritmo in esame.

Nel caso dell'algoritmo di Carcaci, il formato prescelto è costituito da due file .csv:

- **classesterms.csv**, che definisce i termini di riferimento di ciascuna classe. Esso è organizzato in sette righe (tante quante sono le nostre classi) nel formato:

nome_classe; {termini}

- **textsclasses.csv**, che riporta i *tweet* con l'indicazione della/e classe/i alla quale essi appartengono, ricavata da una preventiva analisi manuale. Esso è organizzato in tante righe quanti sono i messaggi scelti per il *test set*, e il formato di ogni riga è:

"tweet"; {classi}

La dimensione del *test set* costituisce un parametro importante perché impatta sulla velocità con cui l'algoritmo calcola i risultati e le metriche. Inoltre, dato che in uno scenario reale l'algoritmo dovrebbe lavorare su dati con un *crawling real time*, è chiaro che diventa sensato selezionare finestre specifiche di messaggi di *input* le cui dimensioni, in termini di numero di messaggi, devono discendere da osservazioni circa l'utilità pratica della selezione stessa.

La nostra prima selezione ha generato un test set iniziale di 1000 tweet. Il numero è stato scelto abbastanza spontaneamente, sulla base di osservazioni di carattere probabilisti-

co: è ragionevole infatti pensare che l'informazione su questi 1000 tweet sia di natura varia, sufficiente per coprire tutte le categorie in esame.

La selezione è avvenuta in maniera randomica sulle *keyword* utilizzate nel *crawler*. Si è cercato dunque di prendere un numero consistente di messaggi da quelli estratti con ciascuna di queste *keyword*, senza preoccuparsi della loro effettiva pertinenza con il *topic*. L'inclusione di messaggi non pertinenti, infatti, può risultare utile per verificare che l'algoritmo sia in grado di scartare, ovvero di non classificare, una certa percentuale di messaggi inutili.

Costruito il *test set*, si è proceduto alla classificazione manuale. Per ognuno dei *tweet* si è perciò indicata la classe o le classi nelle quali tale *tweet* dovrebbe ricadere. Riguardo a questa operazione, si possono seguire due strategie. La prima prevede che la scelta delle classi, fatta manualmente, segua i medesimi criteri di rigida inclusione dell'algoritmo, ovvero si consideri i termini presenti nei messaggi. I test condotti in questo modo produrrebbero risultati di *precision* assoluta, visto che opereremmo nella scelta nel medesimo modo in cui opera l'algoritmo. Per una buona valutazione delle sue metriche, invece, preferiamo seguire una strada differente, operando una classificazione manuale che sia logica, piuttosto che lessicale. Pertanto, la classificazione manuale da noi prescelta non ha considerato i termini dei *tweet*, ma piuttosto il loro senso logico e intuitivo.

Una volta selezionati i dati, essi sono stati esportati nel formato corretto, e passati finalmente all'algoritmo.

5.5.4 - Riduzione del *test set*

La prima esecuzione sui 1000 *tweet* ha purtroppo evidenziato tempi biblici per la generazione dei risultati, dell'ordine di minuti. Fermo restando che il tempo di esecuzione dipende anche dall'hardware utilizzato, al verificarsi di casi simili la strategia da seguire è quella della riduzione del *test set*.

Si è dunque proceduto a passare da 1000 a 337 *tweet*, operando una nuova selezione randomica sul *test set* originario. La scelta del nuovo ordine di grandezza non è stata casuale, ed è derivata da precedenti osservazioni di Carcaci, ricavate dall'esperienza, riguardo alle dimensioni ottimali dei *dataset* di ingresso.

La nuova esecuzione ha classificato correttamente 190 *tweet* su 337 e ha prodotto i seguenti risultati in termini di metriche prestazionali:

$Precision = 0.449$

$Recall = 0.423$

$Error = 0.226$

$Accuracy = 1 - Error = 0.774$

L'algoritmo ha agito in un tempo globale di 4.953 secondi, tra *parsing* e classificazione. Per quanto la velocità di esecuzione si sia rivelata adeguata, lo stesso non si può dire delle metriche prestazionali. Le ragioni si possono ricercare in una classificazione manuale troppo blanda, in una scelta non adeguata o incompleta dei termini associati alle classi, o nella presenza fra i *tweet* in ingresso di troppi messaggi non pertinenti.

5.5.5 - Introduzione di un *feedback* nell'algoritmo

Per risolvere il problema della scelta incompleta dei termini associati alle classi, si può agire elaborando un opportuno *feedback* nell'algoritmo, che sia in grado di popolare automaticamente l'associazione termini-classi. In pratica, se l'algoritmo fosse capace con opportune strategie di ricavare autonomamente i termini afferenti a ciascuna classe, esso aumenterebbe le proprie metriche prestazionali, a patto naturalmente che l'apprendimento fosse eseguito in maniera idonea e che il *feedback* fosse correttamente strutturato.

Il problema da affrontare consiste nell'identificare quanto un termine, non precedentemente incluso in una classe, risulti attinente ad essa, sulla base della rilevanza che esso assume nei messaggi presenti in tale classe. Analogamente, si deve poter eliminare i termini che, pur essendo stati indicati come attinenti, dall'analisi dei dati risultino invece così poco ricorrenti da essere, di fatto, inutili.

Si può procedere identificando, per ciascun termine:

- un peso, che può essere utilizzato come criterio di inclusione/esclusione per una classe;
- un'attinenza, che definisce quanto un termine attiene a una certa classe;
- una rilevanza, che definisce quanto un termine è rilevante per una classe, ovvero quante volte appare nei messaggi di tale classe.

Un termine può quindi essere incluso o escluso qualora sorpassi determinate soglie opportunamente identificate. Tali soglie sono valori prefissati di peso, varianti tra 0 e 1,

che possono essere scelti con considerazioni matematiche empiriche, sulla base delle valutazioni prestazionali dell'algoritmo e della bontà dei risultati trovati. In generale, fissata una soglia di inclusione (*inclusion percentage*), una corrispondente soglia di esclusione (*exclusion percentage*) si ricava come valore proporzionale di essa, tramite la moltiplicazione con una costante che varia tra 0 e 1.

L'algoritmo deve procedere modificando opportunamente il peso di ciascun termine grazie alle considerazioni sull'attinenza (preventivamente definita o posteriormente determinata) di quel termine a una classe, basandosi sulla rilevanza che lo stesso assume nei messaggi di tale classe.

Per ogni classe l'algoritmo estrae dunque l'elenco dei documenti che contengono almeno un termine preventivamente associato a essa, analizza le parole in tali documenti, e aggiorna una matrice, denominata di attinenza, secondo la rilevanza dei termini all'interno della classe.

A seguire, i pesi dei termini vengono modificati in questo modo:

1. ogni volta che un termine attinente a una classe è contenuto in un testo di tale classe il suo peso è rinforzato positivamente con un valore opportuno, poiché il termine effettivamente opera positivamente per quella classe. Tale rinforzo viene aggiunto a un valore iniziale diverso da zero in quanto il termine è attinente alla classe;
2. ogni volta che un termine contenuto in un testo di una classe non è indicato come attinente alla classe il suo peso è comunque rinforzato positivamente con un valore opportuno poiché il termine può essere utile per la classe anche se non è specificato nei *metadata* relativi ad essa. Il valore iniziale è in questo caso zero, il che significa che raggiungerà il valore soglia solo se compare con buona frequenza in testi in cui compaiono altri termini specificati nei *metadata*, ovvero se assume rilevanza;
3. ogni volta che un termine specificato come attinente a una classe non è contenuto in un testo di tale classe il suo peso è ridotto di un valore opportuno, poiché è facile che, se non compare in molti testi dove sono contenuti gli altri termini, avrà una bassa rilevanza per quella classe e dovrà essere escluso.

In seguito, in base al superamento del valore soglia prescelto, si includeranno perciò nuovi termini attinenti e i relativi testi, oppure al contrario si escluderanno i termini meno attinenti e i testi che li contengono.

5.5.6 - Esecuzione dell'algoritmo con *feedback*

Un'ultima esecuzione dell'algoritmo, questa volta provvisto di *feedback*, è stata eseguita su un'ulteriore riduzione del *test set* a 90 *tweet*, meglio raffinati in termini di classificazione manuale.

Viene fornita una tabella che riporta i valori dei tempi di esecuzione delle varie fasi, rispetto a valori diversi della soglia di apprendimento.

In tabella, t_{par} indica il tempo di *parsing*, t_{rel} indica il tempo di calcolo della matrice di rilevanza + sua riduzione matriciale, e t_{exe} indica il tempo di classificazione. Il tempo totale dell'algoritmo è dato da $t_{par} + t_{rel} + t_{exe}$.

Tabella 5 – Tempi di esecuzione dell'algoritmo di Carcaci

INCLUSION PERCENTAGE	Metriche			
	t_{par} (s)	t_{rel} (s)	$t_p=t_{par}+t_{rel}$ (s)	t_{exe} (s)
0.1	0.687	0.328	1.015	0.718
0.2	0.640	0.345	0.985	0.655
0.3	0.655	0.329	0.984	0.671
0.4	0.655	0.329	0.984	0.718
0.5	0.624	0.329	0.953	0.655
0.6	0.640	0.329	0.969	0.656
0.7	0.624	0.329	0.953	0.656
0.8	0.639	0.344	0.983	0.639
0.9	0.640	0.329	0.969	0.640

Anche i risultati qualitativi sono forniti in una tabella dove è riportato il valore di soglia relativo all'apprendimento (Tabella 6).

Tabella 6 - Risultati qualitativi dell'algoritmo di Carcaci

INCLUSION PERCENTAGE	Metriche		
	Precision: P	Recall: R	Error: E
0.1	0.634	0.837	0.219
0.2	0.647	0.834	0.208
0.3	0.647	0.826	0.211
0.4	0.646	0.821	0.214
0.5	0.641	0.782	0.220
0.6	0.658	0.783	0.206
0.7	0.651	0.712	0.225
0.8	0.651	0.637	0.230
0.9	0.662	0.592	0.229

Sulla base di questi risultati, è possibile osservare come l'algoritmo privilegi leggermente la *recall* per valori bassi di *inclusion percentage*: questo significa che esso assegna effettivamente i testi nelle classi con una precisione accettabile. A inficiare sul risultato non impressionante della metrica di *precision* è invece il fatto che i testi sono brevi, e risulta difficoltoso classificarli con precisione massima in tempo reale; tuttavia la classificazione ha valori di precisione sufficienti. La metrica di *recall* è invece ottima: l'algoritmo effettivamente rileva i testi afferenti alle diverse classi.

5.5.7 - Alcuni commenti relativi all'algoritmo

Seguono alcuni commenti riguardo alle problematiche critiche evidenziate durante l'utilizzo dell'algoritmo sintattico, che possono valere come osservazioni generali.

Le soglie di apprendimento

L'algoritmo esposto fa uso di due soglie per l'apprendimento dei termini di riferimento per una data classe, l'una di inclusione, l'altra di esclusione. Le soglie sono proporzionali tra loro quindi sono in effetti identificabili da un singolo parametro. Non è possibile sapere a priori quale sia il valore che garantisce le prestazioni migliori, conviene piuttosto calcolarlo con l'esperienza, in riferimento al problema considerato.

Questo valore dipende però da due elementi fondamentali:

1. **Dimensione dell'insieme di testi in ingresso:** per valori molto bassi della soglia di inclusione un insieme di testi molto numeroso potrebbe portare ad una continua aggiunta ed eliminazione di termini, generando instabilità nell'algoritmo. Per valori alti della soglia di inclusione, invece, l'algoritmo tenderà ad apprendere pochi termini nuovi senza migliorare le prestazioni su testi brevi. In particolare per $\text{INCLUSIONPERCENTAGE} = 1$ l'algoritmo non apprenderà alcun termine nuovo.
2. **Numero di termini di riferimento per le classi:** una classe alla quale siano associati pochi termini all'interno dei *metadata* tenderà a modificare poco i pesi degli altri termini, rendendo difficile l'apprendimento. Questo farà preferire una soglia bassa di $\text{INCLUSIONPERCENTAGE} \cong 0$. Viceversa una classe con tantissimi termini predefiniti tenderà a modificare molto i pesi relativi agli altri, rendendo l'apprendimento fin troppo invasivo.

Riduzione della complessità di calcolo

Dopo aver effettuato il *parsing* dei testi è possibile, opzionalmente, ridurre le dimensioni dello spazio di analisi tramite operazioni di *stemming* e *stop-word removal*. Queste attività riducono la dimensione della matrice di rilevanza migliorando in prima approssimazione la velocità di calcolo. [Carcaci, 2012]

Lo *stemming* consiste nell'effettuare la rimozione dei suffissi nelle parole in modo da considerare i termini nella loro forma radice (denominata *tema*), riducendo così la complessità dell'albero sintattico. A seconda della lingua considerata, esistono *stemmer* predefiniti che possono essere sfruttati all'occorrenza.

Lo *stop-word removal*, invece, consiste nell'eliminazione preventiva di tutte quelle parole ricorrenti nel linguaggio dei testi di *input*, come p.es. articoli e preposizioni, del tutto inutili ai fini dell'analisi sintattica.

5.6 - I dati in uscita

Per concludere la nostra metodologia, ci limitiamo a riassumere gli elementi di uscita ricavati dall'analisi manuale e dall'elaborazione dell'algoritmo.

L'algoritmo ha suddiviso i testi in ingresso, nel nostro caso *tweet*, all'interno di classi predefinite da noi individuate a seconda dei nostri scopi. La classificazione è avvenuta con un'accuratezza definita da metriche qualitative fornite dall'algoritmo stesso.

Il tempo di esecuzione costituisce un termine fondamentale per valutare la possibilità di utilizzare l'algoritmo in un contesto di analisi dei dati in tempo reale: è dunque necessario calcolarlo con accuratezza, in modo da modificare opportunamente i *dataset* di *input* utilizzati al fine di operare al meglio.

L'algoritmo ha poi generato nuovi termini di inclusione nelle classi, modificando l'analisi manuale effettuata primariamente e il contesto stesso di questa analisi.

È necessario operare sui dati di uscita con una nuova validazione manuale, così da verificare che i nuovi dati forniti siano utili allo scopo.

L'ampliamento dei termini fornito in maniera automatizzata, in generale, andrà a modificare il dominio lessicale e ontologico del sistema, e costituirà la base di successivi studi multidisciplinari sullo stesso.

5.7 - Estendibilità ad altri contesti di emergenza

La metodologia fornita si è concentrata sulle alluvioni, ma nulla vieta la sua estendibilità a scenari differenti. Tale estendibilità si può realizzare semplicemente ridefinendo il dominio, ovvero effettuando una nuova analisi manuale sui termini più utilizzati nello specifico contesto in esame, e ricostruendo dunque l'associazione tra classi e termini. L'algoritmo può essere riutilizzato tale e quale.

5.8 - La geolocalizzazione

Uno delle questioni principali quando si parla di emergenze, è quella della geolocalizzazione, ovvero del posizionamento di un dato messaggio nello spazio, in modo da comprendere a quale punto si riferisca una specifica segnalazione. La geolocalizzazione su Twitter è talvolta esplicita, grazie a opportuni strumenti forniti dal sito, ma solo una bassissima percentuale di *tweet*, nella pratica, è oggi geolocalizzata.

Nella nostra metodologia abbiamo proposto un modo per effettuare la geolocalizzazione sulla base dell'analisi sintattica del messaggio, grazie all'utilizzo della categoria "Luogo". Una possibile miglioria di questo sistema potrebbe venire dall'uso esplicito, fra i termini associati alla categoria, dei nomi delle strade e dei luoghi di interesse più famosi

delle principali città italiane. Questo consentirebbe una rilevazione e una categorizzazione più fine per quei *tweet* che trasportano un'informazione di localizzazione.

Qualcosa di ancor più utile si potrebbe ottenere definendo una classe per ogni luogo identificabile, andando poi a strutturare tutte queste classi in maniera intelligente e aggregata. Ovviamente molti messaggi rischierebbero in tal caso di finire in più classi: unendo però il contenuto informativo di ognuna si potrebbe ottenere un'informazione precisa dei luoghi di interesse per il sistema.

Conclusioni

Il lavoro presentato, inserito nell'ambito del progetto TORCIA, ha voluto fornire una possibile metodologia per la definizione di un dominio lessicale e sintattico in un contesto di emergenza, allo scopo di utilizzare detto dominio in un motore di analisi dei messaggi provenienti dai *social network*. L'analisi costante del parlato sociale è infatti risultata utile durante la gestione di diverse emergenze, avendo permesso alle autorità di avere un quadro più preciso della situazione, grazie al paradigma del *crowdsourcing*.

La strategia seguita per la definizione del dominio è stata di tipo *bottom-up*, ovvero si è partiti da un'analisi manuale sui messaggi veicolati da Twitter, in modo da definire i termini più ricorrenti durante un'emergenza – nel nostro lavoro, un'alluvione – e proporre una possibile categorizzazione utile a restituire un'informazione strutturata.

L'utilizzo di un motore sintattico reale, sviluppato da un altro tesista del Politecnico di Milano, ha poi permesso di effettuare le prime valutazioni qualitative sul nostro studio, e di porre le basi per un ampliamento successivo dello stesso.

È stata proposta una strategia per migliorare l'algoritmo, in modo da introdurre un *feedback* che permetta di ampliare automaticamente il dominio, grazie alla ricorrenza e all'importanza dei termini ricavati dall'analisi automatica dei testi. Questa strategia può essere migliorata, al fine di trovare un compromesso tra affidabilità del sistema e qualità dell'*output* fornito. I test effettuati hanno permesso di considerare gli aspetti cardine sui quali lavorare, come il tempo di elaborazione e le metriche qualitative, che consentono di valutare se l'algoritmo sia in grado o meno di restituire risultati corretti.

La metodologia proposta si è concentrata solo su contesti specifici, ma una semplice strategia di riadattamento ad altri scenari permetterà la sua applicazione a qualsiasi ambito critico: si è osservato come i passi da seguire si mantengano costanti e di come sia sufficiente una ridefinizione lessicale compatibile con lo scenario di interesse.

Bibliografia

Asproni B., “Le alluvioni in Piemonte del 1994 e del 2000”, 13 Novembre 2007,
<http://www.meteoWebcam.it/articolo-meteo-storia/16/Le-alluvioni-in-Piemonte-del-1994-e-del-2000.html>

Balestreri G., “Twitter vola a 500 milioni di utenti. Facebook vede la Borsa, ma frena”,
Repubblica.it, 22 Febbraio 2012,
http://www.repubblica.it/economia/finanza/2012/02/22/news/twitter_utenti_record-30309912/

Baree M., “Citizen Voices”, Forbes Magazine, 20 November 2008

BBC News, “Japan earthquake: Tsunami hits north-east”, 11 March 2011,
<http://www.bbc.co.uk/news/world-asia-pacific-12709598>

Bennato D., Benothman H., Panconesi A., “La diffusione delle informazioni online - Il caso Twitter”, AIS – Associazione Italiana di Sociologia, IX convegno nazionale, Milano, 23-25 Settembre 2010

Bignam D.F., “Protezione civile e riduzione del rischio disastri”, Maggioli editore, 2010

Cameron M., “Tapping into Social Media to Build Emergency Situation Awareness”,
CSIRO ICT Centre, 2012

Cameron M., Power R., “Emergency Situation Awareness from Twitter for Crisis Management”, Lione, Aprile 2012

Carcaci C., "Classificazione sintattica con metodi insiemistici", Dipartimento di Elettronica e Informazione, Politecnico di Milano, 2012

Cha M., Haddadi H., Benevenuto F., Gummadi K.P., "Measuring User Influence in Twitter: The Million Follower Fallacy", ICWSM 2010: Proceedings of international AAAI Conference on Weblogs and Social, 2010

Denti E., "Linguaggi e grammatiche", DEIS, Università degli studi di Bologna, 2010

European Commission, "Green Paper On a European Programme For Critical Infrastructure Protection", COM(2005) 576 final, Brussels, 17 Novembre 2005

Fischetti M., "Drowning new Orleans", Scientific American, 285(4):76-85, 2001

Geri F., Luccone M.G., Speranza C.M., "L'early warning e il ciclo dell'emergenza nella gestione dei rischi indotti da eventi naturali su impianti a rischio di incidente rilevante"

Hand D., Mannila H., Smyth P., "Principles of data mining", The MIT Press, 2001

Howe J., "The rise of Crowdsourcing", Wired, Issue 14.06 - June 2006

Howe J., "Crowdsourcing", Random House Inc., 2009

ISCOM, "Gestione delle emergenze locali", Ministero delle comunicazioni

Jul S., "Structuring the problem space of user interface design for disaster response technologies", Information Systems for Emergency Management, 2009

Kaplan A.M., Haenlein M., "Users of the world, unite! The challenges and opportunities of social media", Business Horizons, Vol. 53, Issue 1, p. 59-68, 2010

Kwak H., Lee C., Park H., Moon S., “What is Twitter, a Social Network or a News Media?”, Proceedings of the 19th International World Wide Web (WWW) Conference, Raleigh NC (USA), 26-30 Aprile 2010

Lerman K, Ghosh R., “Information contagion: An empirical study of the spread of news on Digg and Twitter social networks”, USC Information Sciences Institute, Marina del Rey CA (USA), 2010

Lesperance, Godinez, Olson, “Social networking for emergency management and public safety”, PACIFIC NORTHWEST NATIONAL LABORATORY, 2010

Lewis D.D., “Evaluating text categorization”, Computer and Information Science Dept., University of Massachusetts

Lindsay B.R., “Social Media and Disasters: Current Uses, Future Options, and Policy Considerations”, 6 Settembre 2011

Marya R., “Myspace reduces staff by 47%”, mashable.com, 11 Gennaio 2011, <http://mashable.com/2011/01/11/myspace-reduces-staff-by-47-percent/>

Mazzoni E., “Reti sociali e reti virtuali: la Social Network Analysis applicata alle interazioni su Web”, in A. Salvini (a cura di) “Analisi delle reti sociali. Teorie, metodi, applicazioni”, Franco Angeli Edizioni, 2007

Memoli R., “L’analisi delle reti sociali”,

<http://w3.uniroma1.it/dcnaps/memoli/documenti/Le-reti-sociali.PDF>

Morrow N., Mock N., Papendieck A., Kocmich N., “Independent Evaluation of the Ushahidi Haiti Project”, DISI – Development Information Systems International Ushahidi Haiti Project, 8 Aprile 2011

O'Hear S., "Blur Group trades Madison Avenue for its Creative Services Exchange", TechCrunch.com, 22 Febbraio 2011, <http://techcrunch.com/2011/02/22/blur-group-trades-madison-avenue-for-its-creative-services-exchange/>

Pace I., "CroSafe: an Online Social Network for helping the Emergency Mitigation phase through Crowdsourcing", Dipartimento di Elettronica e Informazione, Politecnico di Milano, 2011

Piselli F., "L'analisi di network nelle reti sociali", Donzelli Editore, 2001

Rogers E.M., "Diffusion of informations", Free Press, 1962

Rossitto R., "I numeri di Twitter in Italia", Wired, 26 Settembre 2012, <http://daily.wired.it/news/internet/2012/09/26/numeri-twitter-italia.html>

Salerno J., Hinman M., Boulware D., "Building a framework for situation awareness", AIR FORCE RESEARCH LAB ROME NY INFORMATION DIRECTORATE, Luglio 2004

Wikipedia, "Alluvione", <http://it.wikipedia.org/wiki/Alluvione>

Wikipedia, "Alluvione della Lombardia del 1995", http://it.wikipedia.org/wiki/Alluvione_del_1995_%28Lombardia%29

Wikipedia, "Ciclone tropicale", http://it.wikipedia.org/wiki/Ciclone_tropicale

Wikipedia, "Crowdsourcing", <http://it.wikipedia.org/wiki/Crowdsourcing>

Wikipedia, "Lista di alluvioni e inondazioni in Italia", http://it.wikipedia.org/wiki/Lista_di_alluvioni_e_inondazioni_in_Italia

Wikipedia, "Rete sociale", http://it.wikipedia.org/wiki/Rete_sociale

Wikipedia, “Social media”, http://it.wikipedia.org/wiki/Social_media

Wikipedia, “Uragano”, <http://it.wikipedia.org/wiki/Uragano>

Wikipedia, “Uragano Gustav”, http://it.wikipedia.org/wiki/Uragano_gustav

Wikipedia, “Terremoto”, <http://it.wikipedia.org/wiki/Terremoto>

Wikipedia, “Terremoto e maremoto del Tōhoku del 2011”,
http://it.wikipedia.org/wiki/Terremoto_e_maremoto_del_T%C5%8Dhoku_del_2011

Wikipedia, “Tromba d’aria”, <http://it.wikipedia.org/wiki/Tornado>

Wikipedia, “Vulcano”, <http://it.wikipedia.org/wiki/Vulcano>