ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

CAMPUS DI CESENA

SCUOLA DI SCIENZE

CORSO DI LAUREA TRIENNALE IN INGEGNERIA E SCIENZE INFORMATICHE

APPLICAZIONE WEB PER VISUALIZZARE E GESTIRE DATI ESTRATTI DA TWITTER

Tesi in

Tecnologie Web

Relatore:
Prof.ssa Paola Salomoni

Presentata da:
Giulia Lucchi

Correlatore:
Dott.ssa Catia Prandi
Dot. Stefano Cacciaguerra

Sessione II
Anno Accademico 2015/2016

A mio nonno Renzo
che quando mi lamentavo per lo studio ripeteva sempre
"Ma lassa sté . . . ma va a stùdie valà."

Elenco delle figure

Elenco dei codici

Indice

Introduzione

Gli ultimi anni sono stati un periodo di rivoluzione e innovazione, tanto che gli esperti hanno dato anche un nome preciso a questa fase di evoluzione: il "Web 2.0". Questo ha portato con sé l'esplosione dei Social Media e l'incremento delle informazioni disponibili sul Web, sotto forma di User Generated Content [49]. In questo scenario diventa l'utente la figura centrale, in quanto diviene egli stesso il creatore dei contenuti di questo "nuovo" Web. Questa partecipazione dell'utente e conseguente aumento di dati sfocia nella nascita dei "Big Data". Quest'ultimo termine ha impatto su tutti i aspetti collegati al web. Risulta interessante come ora, infatti, si abbia una crescita esponenziale di dati di natura destrutturata, diventati così il nuovo tesoro dell'informazione, contenente testi, informazioni e opinioni. Tutto questo quindi ha portato con sé l'esigenza di nuovi strumenti, che permettano un'analisi e un'estrazione di informazioni significative [50].

Il "Web 2.0" può essere definito un cambiamento tecnologico, ma anche culturale. L'impatto sociale del Web è una conseguenza anche dell'utilizzo e la diffusione dei Social Media, fra cui abbiamo anche i social network, blog, community e microblogging. Lo studio dei social network e dei microblog si è rilevato, fin dall'inizio, di grande utilità, poiché ha dato la possibilità di monitorare le persone, le loro relazioni e i pensieri che in modo costante ormai pubblicano.
Queste piattaforme rappresentato un enorme raccoglitore di contenuti di svariato genere, che hanno un grandissimo potenziale per ricerche di ogni tipo. Su questi dati quindi si possono fare ricerche sull'umore, su malattie ed epide-

mie, estrapolare informazioni o comportamenti tipici della società, prevedere situazioni che riguardano la società o il mondo e infine anche analizzare situazioni di emergenza in cui c'è bisogno di aiuto.
La tesi infatti si concentrerà sull'impatto dei social media su situazioni di emergenza. Una volta estrapolati i dati, il passo successivo consiste nel dare una forma a questi dati e riuscire, attraverso i nuovi strumenti, nati con l'avvento dei cambiamenti sopracitati, ad analizzarsi per trarre conclusioni e risultati utili, andando al cuore dell'informazione.

La mia analisi si concentra prevalentemente su Twitter, uno dei microblog più famoso e utilizzato a livello mondiale [34], e sulla visualizzazione del traffico dei dati che è stato prodotto durante la grossa scossa sismica in Centro Italia il 24 Agosto 2016 e i giorni successivi a questo evento. Twitter si è rivelato l'ambiente più "serio e sincero", ed è per questo che i suoi contenuti, in contrapposizione agli altri social network, rappresentano in modo più veritiero il pensiero delle persone in quel particolare evento.

Il lavoro svolto mira, quindi, alla progettazione e allo sviluppo di un sistema d'analisi, effettuato tramite un'insieme di tecnologie diverse, in grado di esaminare ciò che gli utenti pubblicano. In generale l'idea di base dello studio consiste nell'analizzare un breve testo, nel nostro caso quindi si parla di "tweet", e dimostrare o confutare l'utilità concreta di Twitter nelle situazioni di emergenza, come può essere quella dei terremoti avvenuti nel Centro Italia, evento alquanto significativo e attuale. Per fare questo si è deciso anche di ampliare quest'opportunità a tutti coloro che possiedono un insieme di tweet già completo, su cui effettuare una ricerca, senza possedere specifiche abilità di programmazione tramite la creazione di un applicazione web online.

Il progetto di tesi in questo documento è stato strutturato considerando sia la componente tecnologica sia quella sociale e culturale relativa al particolare contesto preso in considerazione.

Nel primo capitolo si vuole dare una panoramica dei cambiamenti tecnologici, sociali e culturali partiti con la nascita del "Web 2.0". A seguito di una sua breve introduzione, intendo spiegare i principi base del "Web 2.0",

concepiti e spiegati in modo preciso da O'Really, nel suo documento in cui conia il termine che tutt'ora noi utilizziamo. In seguito andremo a toccare in particolare gli aspetti di nostro interesse: i Social Media (e di conseguenza i social network) e la nascita di quelli che vengono definiti "Big Data". Infine andiamo a concentrate la nostra attenzione su Twitter, la piattaforma protagonista della nostra analisi.

Nel secondo capitolo si vuole andare a illustrare la situazione effettiva degli eventi sismici avvenuti quest'anno in Centro Italia. Ci focalizzeremo, in particolar modo, sulla prima forte scossa del 24 Agosto 2016, oggetto dell'analisi trattata come progetto di tesi, ponendo una particolare attenzione sull'impatto che questo ha avuto sul web in particolare sui social network. Infine andremo a delineare l'obiettivo e i contenuti dell'analisi in questione.

Il terzo capitolo invece risulta il cuore del progetto, in quanto spiegheremo le varie fasi dell'analisi, di conseguenza anche le tecnologie usate, fino ad arrivare ai risultati e alle conclusioni dello studio fatto.

Per concludere, nell'ultimo capitolo andremo ad illustrare in modo più dettagliato i requisiti e il design dell'applicazione, la quale abbiamo creato con lo scopo di rendere possibile a tutti gli utenti di effettuare un'analisi iniziale del problema. L'applicazione si chiamerà "INFORMATION DISCOVERED". A conclusione andremo anche a esporre i possibili sviluppi futuri della presente applicazione.

Capitolo 1

Social Media e Big Data

Il web nel corso del tempo ha subito tanti cambiamenti fino ad arrivare ad ora. Queste trasformazioni convergono in una nuova definizione del web: "web 2.0" [49].
Questa nuova visione del web ha dato all'utente un ruolo attivo e centrale nella creazione del web stesso. Tutto ciò quindi ha avuto anche un impatto a livello sociale, dovuto in maniera consistente alla diffusione e creazione dei Social Media. Con questo termine ci riferiamo ad una miriade di strumenti, i quali danno all'utente modo di partecipare in modo attivo sul Web. In particolare, possiamo notare questo aspetto in modo più visibile sui Social Network, piattaforme che, ad oggi, hanno acquisito maggior rilievo.
Una conseguenza di questa rivoluzione del Web consiste in una crescita repentina dei dati da gestire e per questo nasce anche il concetto di "Big Data".

Fra i Social Media, una delle le cause dell'incremento della quantità di dati, possiamo trovare una miriade di piattaforme social fra cui i "Microblog". Questa forma di Social Media si è diffusa in modo capillare, divenendo rilevante sia per l'impatto sociale e informativo sia dal punto di vista dell'utenza. L'esempio più valido di Microblogging risulta "Twitter",oggi una delle piattaforme più popolari fin dalla sua nascita. In questo capitolo, quindi, faremo una panoramica sul periodo sul del "Web 2.0" concentrandoci in modo par-

ticolare sui temi già sopra citati, concentrandoci infine su Twitter, un social media e in particolare microblog molto famoso e utilizzato.

1.1 Web 2.0

Il Web, fin dalla sua nascita, è sempre stato in perpetuo movimento, poiché non ha mai smesso di crescere e migliorare. Per questo il "web 2.0" può essere visto come un prodotto non finale dell'incessante sviluppo in atto.

Di conseguenza, il "web 2.0" è generalmente definito come una fase dell'evoluzione di Internet e in particolare del World Wide Web [65]. È da mettere in luce, però, l'attuale dibattito aperto sul vero significato e la vera definizione, nel quale c'è persino una parte che lo ritiene uno mero slogan pubblicitario, privo di qualsiasi innovazione [49].

Il termine "Web 2.0" nasce nel 2004, durante una conferenza tra O'Reilly Media e MediaLive International, ad opera di Dale Dougherty e Tim O'Reilly. A O'Reilly, nato del 1954 a Cork, fondatore della O'Reilly Media, casa editrice internazione specializzata nella diffusione di testi di carattere tecnico informatico, viene riconosciuto il merito di aver tentato di dare una definizione chiara ed esauriente del termine. La prima definizione ufficiale di O'Reilly, la quale risulta anche più completa,è la seguente:

> *"Il Web 2.0 è la rete come piattaforma, attraverso tutti i dispositivi collegati; le applicazioni Web 2.0 sono quelle che permettono di ottenere la maggior parte dei vantaggi intrinsechi della piattaforma, fornendo il software come un servizio in continuo aggiornamento che migliora più le persone lo utilizzano, sfruttando e mescolando i dati da sorgenti multiple, tra cui gli utenti, i quali forniscono i propri contenuti e servizi in un modo che permetta il riutilizzo da parte di altri utenti, creando una serie di effetti attraverso un "architettura della partecipazione" e andando oltre la metafora delle pagine del Web 1.0 per produrre così user esperienze più significative"* [43].

Nonostante l'incertezza del significato del termine scelto per denominare questo sviluppo del web, la nostra attenzione deve focalizzarsi su ciò che si porta dietro questo termine e sui concetti di base:

- Web come piattaforma
- Web centrato sull'utente
- Web centrato sui dati
- Enterprise 2.0
- Architettura partecipativa
- Evoluzione tecnologica.

.

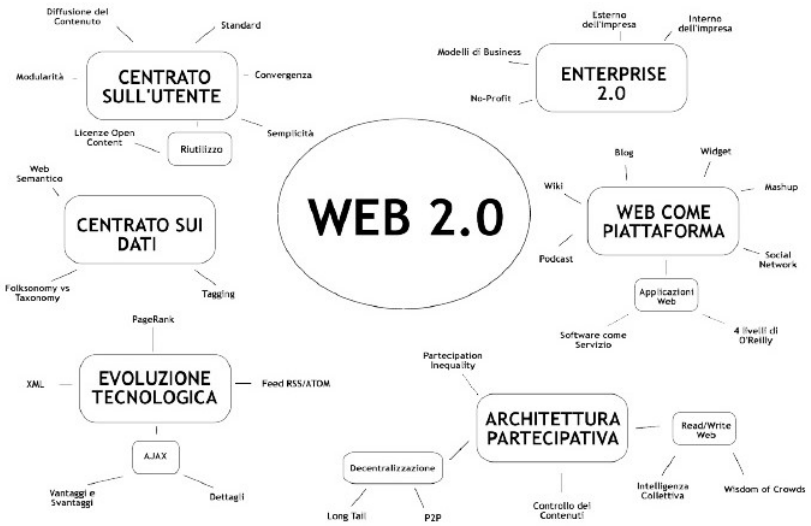


Figura 1.1: Mappa Concettuale Web 2.0 - concetti base [39]

1.1.1 Web come Piattaforma

Il principio del "Web come piattaforma" è il perno attorno al quale O'Reilly fa girare il concetto di "Web 2.0", formulato di già nella prima conferenza

nell'Ottobre 2004. Non si deve quindi considerare "Web 2.0" come un'applicazione, ma bensì il concetto di "piattaforma" suggerisce una visione più globale e dinamica: non si parla più di un insieme di siti, applicazioni e risorse, ma di un insieme di servizi del web a misura dell'utente, quali strumenti comunicativi e applicazioni web [44].

1.1.2 Architettura Partecipativa

Con l'avvento del "Web 2.0" è cambiato anche il modo di vedere il web: ora gli utenti da semplici fruitori del web divengono utilizzatori attivi, poiché gli è lecito creare e diffondere informazioni. Si è quindi costruita una fiducia radicale che consente a gruppi di utenti distribuiti su scala internazionale di contribuire alla creazione della conoscenza. Quindi ora l'associazione fra web e utente diviene bidirezionale, dando al web un valore aggiunto. Tale rivoluzione ha riguardato i diversi settori: dall'educazione, al business, fino ad arrivare alla vita di tutti i giorni, riducendo drasticamente le barriere tecnologiche, che erano presenti in precedenza, tramite anche l'uso di strumenti online e spesso gratuiti [44].

1.1.3 Evoluzione Tecnologica

L'architettura di rete di base sulla quali si reggeva la concezione del web precedente è la stessa di quella attuale. Quella attuale però ha subito modifiche di particolare importanza. Venne modificato in particolare ciò che già in precedenza Holdener criticava:
"Gli strumenti delle pagine web classiche sono in realtà più simili al legno solido o al composto con i quali venivano costruiti i muri del periodo neolitico. Era greggio e semplice, serviva per il suo scopo, ma lasciava molto a desiderare" [57].

Oltre agli strumenti standard e ormai consolidati per creare pagine web, ora entra in scene una nuova tecnica di sviluppo per applicazioni web interattive, denominata AJAX. Tramite questo la pagina web è in grado di recupe-

rare i dati in modo asincrono dal server, senza inferire con la visualizzazione e il comportamento della pagina. Questo scambio di dati in background fra web browser e server consente l'aggiornamento dinamico della pagina [62]. Inoltre c'è stato un intervento anche per quel che riguarda gli aggregatori, in particolare vengono menzionati due principali famiglie di feed: RSS e ATOM. I feed sono un modo rapido ed efficiente di distribuire i contenuti del Web. In particolare sono unità d'informazioni, formattate di genesi in XML, che permettono di interpretare e interscambiare il contenuto fra diverse piattaforme e applicazioni. Inoltre un'importante funzione, sopra citata, consiste nei feed visti come aggregatori di notizie, in grado quindi di effettuare un download di un flusso, così da visualizzare i contenuti in base alle preferenze dell'utente [63].
I primi ad sono stati i feed RSS. Questi sono basati su XML e sono flussi che danno la possibilità di essere aggiornati su nuovi articoli o commenti senza un dover fare un'apposita ricerca manualmente [23]. Attualmente, invece, quelli più utilizzati sono i feed ATOM. La differenza tra i feed RSS e ATOM esiste, ma riguarda soprattutto il codice con cui sono realizzati e il tipo di sintassi e linguaggio utilizzato. Inoltre i feed ATOM, a differenza dei feed RSS, sono stati realizzati con l'idea di essere più robusti a livello strutturale e avere qualche funzionalità in più [42].

1.1.4 Modello Centrato sui Dati

I dati e le informazioni divengono il nucleo principale del web, visto anche la possibilità di un architettura partecipativa. Un' importante questione si riscontra nel recupero dei dati e quindi nella sua conseguente classificazione, in quanto, potendo gli utenti creare contenuti da inserire all'interno del web, si presenterà un incremento della quantità dei dati proposti.

Nasce, quindi, un nuovo concetto tecnologico: folksonomia. Il termine folksonomia, dall'inglese "tassonomia del popolo", sta a indicare la classificazione e la categorizzazione dei contenuti dagli utenti stessi, senza alcuna autorità centrale. Questa può definirsi una "classificazione collaborativa",

in quanto non c'è alcun tipo di gerarchia, ma al contrario la classificazione avviene tramite parole chiave, dette tag. Di conseguenza un numero sempre crescente di classificazioni porterà ad una maggior precisione della folksonomia, rispecchiando i modelli concettuali degli utenti stessi [49][20].

Questa innovazione nel recupero delle informazioni ha indirizzato anche verso una forma di visualizzazione per i tag utilizzati, chiamata tag-cloud. Questa consiste in una lista di tag in ordine alfabetico e con font di diversa dimensione in base alla rilevanza del termine in questione. La rilevanza di questo termine può essere calcolata in due modalità differenti:

- in base alla frequenza dell'utilizzo dell'etichetta all'interno di un determinato sito
- in base al numero di volte per il quale tag viene consultato [36].

Questa grande quantità di dati e questa attenzione per la classificazione e categorizzazione di dati, porta un'innovazione anche nella creazione degli URL.
Considerando che i motori di ricerca si basano sull'analisi semantica e sulla rilevanza dei tag, le nuove tecniche di creazione degli URL devono seguire questa strada, rendendoli indicizzabili [49].

1.1.5 Modello Centrato sull'Utente

Una delle caratteristiche più significative riguardanti il "web 2.0" è la centralità della figura dell'utente nel web. Considerando ciò che è stato detto fino a qui, si può arrivare alla seguente conclusione: l'architettura partecipativa e utente visto come valore aggiunto è la prova che, senza l'utente, il web, considerato come lo consideriamo oggi, non esisterebbe, ma sarebbe rimasto solo un "raccoglitore di informazioni", come lo sono le enciclopedie cartacee.

L'utente quindi è attivo all'interno del web e può essere propulsore di idee innovative, che tutti possono riutilizzare. Questo è reso possibile da un importante aspetto, ossia la semplicità con il quale un utente medio riesce a navigare sul web. Questo è dovuto grazie all'avvenuto un abbassamento

delle barriere d'entrata nei confronti del vecchio web, incoraggiando il riuso delle idee e delle informazioni altrui [44][49].

Questa nuova visione ha influenzato anche il settore giuridico che, per incoraggiare l'utente, ha lavorato per creare licenze apposite, quale la Creative Commons License. Quest'ultima da una parte incoraggia l'utente e la sua "libertà" di condivisione e fruizione di informazioni e dall'altra salvaguarda i diritti, la proprietà intellettuale e la fonte dell'idea in questione [7].

L'insieme dei contenuti apportati dagli utenti nel web viene definito "User Gernerated Content". Gli UGC sono la prova tangibile del valore democratico e partecipativo del "Web 2.0". È proprio questo aspetto che incentiva l'utilizzo dei social media, portati alla luce grazie a tutte queste innovazioni e miglioramenti del "Web 2.0" [49].

1.1.6 Enterprise 2.0

L' "Enterprise 2.0" può essere definito come un nuovo approccio al web, nel quale avviene una rottura dei classici sistemi di organizzazione, un distacco dai rigidi stereotipi ed un'apertura ad una collaborazione a tutto tondo [13]. Nasce quindi una progressiva evoluzione sociale ed organizzativa, che ha come obiettivo la connessione delle persone tramite gli strumenti che in questo periodo si sono sviluppati. Questa "connessione sociale" contribuisce anche a migliorare le interazioni con clienti, fornitori, partner e consulenti, portando così ad un incremento dei processi, prodotti e servizi, anche in un contesto aziendale.

Un ruolo fondamentale nella realizzazione di un"Enterprise 2.0" è quello delle tecnologie legate al "Web 2.0", come blog, wiki, social network, RSS e tutti quegli strumenti diffusi e a portata dell'utente medio [49].

1.2 Social Media

La vera rivoluzione del web quindi avviene anche in campo sociale, in quanto il web, da una vetrina nel quale rimediare informazioni, diviene un ambiente

partecipativo e interattivo. Il conseguente sviluppo e diffusione delle User Generated Content, già visto nel paragrafo 1.1.5, contribuisce alla nascita di tecnologie apposite, definite Social Media.

Con Social Media quindi si intende un insieme di strumenti con l'obiettivo di realizzare, consultare e condividere informazioni, opinioni e conoscenze in rete [58]. La novità che portano e la differenza maggiore con i mass media consiste nella fonte delle informazioni, in quanto nei mass media i contenuti sono generati e proposti da un'organizzazione centrale, solitamente un' organizzazione mediatica, al contrario dei social media in cui qualsiasi persona ha la possibilità di inserire contenuti.
Gli elementi innovativi e cardine dei Social Media sono:

- Partecipazione;
- Conversazione bidirezionale;
- Persistenza;
- Trasparenza;
- Community.

I vari elementi sono autoesplicativi e sono la conseguenze anche dell'evoluzione del web sopra ampiamente descritta. Non si può spiegare e vedere questi termini in modo distaccato l'uno dall'altro, tanto che la conversazione, che diviene di tipo bidirezionale viene introdotta per eliminare la percezione dell'utente come unità passiva, incoraggiando quindi quella che è la partecipazione. La conversazione non è l'unico contributo ad una maggior partecipazione, ma ne fanno parte anche l'utilizzo di feedback da parte degli utenti, tramite commenti, condivisioni o classificazioni [40].
Focalizziamoci sull'elemento "Community". Quest'ultima consiste in un insieme di persone interessate ad un determinato argomento ed aventi un approccio comune. La community è uno strumento che, se usato efficientemente, permette di creare interazioni fra i vari utenti partecipanti. La Community è un molto significativa, in quanto ha la capacità di riuscire a creare

un area di interazione su argomenti riguardanti vari settori e di conseguenza possono essere utili in un contesto di ricerca o in uno aziendale [66].

1.2.1 Tipologie

I social media possono essere considerati come un "ecosistema dinamico" poiché può essere visto come un insieme di organismi che interagiscono fra loro, classificandosi in base alle proprie caratteristiche e funzioni comuni [40]. Negli ultimi anni infatti sono state fatte diverse classificazioni dei social media. Fra le prime classificazioni trovate, abbiamo quella in base alla tipologia, in particolare quella di Scloble con la sua "starfish". Questa è stata la base da cui sono partite tutte le altre definizioni.

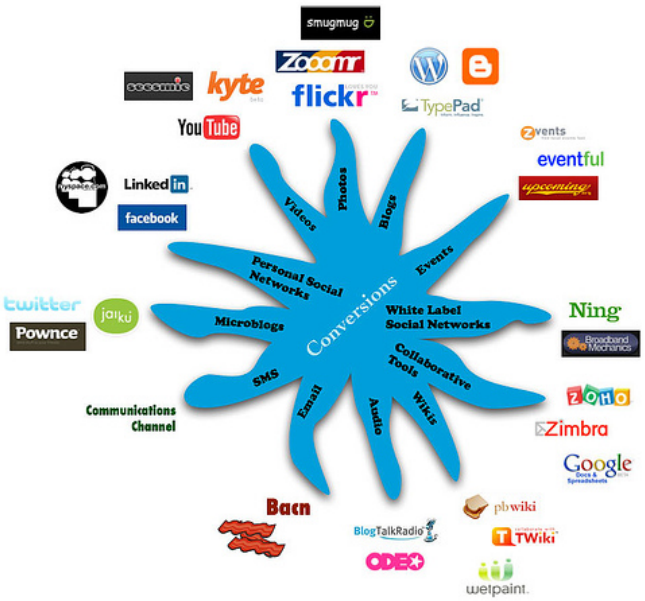


Figura 1.2: Diagramma di classificazione "Starfish, Sclobe 2007 [52]

Come si nota dalla figura 1.2, vengono rappresentate 12 tipologie diverse di Social Media, in base alla struttura:

- video;
- photo;
- blog;
- events;
- collaborative tools;
- wiki;
- audio;
- email;
- sms;
- microblogs;
- social networks.

Un' ulteriore classificazione significativa è quella in base ai servizi che vengono offerti all'utente e al loro conseguente utilizzo. Quella più completa è la versione 2012 del "Social Media Landscape", dove le classi in questione sono quelle ben evidenti nella figura seguente 1.3.

Figura 1.3: Diagramma di classificazione "Social Media Landscape", 2012 [48]

1.2.2 Social Network

La nuova affermazione del web , con tutto quello che ne consegue, vede la nascita dei veri e propri Social Network. Quest'ultimi vedono in loro un concentrato di alcuni degli aspetti più significativi, fra cui anche la semplicità, la condivisione, la partecipazione e l'interattività.

I Social Network sono definiti letteralmente come "rete sociale", in quanto si fa riferimento ad un gruppo di persone connesse fra loro tramite relazioni che possono essere di varia natura. Si nota, quindi, come questo "strumento" sia incentrato sull'utente, in particolare sul suo aspetto più sociale e "condivisibile".

La caratteristica più visibile nei Social Network è la creazione di contenuti

da parte dell'utente che andrà ad incrementare le proprie User Gernerated Content, divenendo così l'utente attivo, in particolar modo in questo tipo di ambienti [49].

I Social Network risultano piattaforme accessibili a chiunque, questo poiché non servono strumenti esclusivi e impossibili per accedervi ed inoltre il funzionamento è estremamente semplice. Per registrarsi ad un qualsiasi social network, ci sono pochissimi step guidati da seguire, che ci obbligano a fornire una serie di informazioni più o meno personali , così da creare un proprio profilo personale con il quale presentarsi agli altri, divenendo il proprio "biglietto da visita" virtuale.
Con questa scheda personale possiamo collegarci con qualche altro profilo che riteniamo opportuno, così da poter condividere e visualizzare anche i loro contenuti. È per questo che viene definita "rete sociale".

Gli elementi principali che caratterizzano un Social Network sono:

- *Creazione di un profilo*: come sopra accennato, all'interno di un profilo personale troviamo le nostre informazioni personali, ma anche tutto quel materiale che decidiamo di pubblicare sul profilo, in gergo sotto il nome di "post", ossia testo libero proveniente dall'utente in questione, video, foto, documenti, articoli e tutto ciò che più ci interessa mettere agli occhi di tutti.

- *Realizzazione di una catena*: i legami fra i vari membri di un social network vengono visti come una catena tra parenti, amici, conoscenti o persone di nostro interesse, che possiamo consultare in qualunque momento.

- *Gestione di commenti*: Ogni membro del social network può creare commenti o dare risposte a qualsiasi post pubblicato, così da poter creare una vera interazione fra tutti gli utenti. Oltre ai commenti e ai contenuti pubblicati, possiamo anche lasciare uno o più messaggi pubblici sul profilo di un altro utente [6].

1.2.3 Generazione dei Social Media

Il facile accesso al social networking ha influenzato in maniera molto significativa anche la società, tanto che ormai ogni situazione, evento, discussione pubblica o personale che sia viene affrontata sui social network, grazie a quello che un punto cardine di queste piattaforme social: la condivisione. I soggetti principali all'interno di un social network infatti non sono gli individui stessi, ma le relazioni fra gli individui, le loro azioni e il loro comportamento.

Il comportamento infatti risulta molto importante ed è anche per questo che la "We are Social" effettua analisi precise e approfondite su tutto ciò che riguarda la relazione fra social network e utente.
La "We are Social" è un' importante agenzia che combina un'innata comprensione dei social media con competenze di comunicazione e marketing e che si orienta allo sviluppo di progetti creativi, innovativi ed efficaci [4].
La ricerca di nostro interesse sulla quale ci focalizziamo si chiama "Digital in 2016". "Digital in 2016" consiste in una ricerca annuale nella quale vengono raccolti dati legati all'utilizzo dei canali social, dei dispositivi mobili e di ciò che riguarda lo scenario digitale a livello globale [35].

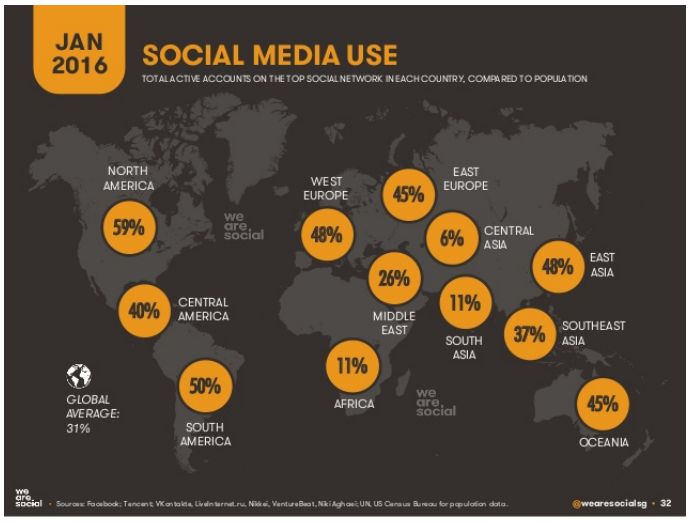


Figura 1.4: DIGITAL IN 2016: Statistica sull'uso dei Social Network- We are Social 2016 [54]

Nella figura 1.4, possiamo notare le percentuali dell'uso dei Social Network nel mondo: l'Italia è uno dei paesi in cui il valore è fra i più alti. Inoltre fra le analisi disponibili online emerge come le ore trascorse su queste piattaforme stiano incrementando sempre più. Questo fenomeno viene collegato a quello del "multi-networking" che si sta diffondendo ampiamente: in media le persone possiedono almeno un account su 5,39 social network e ne utilizzano attivamente 2,75, così dice la "Global Web Index" in una ricerca del 2015.

Nel Giugno 2016, la "Global Web Index", con i dati relativi a Q2 2016, ha focalizzato la sua attenzione sull'Italia. Emerge che l'Italia è allineata alla media globale per l'utilizzo e l'iscrizione dei social network, in quanto l'utente medio possiede profili su oltre 6 piattaforme social [14].

1.3 Big Data

L'evoluzione tecnologica insieme all'avvento del "Web 2.0" e tutto ciò che ne consegue è causa di un incremento esponenziale di dati, tanto da superare l'ordine dei Zettabyte. Questa situazione porta alla nascita di un nuovo termine: "Big Data".

Con il termine "Big Data" intendiamo quindi un'enorme aggregazione di dati che necessitano di essere trattati per mezzo di tecniche differenti da quelle tradizionali utilizzate per dataset di piccole dimensioni [50].
Per completare la definizione di Big Data è essenziale citare le caratteristiche principali [53]:

- **Volume**: ci si riferisce alla capacità di acquisire, memorizzare e accedere a grandi volumi di dati e ne rappresenta la dimensione effettiva del data-set. Inizialmente questa caratteristica poteva essere vista come un problema, ora il problema non sussiste in maniera massiccia, in quanto tecnologie, quali cloud e virtualizzazione aiutano nella gestione di grossi volumi.

- **Velocità**: si riferisce alla velocità con il quale i dati vengono generati. Questa situazione porta alla necessità di effettuare analisi in tem-

po reale, per ottenere una valutazione più significativa della ricerca in questione.

- **Varietà**: Le fonti di questi dati sono differenti ed eterogenee, ne consegue che anche i tipi di dati avranno un cambiamento fondamentale. Nei Big Data infatti non abbiamo sono dati strutturati, oggetto dei sistemi più tradizionali come i database, ma al contrario abbiamo una maggioranza di dati destrutturati, ossia dati che non possiedono una struttura predefinita. Quest'ultimi, negl'ultimi anni, sono la tipologia di dati più diffusa ed è per questo che è stato utile sviluppare nuove tecnologie per analizzare meglio questi tipi di dati.

- **Veridicità**: questa caratteristica si è aggiunta solo in seguito con il tempo e consiste nel fatto che tutti i dati raccolti costituiscono un valore e un potenziale sia per un azienda, un'organizzazione o una ricerca. È per questo che la veridicità dei dati diventa un requisito fondamentale, affinché i dati possano essere utilizzati così da avere un grande impatto sulla nostra attività.

I Big Data quindi possiedono un grandissimo impatto e vantaggio sulla vita quotidiana, portando con sé grandi opportunità riguardanti in particolar modo tre settori: business, finanziario e tecnologico.
A livello di business, hanno la possibilità di creare nuovi modelli, riuscendo così a migliorare l'andamento dell'azienda, focalizzandosi su potenziali vantaggi relativi alla competitività, alla redditività, alla tempestività e all'aumento di efficacia nei processi decisionali.
L'analisi dei Big Data inoltre può anche migliorare l'aspetto più finanziario, in quanto, attraverso tecniche e algoritmi appositi per l'analisi di questa moltitudine di dati, si riuscirebbe a portare un grande incremento dei guadagni e una relativa riduzione dei costi,
Infine, per quanto riguarda il settore tecnologico, si può notare come si è propensi ad una sempre maggiore ottimizzazione e creazione di tecnologie per l'analisi, tentando di ottenere qualità dei dati impeccabile [19].

1.3.1 Social Media e Big Data

Figura 1.5: DIGITAL IN 2016: Dati sull'accesso a Internet - We are Social 2016 [34]

Nel mondo di oggi, la maggior parte della popolazione mondiale ha la possibilità di accedere a Internet, qualsiasi sia la modalità. Questo è stato confermato anche nel Gennaio 2016 dalla "We are Social" con la ricerca "Digital in 2016, sopra citata.
I dati che vediamo nella figura 1.5 sono in crescita rispetto a quelli misurtati nel 2015, quando erano 3 miliardi le persone ad accedere a Internet contro i 3.4 miliardi di oggi e quando gli account attivati sui canali social erano 2 miliardi, a fronte degli attuali 2.3 miliardi. Il numero di utenti totali e quello degli utenti attivi sui canali social è aumentato quindi per entrambi del 10%. In seguito, nella figura 1.6 vediamo la situazione invece sulla popolazione globale, per farci un idea di quanti hanno la possibilità di un accesso sul web non più a livello globale, bensì sul territorio italiano [54].

Figura 1.6: DIGITAL IN 2016: Dati sull'accesso a Internet in Italia - We are Social 2016 [34]

Queste statistiche, quindi, dimostrano come sempre più persone, attraverso l'uso dei social media, tendono a registrare online la nuda verità su comportamenti individuali e collettivi, toccando così le diverse dimensioni sociali della nostra vita: dai desideri, agli stili di vita, alle opinioni e infine persino alle relazioni.

I Big Data e il rinnovamento del web, quindi, rendono analizzabile la società e la rete sociale in modo attuale e real-time. Tutto questo quindi sta andando verso la nascita di una nuova scienza dei dati: il social mining. Il social mining riesce quindi, con le informazioni degli utenti condivise e pubblicate sui social network, a misurare e a prevedere eventi sociali di qualsiasi genere. Si deve però tenere conto anche della qualità dei dati e della loro rappresentatività [19].

1.3.2 Microblogging: definizione

Il microblogging è una forma di pubblicazione costante di piccoli contenuti nel web, sotto forma di brevi messaggi di testo, immagini, video, audio, ma anche segnalibri, citazioni, appunti. Questi contenuti vengono pubblicati nei

Social Network, visibili a tutti, nei limiti delle impostazioni personali della privacy, o soltanto alle persone della propria Community [35].
Il microblog può essere definito come un'evoluzione del blog che si va a intersecare con il Social Network. Il blog infatti nasce con lo scopo di creare un proprio spazio virtuale, per esprimere un punto di vista personale e raccontare qualcosa di sé. Nel microblog questo non basta, in quanto si unisce all'obiettivo principale dei Social Network, quale rendere esplicita la propria rete sociale, creandosi una lista di utenti con cui condividere contenuti e attraversando anche le liste di connessione fatte anche dagli altri utenti [9].

Il microblog possiede un' alta componente sociale, in quanto è possibile inviare con un'alta frequenza brevi messaggi che possono essere visualizzati da chiunque conosca l'indirizzo del nostro microblog, rendendo possibile seguire digitalmente una persona e sapere sempre che cosa stia facendo o pensando.
L'alta componente sociale si percepisce nel modo in cui questa forma di Social Network segue le inclinazioni della società. La caratteristica principale infatti consiste nella velocità, in quanto segue la frenesia della vita e l'orientamento delle persone verso un soddisfacimento immediato dei propri desideri. La velocità è ben evidente nelle attività principali di un microblogging, in particolare nel pubblicare messaggi molto brevi, leggere i contenuti dei propri contatti, utenti, aziende o personaggi televisivi che siano e ottenere o creare in tempo reale feedback e spunti in risposta a contenuti appena letti [9].

Un altro aspetto importante del microblog, visto come fonte inesauribile di informazioni, è l'alta frequenza d'aggiornamento della piattaforma. Questo può essere considerato in parte positivamente vista la grande quantità di informazioni utili, ma d'altro canto c'è da considerare che il costante aggiornamento porta l'utente ad essere sommerso da una grande mole di contenuti ancor prima di avere letto quelli precedenti [8].
Per quest'ultima considerazione, la pubblicazione di messaggi può essere considerata come una "lotta alla visibilità". Chi scrive infatti deve fare in modo che quanto scritto attragga l'attenzione del lettore. Non basta un messag-

gio accattivante, ma bisogna utilizzare il contenuto da pubblicare come uno strumento di comunicazione, dandogli così anche il compito di trasmettere emozioni, molto importanti per questa forma di pubblicazione in quanto i contenuti del microblogging sono personali e orientati all'aspetto più "intimo" della persona.

1.3.3 Microblogging e Analisi di stati d'emergenza

Con i social network possiamo avere due approcci allo stato di emergenza da analizzare:

- Dal punto di vista dell'ente, di chi presta il soccorso e chi deve decidere come agire.
- Dal punto di vista della vittima dell'evento che ha causato l'emergenza.

Andando per ordine, per capire meglio il punto di vista di chi deve agire per conto dell'emergenza, andiamo a chiarire il vero significato di "stato d'emergenza", utilizzando la chiara definizione trovata nell' Oxfort Poket Dictionary : *"Emergenza è un termine che descrive uno stato. È un termine gestionale e richiede di prendere delle decisioni e di effettuare dei follow-up rispetto a delle misure straordinarie"* [61]. Inoltre lo stato di emergenza deve essere dichiarato o imposto dalle autorità, definendo l'emergenza in termini di tempo e di spazio, le regole di coinvolgimento e la strategia di fronteggiamento.
Ora i microblogging e i social network, in particolare Twitter, vengono utilizzati da questi soggetti come supplemento alle forme di comunicazione tradizionale per la raccolta d'informazioni di qualsiasi genere, comprese le strategie d'azione. Negli ultimi anni, social media e microblog come Facebook, Twitter, Instagram e You Tube hanno dimostrato il loro valore durante le emergenze come veicoli d'informazione continua e real time, permettendo di condividere informazioni, sforzi umanitari, richieste di soccorso e consentendo di contattare le agenzie responsabili della gestione delle emergenze.
Inoltre l'uso di questi social media vanno ad interagire con il mondo del web,

senza prendere una posizione dominante nell'arena mediatica. Lo spicco di queste piattaforme social negli stati d'emergenza è stato accertato anche dall'ONU nel vademecum per l'emergenza, nel quale viene spiegato come utilizzare Facebook e Twitter nei casi di emergenza come calamità naturali o attentati terroristici [37].

Una piattaforma di questo genere, però, è anche uno strumento potenzialmente sfruttabile per comunicare e condividere esperienze e informazioni personali, quasiasi esse siano. É per questo motivo che si può considerare anche l'approccio all'emergenza dal punto di vista della vittima stessa.
Gli utenti del web quindi possono diventare dei "sensori sul territorio", poiché elaborano l'input, ossia lo stato di emergenza, creando come output la risposta singolare causata dall'input. Quindi creano un valore aggiunto all'emergenza in quanto monitorano con i loro contenuti e le loro informazioni condivise in maniera capillare il territorio in cui vivono.

Ci sono degli strumenti, oltre ai soliti messaggi o contenuti multimediali, direttamente utilizzabili sui social network e nei microblog per rendere la comunicazione e l'informazione più strutturata e semplice da estrarre:

- **Hashtag**: etichette precedute del simbolo cancelletto "#" fondamentali nella comunicazione su social network, sopratutto in situazioni d'emergenza. Vengono utilizzati per contrassegnare argomenti o eventi e inseriti all'interno dei proprio messaggi personali.
  In caso di emergenza, sono proprio coloro che hanno il compito di gestire e soccorrere che li creano per riuscire a compiere una ricerca delle informazioni in modo più veloce e semplice.

- **Retweet o Condivisioni**: termini riguardanti corrispettivamente Twitter e Facebook. Consistono nel ri-pubblicare qualcosa scritto da un altro utente per dargli più visibilità ed importanza. Anche questa funzionalità potrebbe in caso di emergenza essere una buona soluzione per velocizzare l'estrazione di informazione

- **Safety Check o Twitter Allert**: sistemi in grado di accertare se l'utente sia sopravvissuto o disperso all'evento d'emergenza. Esso viene attivato in automatico per le persone localizzate nel luogo della disgrazia

- **Geolocalizzazione**: la geolocalizzazione dei messaggi pubblici all'interno dei social network e dei microblog è molto utile, in particolar modo per le analisi delle situazioni da parte di enti atti al soccorso o al protezione. Questo strumento viene utilizzato poco fra coloro che utilizzano i social network , ma sarebbe in questi casi una risorsa molto utile [37].

Le due piattaforme social più adatte e utilizzate per questo tipo di situazione risultano Twitter e Facebook. Questa constatazione non è stata fatta senza coscienza, ma seguendo le statistiche sull'utilizzo più frequente fra i microblog e i social network. La ricerca di cui si parla si riferisce alla "Digital in 2016" fatta da "We are Social", già citata precedentemente [34]. Nonostante questa parità di importanza, il microblog più utilizzato già in situazioni di calamità naturali e in stato di emergenza è stato Twitter.

1.4 Twitter

Twitter è un servizio gratuito di social networking e microblogging, creato nel Ottobre 2006 da Biz Stone, Evan Williams, Noah Glass, Jack Dorsey e alcuni altri membri di Odeo, società del progetto "Twttr", idea primordiale di Twitter, uniti nella Obvious Corporation di San Francisco. Solamente nell'Aprile del 2007 Twitter diviene una società indipendente.
Questa piattaforma in poco tempo acquista grande popolarità e fama, incrementando sempre più, fino ad arrivare ad un attuale rallentamento di crescita, in quanto dal Settembre 2015 l'utenza è incrementata solo del 12,7%, registrando 320 milioni di utenti attivi al mese, al di sotto per esempio del grande incremento che interessa facebook [64].

Oltre ad essere un modo per comunicare ed un mezzo per l'estrazione d'informazione, Twitter risulta interessante e quindi più attrattivo anche per l'iscrizione al servizio di una grande quantità di personaggi famosi, aziende, società e associazioni di rilevanza mondiale.

Passando al punto di vista più tecnico, Twitter è una piattaforma che offre varie funzionalità, fra cui la possibilità di scrivere e pubblicare dei messaggi. Questi messaggi vengono definiti del gergo "tweet", che a differenza di altri social network come Facebook, essendo una forma di microblogging, possono essere di una lunghezza non superiore ai 140 caratteri per messaggio. Prima di iniziare a parlare e analizzare questo social network a livello di dati, è necessario sapere la meccanica di base che c'è sotto. Ogni utente registrato, possiede un profilo nel quale può inserire le sue informazione, che gli altri utenti, in base alle proprie impostazioni di privacy, possono vedere. Oltre ad un profilo, si può tessere una rete di connessioni, collegandosi al profilo di altri utenti: coloro al quale l'utente sceglie di connettersi vengono definiti "Following", mentre coloro che richiedono una connessione all'utente in questione sono definiti "Follower". In italiano, quando un utente fa una richiesta di connessione ad un altro utente si usa il verbo "seguire". Il termine connessione non è esattamente il più appropriato, in quanto in questo caso ha un accezione di relazione, ossia un utente può vedere tutti i contenuti dei propri following e, in caso di profilo pubblico, allora anche i profili di coloro che ancora non appartengono alla lista dei suoi following.
Nella timeline dell'utente quindi sono presenti solamente i propri following e per questo si può dire che Twitter non ha connessioni bidirezionali poiché la rete che sta alla base è una rete asimmetrica. Questo significa che, a differenza di Facebook che ha il concetto di "amicizia", Twitter possiede solo il concetto di "follower" e "following", che spesso possono non coincidere,
Dall'aprile del 2009, Twitter ha cambiato la sua interfaccia web, in quanto ha aggiunto alcune delle funzionalità che lo rende più "utile e funzionale". Viene aggiunta la barra di ricerca e un riassunto di temi di attualità, chiamato Trending Topics, ossia le notizie più frequenti e comuni che compaiono

nei tweet.
Un elemento molto attivo e rilevamente in twitter consiste negli "Hashtag", citato in precedenza. Queste etichette con parole o serie di parole senza spazi preceduto da un cancelletto creano un collegamento ipertestuale a tutti i tweet e i contenuti che contengono quel determinato hashtag. Nel 2010, avendo notato l'importanza dell'hashtag sia come elemento di gestione dei dati sia come facilitatore per le ricerche, hanno inserito nella prima pagina le "Tendenze", ossia gli hashtag più frequenti e qualche anno dopo son state personalizzate in base alla geolocalizzazione del profilo.
Inoltre un ultima funzionalità è quella dei "Preferiti". Col il termine "Preferito" intendiamo dare un approvazione al tweet scritto da un altro dei nostri follower. In concreto basta fare un click nel bottone a forma di cuore che si trova sotto al messaggio pubblicato. Questi "cuoricini" potrebbero sembrare un di più nel meccanismo di base di Twitter, ma risulta importante sia per l'approvazione e per misurare la visibilità di un contenuto, sia per l'impatto sulle relazioni sociali fra utenti [22] [41].

1.4.1 Contenuti dei tweet

Ogni minuto vengono pubblicati 278mila tweet da 140 caratteri ciascuno. Questi contenuti sono in un qualche modo soggetti ad una dicotomia riguardante il loro argomento: da un lato abbiamo quella che viene definita "l'intimità d'ambiente", ossia l'effetto provocato dalla condivisione di piccoli momenti della vita quotidiana e d'altro lato abbiamo informazioni e discussioni pubbliche su eventi o situazioni attuali nella società [45].
I tweet pubblicati possono essere ricondotti a 6 categorie: status personale, conversazioni, retweet, self-promotion, spam e news. Inoltre è da tenere in conto che all'interno dell'utenza media di Twitter troviamo anche aziende, società e associazioni, che invece hanno come scopo più grande quello promozionale e conoscitivo, visto come farsi conoscere dagli altri.

Essendo i tweet un insieme eterogeneo sia per argomento sia per finalità, il linguaggio utilizzato per i tweet cambia fra un tweet e l'altro e, un po'

come nella realtà, dipende dal contesto, dal contenuto e dall'utenza del tweet stesso. Senza dubbio l'uso di emoticon, di dialettismi, di modi di dire o di gif sono soliti di quei contenuti più personali e rivolti a tutti i proprio follower. Diverso è il linguaggio e il modo in cui un'azienda si pone verso i proprio potenziali clienti.

Capitolo 2

Emergenza in Analisi

L'Italia è un Paese ad elevata sismicità e come afferma anche la Protezione civile viene definito addirittura "uno dei Paesi a maggiore rischio sismico del Mediterraneo, per la sua particolare posizione geografica". L'Italia infatti è situata al margine di convergenza tra due grandi placche, quella africana e quella euroasiatica, che dividono esattamente il Paese in corrispondenza degli appennini, sopratutto nella parte centro-meridionale [11].
Una maggior concentrazione di rischio si concentra nella parte centro meridionale, esclusa la Sardegna, la quale non risente di particolari eventi sismici.

In questo capitolo, introdurremo gli eventi sismici avvenuti quest'anno nel Centro Italia. Ci concentreremo, in particolar modo, sull'oggetto della nostra analisi: la prima forte scossa di terremoto avvenuta il 24 Agosto 2016. Oltre alla descrizione e ai dati riguardanti questo evento, andremo a discutere sugli effetti del sisma e sulla reazione degli utenti sui social media.
Infine esporremo gli obiettivi e i contenuti dell'analisi che verrà poi sviluppata nel capitolo seguente.

2.1 Eventi Sismici Centro Italia 2016

Ad oggi, il Centro Italia è divenuto teatro di scosse sismiche non indifferenti. Con la denominazione "Eventi Sismici Centro Italia 2016" ci si riferisce alla

sequenza di terremoti iniziati il 24 Agosto 2016 con epicentro situato lungo la Valle del Tronto, tra i comuni di Accumoli e di Arquata. Per ora, il sisma più forte avvenuto in questa serie di eventi è quello del 30 Ottobre 2016 con epicentro nei comuni di Norcia e Preci, in provincia di Perugia [31].

Dal primo evento sismico in questione, l'INGV [24], l'Isituto Nazionale della Geologia e Vulcanologia, effettua dei periodici aggiornamenti per quanto riguarda la situazione del Centro Italia. Attualmente, l'aggiornamento più recente è quello in data 6 Novembre 2016 alle ore 17 [32]. L'aggiornamento citato prende in considerazione due punti di vista, basati sugli eventi più rilevanti sopracitati: la prima scossa della sequenza sismica e quello più potente. Queste rilevazioni sono localizzate dalla Rete Sismica Nazionale dell'INGV, che consiste nell'installazione di circa 350 stazioni sismiche su tutto il territorio nazionale. Queste stazioni sono postazioni fisse, dotate di strumenti che rilevano ogni minimo movimento del suolo [28].

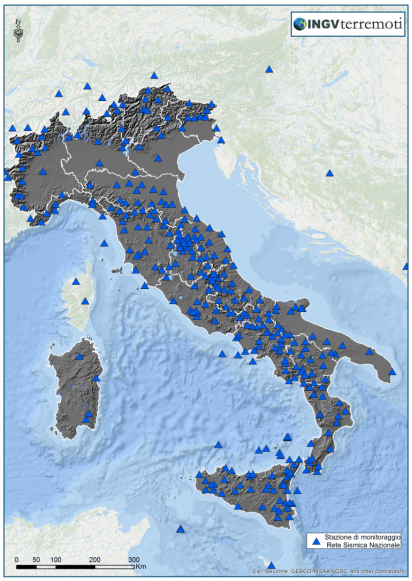


Figura 2.1: Rete Sismica Nazionale gestita dall'INGV.- Istituto Nazionale della Geologia e Vulcanologia [28]

A partire dal primo evento del 24 Agosto 2016, come si nota nella figura 2.2, si sono registrati un numero complessivo di scosse pari a circa 23.900.

Alle ore 17:00 di oggi, 6 novembre, sono circa 682 i terremoti di magnitudo compresa tra 3 e 4, 41 quelli di magnitudo compresa tra 4 e 5 e 5 quelli di magnitudo maggiore o uguale a 5 [32].

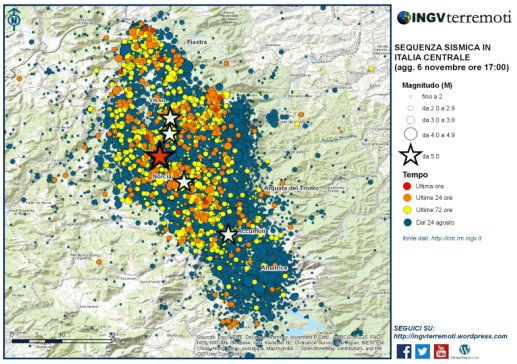


Figura 2.2: Aggiornamento 6 Novembre 2016 ore 17.00 dal 24 Agosto 2016 - INGV [32]

Prendendo in analisi invece gli eventi sismici dal 30 Ottobre 2016 la situazione a livello nazionale è quella che si può vedere nella figura 2.3 che segue.

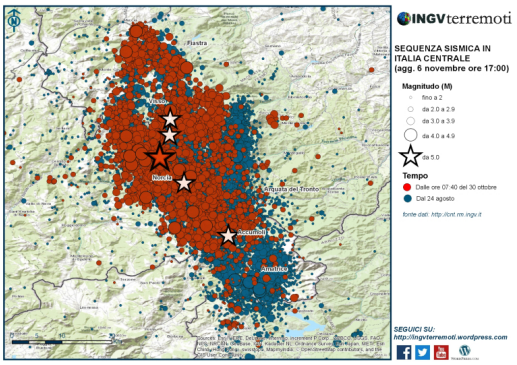


Figura 2.3: Aggiornamento 6 Novembre 2016 ore 17.00 dal 30 Ottobre 2016 - INGV [32]

La parte rossa raffigura l'andamento degli eventi sismici avvenuti dopo quello del 30 ottobre 2016, e questi sono complessivamente oltre 300, fra cui

circa 335 i terremoti di magnitudo compresa tra 3 e 4 e 20 quelli di magnitudo compresa tra 4 e 5.

2.2 La prima forte scossa: 24 Agosto 2016

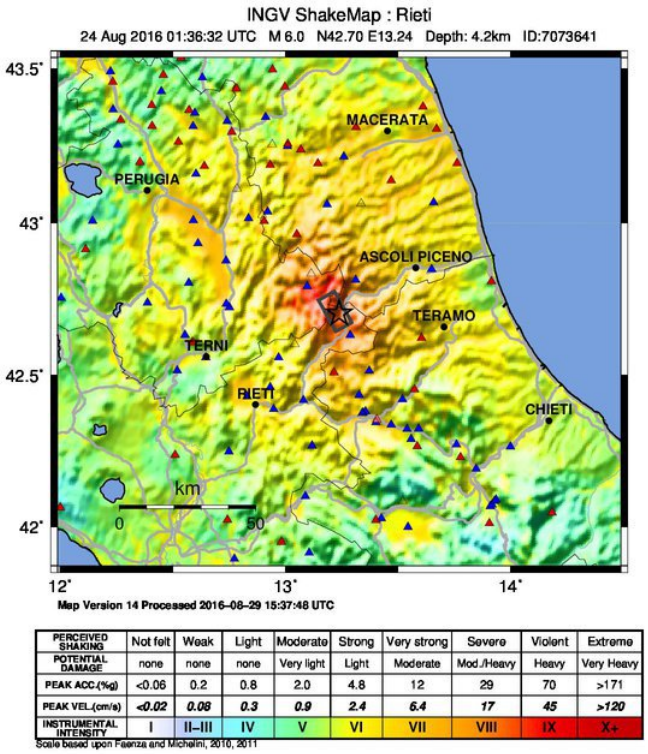
La prima di questa serie di eventi sismici è avvenuta il 24 Agosto 2016 alle ore 03:36 italiane fra le provincie di Rieti ed Ascoli Piceno. L'epicentro del terremoto è localizzato nel comune di Accumoli, in particolare con una latitudine di 42.70N e una longitudine pari a 13.23E; mentre l'ipocentro è a 4 chilometri di profondità.

Secondo le misure date dall'INGV l'evento principale e quindi la magnitudo associata al terremoto è una magnitudo di 6.0 con un incertezza di 0.3. Infine l'INGV, utilizzando i modelli globali ed un'altra tecnica di analisi dei dati ottiene un valore di magnitudo pari a 6.2. [26].
I comuni entro 10 km dall'epicentro sono Arquata del Tronto (AP) e Accumoli(RI), ma ci sono anche molti altri comuni che son stati toccati in modo significativo, nonostante una maggior distanza dall'epicentro e sono: Amatrice (RI), Cittareale(RI), Norcia(PG), Acquasanta Terme(AP), Cascia(PG), Montegallo(AP), Montereale(AQ), Campotosto(AQ), Capitignano(AQ), Castelsantangelo Sul Nera(MC), Valle Castellana(TE), Posta(RI), Borbona(RI), Monteleone Di Spoleto(PG), Montemonaco(AP), Poggiodomo(PG), Preci(PG), Rocca Santa Maria(TE), Cortino(TE), Leonessa(RI), Roccafluvione(AP), Ussita(MC), Visso(MC).

2.2.1 Misurazione ed effetti del sisma

Fino ad ora si è presa in considerazione una delle varie misurazioni possibile, ossia quella attraverso la scala Richter.
Tale scale valuta l'intensità obiettiva del terremoto, secondo la quantità di energia liberata. Non ha livelli o gradi, ma fa riferimento solamente ad indici, chiamati magnitudo, che è esattamente un rapporto logaritmico fra l'ampiezza massima di una scossa e il logaritmo di una scossa campione. Il

magnitudo va in una scala da 0 a 8.2 che è il massimo registrato fino ad ora ed esso è corrispettivamente anche il grado di gravità. Fino a un magnitudo di 3.5 la scossa viene registrata ma non sentita e gli effetti del sisma vanno incrementando fino ad 8.2, che rappresenta un terremoto che può causare danni veri su vaste aree fino a svariate centinaia di chilometri dall'epicentro. Oltre alla scala Richter, esiste la scala Mercalli, che invece valuta l'intensità del terremoto sulla base di gradi che rappresentano gli effetti sull'ambiente.



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Mod./Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	<0.06	0.2	0.8	2.0	4.8	12	29	70	>171
PEAK VEL.(cm/s)	*<0.02*	*0.08*	*0.3*	*0.9*	*2.4*	*6.4*	*17*	*45*	*>120*
INSTRUMENTAL INTENSITY	I	II–III	IV	V	VI	VII	VIII	IX	X+

Scale based upon Faenza and Michelini, 2010, 2011

Figura 2.4: ShakeMap del 24 Agosto 2016 - INGV [1]

Le due tipologie di valutazioni sopra descritte possono avere correlazioni, ma non sempre un alta magnitudo va ad indicare un alto grado nella scala Mercalli. Si confronti, per esempio, un terremoto ad altissimo magnitudo localizzato in un deserto e uno di media magnitudo verificatosi invece in un zona densamente abitata, anche dove ci sono molte costruzioni. Con la

scala Mercalli, rispettivamente il primo terremoto avrà un' intensità minore di quello avuto per secondo nonostante la magnitudo. È infatti opportuno considerare entrambe le misurazioni, sopratutto per verificare anche gli effetti che il territorio ha sulle persone e sull'ambiente.

Prendiamo quindi ora in considerazione le mappe di scuotimento, o ShakeMap, che forniscono una visualizzazione dei risultati della scala Mercalli. La figura 2.4 è già esplicativa considerando la legenda posta sotto l'immagine. È da precisare che tale mappa adotta la scala di colore per le intensità tramite la dicitura WEAK-STRONG-SEVERE (debole-forte-severo): strong con il verde e giallo in cui la scossa si avverte molto distintamente e forte con danni lievi, severe con il colore che varia dall'arancio al rosso intenso e qui lo scuotimento risulta assai potente da fare gravi danni e infine weak dal bianco all'azzurro, che è la situazione più favorevole in cui viene avvertito ma non ci sono danni. Possiamo infatti notare come nell'epicentro, rappresentato dalla stella in figura, il grado della scala utilizzata sia più alto. Inoltre i triangoli rossi indicano le stazioni accelerometriche e velocimetriche dell'INGV e quelli blu le stazioni accelerometriche del Dipartimento della Protezione Civile [1].

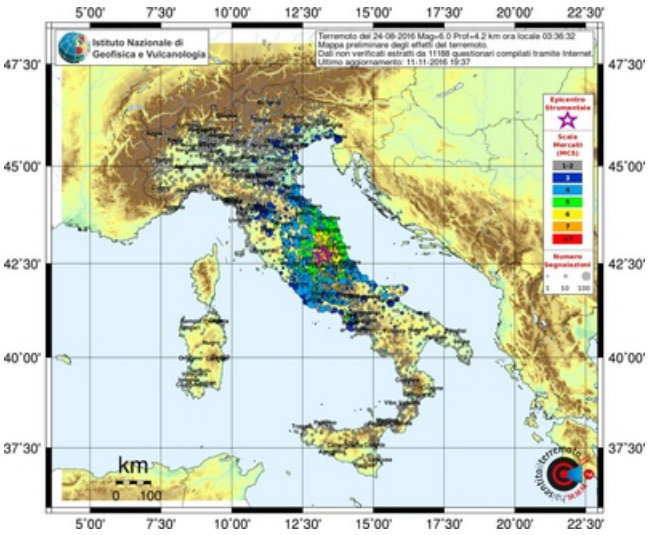


Figura 2.5: ShakeMap dai questionari del web - INGV [25]

Infine è alquanto significativo tenere in considerazione anche le mappe di

scuotimento create in base a questionari fatti sul web e non sono in base alla velocità e accelerazione come in precedenza.
Nella figura 2.5 vediamo infatti l'intensità di questo terremoto, vista con gli occhi delle vittime dello stesso. Questa mappa viene aggiornata ogni volta che gli utenti compilano il questionario. La dimensione dei cerchi in figura sono proporzionali al numero di questionari fatti per ciascun comune, mentre i colori vanno in base all'intensità di come è stato sentito il terremoto [25].

2.2.2 Reazioni sui Social

Il web è diventato un canale universale di comunicazione principale per qualsiasi evento.
Alcune persone davano il proprio sostegno concreto e il proprio aiuto sui luoghi del sisma, altre invece provavano a contribuire dietro ad un computer. Le modalità di intervento possono essere tante e svariate: da chi condivide informazioni necessarie, come i numeri della Protezione Civile, chi invita a tenere la linea Wi-Fi aperta, chi va elenchi di luoghi per donare e chi avvisa de bisogno di beni di prima necessità. Fra tutti gli utenti che pubblicano contenuti riguardanti il terremoto avvenuto, non tutti lo fanno per dare un aiuto, o meglio non per un aiuto diretto. È facile notare come molti contenuti pubblicati siano pubblicati per pregare per le vittime, per esprimere la propria empatia o semplicemente per fare polemiche e per mettersi in mostra ed avere visibilità.

Nonostante questa parte dei persone appena descritte, si è notato come anche i personaggi famosi o addirittura anche all'estero, come in Russia, abbiano sentito l'esigenza di contribuire per gli eventi sismici in questione. Questo dimostra come il web riesce anche a unire più parti del mondo per intervenire ad un emergenza di questo livello.
Come ci informa "il Giornale.it",anche i vip stanno aiutando la Croce Rossa e la Protezione Civile al fine di essere d'aiuto alle popolazioni terremotate. Cita molti nomi di personaggi famosi e non solo di nazionalità Italiana "i vip non restano insensibili di fronte alla tragedia: in tanti stanno infatti

aiutando a condividere gli appelli di Croce Rossa e Protezione Civile, al fine di essere d'aiuto alle popolazioni terremotate. Da Belen Rodriguez a Stefano De Martino, da Simona Ventura a Elisa, fino ad Alessandra Amoroso e i Negramaro, i personaggi del mondo dello spettacolo non possono fare a meno di esprimere vicinanza alle persone colpite dal disastro" [16].

Sui social network, anche in Russia si parla del terremoto che ha colpito l'Italia Centrale. Dalla Russia infatti vengono inviati messaggi di solidarietà e di incoraggiamento, persino con l'account dell'Ambasciata della Federazione Russa. I mittenti non sono solamente istituzioni o persone di rilievo, come il capo del commissione estera Duma Pushkov e il presidente Putin, ma anche persone comuni hanno lasciato grandi messaggi di solidarietà.
Inoltre hanno anche attivato un numero d'emergenza per i cittadini russi che si trovassero nelle zone compite dal sisma, così da riuscire ad aiutare anche a livello più concreto [33].

Purtroppo il web, in particolar modo i social network, non sono totalmente un ambiente pieno di buon senso. C'è stata una strumentalizzazione dell'accaduto per fini esclusivamente politici, fino ad vere persino connotati razzisti.
Altre manifestazioni di questo genere sono i contenuti pubblicati dai selfie-addicted e le "mezze celebrità", definite così nell'articolo del "il Giornale.it". Questi personaggi pubblicano selfie o un posato con i prodotti del proprio brand di abbigliamento spacciandolo per una preghiera per il terremoto [16].

2.3 Obiettivo

Quest'anno l'Italia è stato uno scenario disastroso per quanto riguarda gli eventi sismici avvenuti nel Centro Italia, come precedentemente già successo nel sisma del 2012 e prima ancora in quello del 2009 all'Aquila.
Queste scosse e questi terremoti divengono l'incubo di tutti i cittadini del Paese, in particolare per coloro che vivono nelle zone più ad alto rischio sismico,

come si nota nella figura 2.6, ricavata dall'INGV tenendo in considerazione anche tutti terremoti avvenuti nella storia.

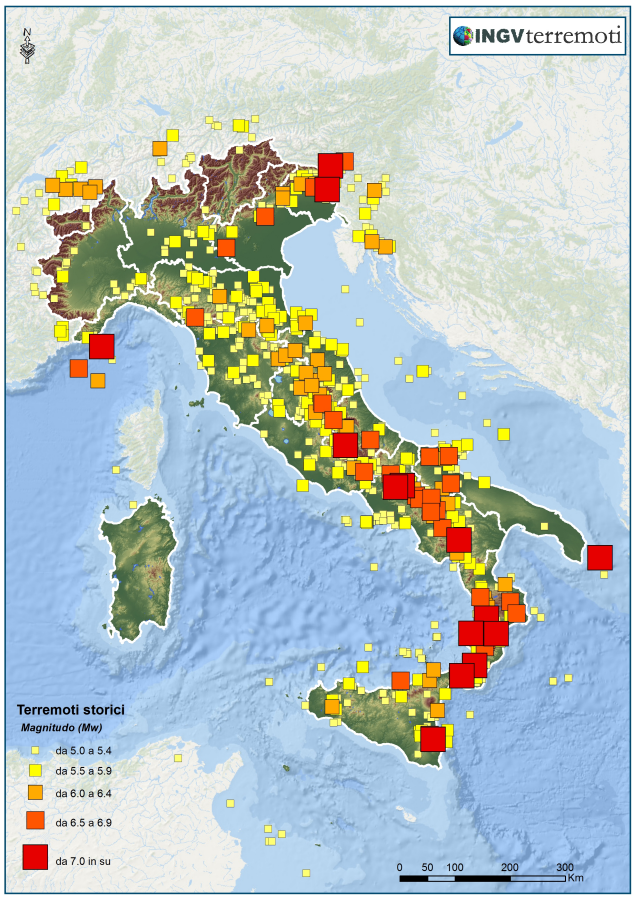


Figura 2.6: Mappa delle zone sismiche - INGV [27]

Oramai il web, in particolare i social network, sono gli strumenti più utilizzati in tutte le situazioni o eventi, nel limite della possibilità di accedervi, un altro problema su cui si è visto adoperarsi a livello nazionale nei casi d'emergenza.

Come si è potuto evincere dal paragrafo 2.2.2, i social network acquisiscono una grande importanza, ma, d'altro canto, si deduce molto bene i rischi del caso, tanto che non tutti contenuti che si leggono sono da prendere per buoni ed utili.

Queste piattaforme quindi sono un insieme eterogeneo di contenuti e l'importante per una buona analisi è saper filtrare i contenuti in modo da non tener in considerazione ciò che è pura polemica o ego personale dell'autore del post pubblicato.

L'analisi quindi riguarda essenzialmente l'evento sismico sopra descritto avvenuto il 24 Agosto 2016. L'evento preso in considerazione iniziale doveva essere il sisma avvenuto nel 2012 in Emilia, ma si è optato per rendere la tesi più attuale possibile, visto che ad oggi le scosse continuano in modo incessante. Infine l'importante è che l'evento preso in considerazione sia attuale e abbastanza significativo e d'impatto per il Paese.
In primo luogo è rilevante specificare quale dei social network è stato preso in considerazione. Questa scelta è stata pensata e dovuta la fatto che questo social network, o meglio microblogging, sia più "serio" rispetto, per esempio, Facebook, l'altro social network più utilizzato in Italia. Twitter infatti viene utilizzato moltissimo anche dalle istituzioni e personaggi rilevanti, poiché c'è meno dispersione che negli altri social network, nonostante le funzioni sono pressoché identiche. Inoltre un altro vantaggio dell'utilizzo di Twitter è la possibilità di comporre dei tweet di soli 140 caratteri, questo porta l'autore del tweet a focalizzarsi in modo più mirato sull'obiettivo del contenuto da condividere, risultando quindi coinciso e diretto.
Ora passiamo al cuore dell'analisi, l'obiettivo principale per questa tesi è riuscire a comprendere se Twitter possa essere d'aiuto in una situazione d'emergenza sismica. L'utilità di questo si è già vista ampiamente, ma l'obiettivo si focalizza sull'utilizzo di Twitter come "strumento operativo". Quindi è interessante capire se è possibile ritrovare informazioni sull'emergenza. Queste informazioni di cui si parla però non sono quelle riguardanti i dati ufficiali, le donazioni o quelli delle istituzioni per le manovre d'azione, ma bensì consistono in richieste d'aiuto o effettivi e concreti effetti del terremoto ricavati direttamente dal profilo dell'utente vittima del disastro.

Infine è sembrato utile fare un'analisi che anche l'utente comune possa capire e possa farlo in modo autonomo attraverso un applicazione online.

L'applicazione però deve mantenere delle buone caratteristiche di utilizzabilità e accessibilità, in quanto chiunque possieda un data-set di tweet possa utilizzare e non solo per eventi riguardanti i terremoti, ma qualsiasi evento in genere in cui si vuole fare la propria analisi.
Quest'ultimo punto è molto importante poiché non ha senso fare una ricerca su uno stato d'emergenza, senza dare la possibilità a tutti i cittadini di poterne usufruire.

2.3.1 Contenuti dell'analisi

L'analisi del data-set raccolto, contenete i tweet in relazione all'emergenza considerata, è un flusso unico in cui vengono fatte analisi testuali e considerazioni sui dati a disposizione.

In primo luogo siamo partirti con l'estrazione del data-set, come descritto in dettaglio nella sezione 3.1. A questo punto si è effettuata un'analisi del testo, pulendo il testo da qualsiasi forma non trattabile. In seguito abbiamo utilizzato ciò per vedere qual erano le parole più utilizzate in questi tweet in modo da avere un'idea sulla tipologia della maggioranza dei contenuti pubblicati su questo argomento.

Successivamente si è passati alla visualizzazione delle parole più frequenti e a filtrare i dati in base all'utilità dell'informazione. Infine si è voluta fare un'analisi statistica sul luogo e il momento in cui i tweet sono stati pubblicati e scritti, sfruttando quindi la geo-localizzazione dei tweet, così da avere un idea dell'andamento temporale e localizzato dell'evento sismico sui social network.

Capitolo 3

Design dell'Analisi

In questo capitolo andremo a descrivere le modalità con le quali si sono effettuate le analisi sul data-set estratto da Twitter e gli strumenti utilizzati. L'analisi è stata effettuata seguendo in particolar modo gli obiettivi descritti nel capitolo precedente. Per comprendere appieno l'analisi fatta, il capitolo inizia focalizzandosi su quello che è il mondo dell'estrazione dei dati nel social network in questione: Twitter.

3.1 Estrazione dei Dati

Twitter fornisce, per gli sviluppatori privati o facenti parte di organizzazioni, la possibilità di realizzare applicazioni che interagiscano con esso stesso, permettendo così di estrarre contenuti relativi a particolari utenti o a parole chiave decise dall'autore dell'applicazione. Tutto questo è possibile poiché poiché Twitter mette a disposizioni delle API, con questo scopo [59].
Le API, quando, come in questo caso, sono usate nel contesto Web, sono definite come un insieme di richieste HTTP, che restituiscono come messaggio di risposta, un contenuto ben strutturato, che utilizza solitamente formati, o definiti linguaggi di markup, come XML o JSON. Attualmente viene usato un paradigma orientato ad un approccio più diretto alla rappresentazione dello stato di trasferimento, finalizzato quindi ad un'applicazione di rete basata

su un protocollo di comunicazione: vengono quindi definite restful API, abbreviato in REST API. Un concetto alquanto importante si ritrova nel fatto che il protocollo utilizzato è indipendente dall'architettura dell'applicazione, perché con essa si interfaccia, ma non si identifica.
Uno dei principi primari delle REST API consiste nel fatto che devono essere identificate con un URI specifico e univoco, contenuto nel corpo o nell'intestazione della risposta. Un altro aspetto importante consiste nell'astrazione, che evita al programmatore di dover utilizzare le API ad un livello più basso, e anche nel fatto che è un sistema livellato in quanto un client non può e non deve sapere a quale server è connesso.

Twitter, per l'utilizzo di queste API, offre la possibilità di utilizzarle anche senza la necessità di autenticarsi in qualche modo e di conseguenza registrarsi al sito "Twitter Developers" [60]. Ovviamente però l'utilizzo dei metodi di autenticazione messi a disposizione, danno allo sviluppatore sicuramente qualche vantaggio. Senza alcun tipo di registrazione infatti si riescono a fare solamente 150 richieste l'ora, anziché 350 richieste l'ora in seguito all'autenticazione.
Twitter mette a disposizione due metodi di autenticazione:

- Application-user authentication;
- Application-only authentication.

Nel primo caso si fa uso del protocollo OAuth 1.0a la richiesta HTTP avviene per conto dell'utente e quindi deve comunicare da quale applicazione si sta effettuando la richiesta e da quale utente. L' application-only authentication viene fatta per conto dell'applicazione, tramite l'utilizzo della sua consumer key e secret key. Viene quindi effettuata una richiesta HTTP POST T all'endpoint oauth2/token per scambiare le credenziali citate. In questo caso si seguono le istruzioni del protocollo OAuth 2.0.
Fra le due modalità la scelta è caduta sul "Application-only authentication", in quanto risulta meno complessa e non esiste alcun tipo di contesto relativo all'utente e non è richiesto che la richiesta HTTP sia firmata. Questo metodo

di autorizzazione però manda all'end point richieste di tipo SSL, usando quindi il protocollo HTTPS.

Fra le REST API, troviamo più tipologie che vengono messe a disposizione. La "Twitter Search API" che ne è parte integrante e permette l'esecuzione di query in tempo reale, però con dei limiti sull'estrazione dei tweet, in quanto è possibile scaricare solamente quelli più recenti. Inoltre abbiamo anche le "Streaming API" che invece offrono funzionalità per accedere in tempo reale al flusso globale di tweet, ma a differenza delle altre API queste richiedono una connessione HTTP persistente, che va così ad impattare tutta l'applicazione.
Twitter estende queste API in librerie finalizzate ai diversi linguaggi, quali ad esempio C, C++, Java, PHP. Per quanto riguarda la mia tesi il data-set utilizzato è stato quello ricavato da un applicazione creata durante un tirocinio fatto precedentemente. Considerando le limitazioni temporali dell'estrazione dei tweet, consistenti nell'utilizzo di tweet di un massimo di sette giorni addietro, la scelta, nonostante le sperimentazioni con le streaming API e le search API, è caduta su un progetto creato attraverso delle librerie java per delle pure richieste HTTP, senza l'utilizzo di API specifiche per estrazioni di dati da Twitter.

3.1.1 Tipologia di dataset

La realtà studiata comprende una sola entità principale: i tweet. La memorizzazione di questi dati è stata fatta seguendo in modo abbastanza fedele la struttura del formato json della risposta. Di conseguenza è necessario che i dati vengano posti in formati altrettanto strutturati, che generalmente sono o database relazionali o file tabulari, predisposti principalmente a trasportare informazioni di entità.
Per questo, la scelta della tecnologia usata ricade in file di tipo "Comma-separated values", con l'estensione .csv. Questo formato infatti viene utilizzato in particolare per l'importazione ed esportazione dei dati in una tabella. Ogni tabella è composta da più righe, le quali rappresentano i record della

base di dati, solitamente visualizzati da una linea di testo. Le righe della tabella a loro volta sono divise in campi, che ne rappresentano invece le colonne. Ogni campo inoltre viene diviso da un carattere separatore particolare, deciso dall'autore della base di dati. Il formato in questione non specifica una codifica di caratteri, ma è l'utilizzatore che deve specificare le varie caratteristiche al momento dell'importazione o esportazione in formato csv dei dati in un programma.
L'entità che raffigura i tweet salvati nella base di dati è composta dai seguenti campi:

- username;
- datatweet;
- retweets;
- txt;
- geo;
- hashtag;
- link.

Il contenuto dei vari campi si può dedurre facilmente dalla denominazione degli stessi.
La motivazione dell'utilizzo di questo formato non è solo per la presenza di un entità unica e la struttura ben definita di questa entità, ma anche per un motivo di successiva implementazione del software per analizzare i testi del tweet.

In un secondo momento si è deciso di utilizzare anche una base di dati sempre strutturata, ossia un database relazionale. Le motivazioni di questa decisione sono varie, anche se non altamente efficiente ma comunque valide, in quanto si necessitava una base che ospitasse più tabelle, come deposito e in modo da poterle gestire contemporaneamente. Questo per il fatto che

dal data-set iniziale si sono filtrati i tweet considerando varie caratteristiche, quali una categorizzazione in base le parole chiavi presenti, in particolari si sono divisi i tweet che riguardavano in modo particolare l'ambito delle donazioni e i tweet che esprimessero solidarietà verso le vittime del terremoto. La parola "deposito" non è stata utilizzata a caso, bensì è stata voluta poiché le tabelle inserite all'interno del database sono scollegate fra loro, quindi si nota l'assenza di associazioni fra le tabelle. A differenza del formato csv, si è aggiunto in ogni tabella un id con la funzione di chiave primaria, che consiste in un numero sequenziale per rendere univoco i record inseriti. Purtroppo non è stato possibile utilizzare uno degli attributi facente parte del data-set iniziale perché l'unica ipotesi plausibile per una chiave primaria poteva essere l'username, essendo univoco per un ogni account, ma bisogna considerare la possibilità che una persona possa scrivere più tweet nell'arco di tempo in questione.
In particolare si è utilizzato un database MySQL, attraverso phpMyAdmin, uno strumento software gratuito scritto in PHP, destinato a gestire l'amministrazione di un database di questo tipo. Ad oggi quest'utilizzo di questa piattaforma è molto diffuso, in quanto mette a disposizione agli utenti una vastità di funzioni per riuscire globalmente a gestire i dati, compresa anche una shell adibita alle interrogazioni sul database.

3.1.2 Parametri di ricerca

A monte dell'estrazione dei dati, avvenuta nelle modalità sopra ampiamente descritte, è stato necessario eseguire un breve studio sui messaggi utilizzati dagli utenti dei social network per annunciare l'evento. L'analisi di questi messaggi ci ha permesso di ricavare un insieme di parole chiave, necessarie per delineare l'evento in questione. Infine i parametri utilizzati, integrati nelle richieste HTTP sono i seguenti:

- **querysearch**: terremoto, emergenza, amatriciana, scossa, scosse, aiuto, terra, trema, norcia, #amatricia, centro italia, ascoli piceno, #ter-

remoto, #terremotocentroitalia, #emergenza, protezione civile, ingv, #terremotoItalia, #ItalyEarthquake, crollo, crollata, crollato;

- **since** : 2016-08-24;
- **lang** : ita.

3.2 Tecnologie utilizzate

L'importanza dell'analisi testuale sta incrementando, visto la continua crescita di applicazioni web. Mentre le classiche applicazioni web si focalizzavano su un "processing and mining raw text", l'avvento del Web 2.0 e la conseguente diffusione di dati non strutturati ha portato la necessità di nuovi metodi per il text mining, che non utilizzassero i dati così com'erano, ma che li modificassero affinché possano riuscire a compiere un'analisi completa e un confronto fra i dati preciso e dettagliato. Per questo uno degli aspetti principali del text mining è il "text cleaning", consistente in un processo di pulizia dei dati prima di iniziare la vera e propria analisi testuale. In questo ambito, uno dei linguaggi e uno degli ambienti di lavoro più potenti e diffusi in particolare nell'ultimo periodo, è il "software R", che in seguito andremo ad illustrare [10].

Inoltre è stato necessario avere anche il supporto di uno strumento in grado di gestire i dati e per questo ci si è indirizzati verso un database relazionale di tipo MySQL. Il MySQL è un software che implementa le funzioni per utilizzare una base di dati relazionale. Per l'uso di questi software esiste un linguaggio di programmazione chiamato SQL, Short Query Language, che utilizza un linguaggio simile a quello naturale per delineare le operazioni da effettuare sui dati e sul database.

3.2.1 Software R

R è definito un linguaggio di programmazione specifico per l'analisi statistica. La denominazione di "linguaggio di programmazione" non è completamente

appropriata, in quanto esso è un ambiente completo di sviluppo per il calcolo statistico e matematico [67].
Il nome deriva dai creatori di R, i quali sono Ross Ihaka e Robert Gentelman, appartimenti al Dipartimento di Statistica dell'Università di Auckland in Nuova Zelanda. R viene considerato una moderna interpretazione di un più vecchio linguaggio di programmazione, chiamato S, il quale riguardava anch'esso l'analisi statistica e grafica. Infine si può affermare che l'ambiente R è una valida alternativa Open Source all'ambiente S, in quanto è offerto come un software libero sotto i termini della General Public Lincense, GNU, della Free Software Foundation. Di conseguenza questa licenza prevede che il software venga fornito sia nel formato compilato e quindi eseguibile da una console, sia in quello sorgente [17].

Il sistema R non è solamente un ambiente di sviluppo statistico, bensì offre un insiemi di funzioni, librerie e package che permettono la gestione, l'elaborazione e l'analisi dei dati, oltre che per la creazione di grafici ed immagini. Questo tool software offre un proprio linguaggio di scripting, utilizzabile tramite una shell per UNIX e una console apposita, definita "R-console" per Windows. in cui è permesso eseguire le operazioni e le funzionalità di nostro interesse con l'uso di specifiche librerie, distribuite in package da installare sulla propria area di lavoro. Questo aspetto di istallazione dei pacchetti messi a disposizione dà un vantaggio all'ambiente di sviluppo rendendolo estendile. I package utilizzati sono distribuiti con la licenza GPL e organizzati in un apposito sito, denominato Comprehensive R Archive Network, CRAN.
La potenza di R la troviamo in primo luogo nella disponibilità di risorse grafiche che permettono un utilizzo del sistema più fluido e intuitivo, ma anche nelle altre sue caratteristiche principali che rivediamo in un'efficace manipolazione e memorizzazione di dati sia testuali sia numerici e in un vero e proprio linguaggio di scripting dotato delle istruzioni di controllo, tipiche di qualsiasi linguaggio di programmazione. Inoltre sono presenti anche molti operatori che permettono di effettuare in modo immediato calcoli su vettori e matrici, visto l'orientamento statistico e matematico del software.

Infine posso dire che la scelta dell'uso di R come ambiente di sviluppo è stata fatta per la diffusione che questo tool ha nell'ambito sopratutto del trattamento delle problematiche del Natural Language Processing, NPL, e del text mining [51].

3.2.2 MySQL

Come descritto nel paragrafo 3.1, si è scelto di memorizzare i dati all'interno di un database. Per i motivi sopra specificati si è scelto di utilizzare MySQL, ossia un RDBMS open source disponibile sia per ambienti UNIX sia per quelli windows.
MySQL supporta la sintassi SQL, che è il linguaggio che abbiamo utilizzato per l'interrogazione al database creato al fine dell'analisi in questione. SQL è oggi considerato in tutto e per tutto uno standard. Infatti il fatto di avere uno standard definito per un linguaggio per database relazionali, apre la strada a quella che è l'intercomunicabilità fra tutti gli elementi che si basano su questo.

L'interazione al database quindi avviene inviando istruzioni SQL al DBMS e ciò può avvenire in due modi:

- invocazione interattiva;
- invocazione tramite un programma applicativo.

L'invocazione interattiva viene utilizzato un programma con lo scopo di ricevere in input istruzioni SQL e inviare un risultato all'utente. Questa è la tipologia che si è utilizzata per compiere un'analisi, poiché attraverso la "shell" integrata in phpMyAdmin abbiamo interrogato il database affinché ci facesse visualizzare l risposta.
L'invocazione tramite un programma applicativo significa le istruzioni SQL sono eseguite nel corso dell'esecuzione del programma in questione e i risultati vengono utilizzati dal programma per produrre il proprio output personalizzato. Quest'ultimo metodo di invocazione è stato utilizzato invece per

l'estrazione dei dai e per inserire immediatamente anche nel database adibito i dati, oltre che creare un data-set anche in formato csv. In questo caso specifico, si è scelto di l'Embedded SQL, inglobando nel programma codice SQL, senza alcun tipo di metodologia di accesso, quali per esempio Object Relational Mapping, ORM.

3.3 Iter progettuale

Definiti in linea generale le tecnologie utilizzate, entriamo nel cuore dell'analisi e andiamo passo per passo a vedere come si è implementata e quale siano i risultati e le considerazioni ottenuti, dallo studio di questi principali aspetti. Il codice mostrato nel capitolo è stato eseguito su due macchine con due processore molto differenti:uno con un processore INTEL PENTIUM con 4GB di RAM e l'altro con un processione INTEL CORE I7 con 16GB di RAM.

Nel progetto di tesi andiamo analizzare più aspetti degli stessi tweet: da un lato il contenuto stesso del tweet e d'altro lato il luogo e la data del tweet. Per la fase di analisi abbiamo utilizzato un data-set di 435 tweet, questo numero così limitato deriva dal fatto che il nostro interesse era volta ad analizzare i tweet in prossimità dell'emergenza presa in considerazione, visto anche gli obiettivi esplicitati nel capitolo 2. Inoltre più si va avanti con il tempo dalla data del 24 Agosto 2016, più i tweet riguardano atti di solidarietà, di donazioni e di informazioni sui dati ufficiali dell'emergenza.

3.3.1 Configurazione di R e caricamento dei dati

Il primo approccio al progetto di tesi consiste nella vera e propria analisi dei contenuti con R, il software sopra delineato che possiede anche specifiche funzioni per l'analisi dei dati, attraverso un proprio linguaggio di scripting.

La prima fase consiste nell'esecuzione di R e caricamento dei dati nel software, Avendo lavorando con Windows, ho utilizzato l'apposita GUI creata per questo sistema operativo. La GUI è composta da una finestra contente un'interprete dei comandi.

Prima di iniziare con l'effettiva implementazione del codice è stato necessario preparare e salvare quella che il software chiama area di lavoro, che altro non è che una console in cui scrivere il codice e su cui vengono installati i package delle librerie necessarie. I package sono essere installati con due modalità attraverso la barra superiore avente una funzione apposta o altrimenti tramite il comando:

```
install.package("packageName")
```

Essendo però un pacchetto aggiuntivo, in ogni caso è necessario dichiarare esplicitamente l'utilizzo di una determinata libreria, così da attivarla preventivamente all'uso immediato.

```
library("libraryName")
```

In seguito si andrà anche a determinare i package e le librerie attivate per ogni blocco di istruzioni. Inoltre ci sono alcuni package come R.utils, R.graphics, R.datasets, R.base, R.stats, R.methods e R.grDevices che vengono caricati quando R viene attivato [56].
Il primo package utilizzato per compiere questo primo step è "R.base", un pacchetto precaricato da R che fornisce classi e metodi utili per la gestione del caricamento dei dati, in particolare viene usata la libreria "data.frame". In questo caso abbiamo semplicemente importato i dati all'interno del software dal file csv contenente i dati in questione.

```
#carico la colonna tweet dal file .csv
tweet <- read.csv( "/R/terremotocentroitaliatw.csv",
                   header = TRUE, sep =";", quote=";")
```

Listing 3.1: Caricamento dei dati

La funzione `read.csv();` ha una serie di input opzionali che servono a spiegare al sistema come leggere il data-set passato in input, che nel nostro caso è in formato csv ed è questa la scelta di questa funzione, oltre ad una maggior sicurezza in confronto all'utilizzo di `read.table("");` che risulta più a

rischio di errori a livello di esecuzione, ma che in pratica possiede la stessa funzionalità e la stessa sintassi. Fra gli input opzionali si è scelto di utilizzare:

- **file**: l'input principale dove si inserisce il path del proprio file, che può essere all'interno della directory di lavoro, settata con la configurazione del software R, oppure in qualsiasi cartella del proprio pc. Nel secondo caso oltre al nome con l'estensione è necessario specificare anche tutto il percorso;

- **header**: è un valore binario che informa il software di dover considerare la prima riga del file di input come una rig di intestazione, ove sono presenti i nomi dei campi del singolo record. Essendo valori binari avremo `header=F` se vogliamo porre l'header a FALSE e quindi indicare che non c'è la riga di intestazione, al contrario avere `header=T` che pone l'header a TRUE;

- **sep**: serve a informare il software il separatore di colonna che divide i vari campi di un record;

- **quote**: consiste nel carattere utilizzato nel file per delimitare le stringhe di testo dei vari record.

Questa funzione crea quindi un oggetto di nome `tweet`, variabile utilizzata per assegnare i valori del data-set dato in input [56].

3.3.2 Preprocessing dei tweet

La variabile `tweet` contiene perciò i dati in input, ma questi sono salvati sotto forma di oggetto di tipo `data.frame`, che sono definiti come collezioni di variabili che condividono molte delle proprietà tipiche delle liste e delle matrici.
A questo punto, il nostro obiettivo è compiere quello che viene definito "pre-processing". Si ricorre a questo procedimento in quanto raramente i dati sperimentali sono pronti immediatamente per essere analizzati nelle fasi successive ed inoltre utilizziamo quello che è il modello definito "bag of words".

Questo modello vede i dati come un insieme di parole in successione e ha come idea alla base che le parole di un testo sono indicative del suo contenuto di conseguenza si può ricavare una semplice rappresentazione strutturata del contenuto.

Per effettuare questo preprocessing dei tweet, R rende disponibile le funzioni della libreria `tm`. Questa libreria però vestisce su tipologia di dati specifica: il `corpus`. Il `corpus` consiste in un tipo di dato che serve per memorizzare un insieme di documenti, che nel nostro caso sono i contenuti testuali dei tweet, ossia i 140 caratteri che l'utente può pubblicare dal proprio profilo. Per definizione, esso è una collezione di testi selezionati e organizzati per facilitare le analisi linguistiche.

```
#creo il corpus
myCorpus <- Corpus(VectorSource(tweet\$text)
```

Listing 3.2: Creazione corpus

Come si intuisce dallo script sopra riportato, il `corpus` è stato costruito in modo particolare dal campo `txt` di ogni record raffigurante il contenuto effettivo dei tweet, selezionato dal data frame `tweet` prima creato.
In primo luogo in una rappresentazione di documenti una trasformazione importante consiste nel normalizzare i termini simili. Queste trasformazioni sequenziali utilizzano tutte sul corpus una funzione chiamata `tm_map`. Quest'ultima funzione si definisce funzione di ordine superiore, in quanto ha come argomento un'ulteriore funzione che applica a tutti gli elementi del corpus in questione.
La normalizzazione di cui si parlava, implica varie funzioni che possono essere applicate a tutti i documenti del corpus. Iniziamo con la prima che permette di convertire tutti i dati in minuscolo. Questo processo viene definito "case-folding". Per fare questo applica `content_traformer()`che è una funzione che ha il potere di modificare i contenuti degli oggetti di R.

```
#conversione in minuscolo
myCorpus <- tm_map (myCorpus,
```

```
                content_transformer(tolower))
```

Listing 3.3: Conversione tweet in minuscolo

Ora con il codice seguente si va a rimuove i segni di punteggiatura, numeri e spazi bianchi extra, ossia quelli che eccedono a quello utile per dividere le parole, affinché rimanga solo una sequenza di una varietà di "parole", o meglio termini, in successione. Insieme alle parole andiamo anche a rimuovere quelli che sono i numeri, in quanto non è di nostro interesse utilizzarli in questa analisi, visto che lo studio, che in seguito andremo a effettuare, considera i "tag" presenti nei tweet, seguendo il modello "Bag of words", prima nominato.
Inoltre per quanto riguarda l'ultima parte del codice, si va ad eliminare le stopword, ossia quelle parole come gli articoli e le preposizioni che prese da sole non danno indicazioni sull'argomento. Inoltre, visto che non tutte le congiunzioni sono presenti in `stopwords("it")`, abbiamo rimosso anche le parole precedentemente decise in seguito alla visione dei tweet estratti. Infine attraverso la funzione `PlainTextDocument()` sono avvenute le modifiche fatte, in quanto ha il compito di trasformare il contenuto del corpus in un documento di testo.

```
# rimozione numeri
myCorpus <- tm_map(myCorpus, removeNumbers)

# rimuovo spazi extra
myCorpus <- tm_map(myCorpus, stripWhitespace)

#rimozione delle stopwords
stoplist <- readLines("C:/stopwords.txt")
myCorpus <- tm_map( myCorpus, removeWords, stoplist)
myCorpus<- tm_map(myCorpus, stopwords("it"))
myCorpus<- tm_map(myCorpus, PlainTextDocument))
```

```
# rimozione punteggiatura
myCorpus <- tm_map(myCorpus, removePunctuation)
```

Listing 3.4: Text cleaning

Come si evince dai commenti del pezzo di codice seguente, vengono eliminati tutti gli url da quelli pubblicati volutamente a quelli delle immagini. Oltre a questi si elimineranno anche gli hashtag, in quanto già presenti in un campo apposta del data-set ed è già possibile da lì fare un'analisi degli stessi.

```
#creo una copia del corpus originale
myCorpusCopy <- myCorpus

#stemming
myCorpus <- tm_map(myCorpus, stemDocument,
                   language="italian")

stemCompletion2 <- function(x, dictionary) {
  x <- unlist(strsplit(as.character(x), " "))
  x <- x[x != ""]
  x <- stemCompletion(x, dictionary=dictionary)
  x <- paste(x, sep="", collapse=" ")
       PlainTextDocument(stripWhitespace(x))
}
myCorpus <- lapply(myCorpus, stemCompletion2,
                   dictionary=myCorpusCopy)
```

Listing 3.5: Stemming

Questa parte è chiamata la fase di stemming. Lo "stem" è una parola o una parte di parola che costituisce la materia prima su cui costruire un analisi statistica. In questo caso consiste nel troncamento di una parola, ricavando così la radice del termine.
Prima di tutto si è creata una copia del corpus, affinché si potesse conservarne una copia da utilizzare in seguito come un dizionario per il completamento

della radice delle parole. Successivamente eseguiamo lo stemming del corpus e successivamente applicando la funzione `stemCompletion2()` ritorniamo a completare le radici dei termini con il dizionario in questione, cancellando le stringhe vuote extra se sono presenti.

3.3.3 Studio delle parole

Dopo questa fasi di preparazione dei contenuti del dataset, si procede con lo studio delle parole più frequenti: introduciamo quindi la Term Document Matrix, o TDM. La TDM è una matrice le cui righe corrispondono ad un termine e ogni colonna ad un testo di un tweet. Di conseguenza, visto lo scopo della matrice, la cella indica il numero di occorrenze delle parole per ciascun tweet. Per visualizzare e gestire questa matrice `tdm` utilizzeremo:

`> inspect(tdm[ 1:20, 1:20])` oppure `> tdm`

Il primo esamina la parte della matrice specificata, mentre la seconda informa sul numero dei termini e quali di questi non sono nulli. Per implementare la costruzione di questa matrice, R ci offre l'opportunità di utilizzare delle funzioni, sempre utilizzando la libreria `tm`. Il comando per creare la matrice richiede, come argomento in input, un oggetto di tipo Vector, creato nel nostro caso inizialmente con la creazione del corpus, utilizzando la funzione `VectorSource`, che prende come parametro i testi dei tweet.

```
#creo la Term Document Matrix
tdm <- TermDocumentMatrix( myCorpus )
tdm1 <- removeSparseTerms(tdmc, 0.99)
```

Listing 3.6: Creazione Term Document Matrix

Oltre alla creazione della matrice con la matrice `tdm1` si identifica la matrice dopo un processo di rimozione di tutti i termini che compaiono in meno dell'1% dei tweet, visualizzati come le colonne della matrice. La funzione ha quindi apportato modifiche alla `tdm`, riducendo la caratteristica della "spar-

sity". Utilizziamo il comando sopra descritto per visualizzare le principali caratteristiche delle due matrici `tdm` e `tdm1`.

```
> tdm
  <<TermDocumentMatrix (terms: 1328, documents: 362)>>
  Non-/sparse entries: 2116/478620
  Sparsity           : 100%
  Maximal term length: 109
  Weighting          : term frequency (tf)

> tdm1
  <<TermDocumentMatrix (terms: 102, documents: 362)>>
  Non-/sparse entries: 1004/35920
  Sparsity           : 97%
  Maximal term length: 13
  Weighting          : term frequency (tf)
```

Listing 3.7: Confronto matrici tdm e tdm1

Come si può notare, questo testing ha apportato grandi differenze fra le due matrici. Nella seconda sono stati filtrati solo i termini con una maggior frequenza, ed è per questo che, se anche di poco si è abbassata anche la caratteristica di sparsità dei termini.

```
#trovo tutti i termini nei documenti
fft<- findFreqTerms(tdm1, lowfreq=5)
write.table(fft, file="/words.csv", quote=F, sep=" ",
            dec=".", na="NA", row.names=T, col.names=T)

#calcolo la frequenza delle parole
tdmat = as.matrix(tdm1)
v = sort(rowSums(tdmat), decreasing=TRUE)
d = data.frame(word=names(v), freq=v)
```

```
#scrivo su file .csv le frequenze e le parole
write.table(v, file="freqWords.csv", quote=F, sep=" ",
            dec=".", na="NA", row.names=T, col.names=T)
```

Listing 3.8: Calcolo parole e relative frequenze

Con il codice sopra riportato, vediamo come trovare le parole in esame e le relative frequenze. Era possibile attraverso librerie apposite, all'interno del package `graphics` di R. Queste sarebbero funzioni ad alto livello che permettono di generare una nuova finestra grafica in modo automatico, passando i giusti argomenti. Nonostante la sua grande potenza si è preferito riportare tutti i dati in file csv, così da studiarli meglio in un secondo passaggio, senza fare qualcosa di altamente automatico e ad alto livello. Per questo si notano nel codice varie `table.write`.

Fatto uno studio delle parole a livello di dati, si è deciso di creare quel che è un `world-cloud`, anche definito inletteratura come "tag-cloud".
Il "tag-cloud" è la rappresentazione visiva dell'analisi appena descritta, in quanto vengono visualizzate i tag, o parole-chiavi. La lista solitamente è in ordine alfabetico e vengono attribuiti font e dimensioni diverse a seconda dell'importanza delle parole. La dimensione quindi raffigura il peso del tag in questione e nel nostro caso si identifica con la frequenza della parola nei documenti della matrice [36]. Il codice raffigurante questo passaggio è il seguente.

```
library(wordcloud)
library(colorspace)
wordcloud(d$word, d$freq, min.freq=900,
          random.color=TRUE, colors=rainbow(7))
```

Listing 3.9: World-cloud

Per creare questa tag-cloud, utilizziamo la libreria predisposta da R, chiamata `worldcloud`, insieme alla libreria `colorspace`, che ha lo scopo di colorare le parole stampate nella world-clud, ed in questo caso avendo predisposto l'opzione `random.color=TRUE` si avranno appunto colori casuali [55].

Come parametri vengono passate le parole e la frequenza della matrice `d` creata precedentemente.

3.3.4 Studio sulla geo-localizzazione e temporizzazione dei tweet

Un altro aspetto, a livello teorico, alquanto importante da analizzare consiste nella geo-localizzazione dei tweet. È importante comprendere dove questi contenuti vengono pubblicati, per avere una stima di quanta è la possibilità delle vittime del terremoto di interfacciarsi a internet e quindi utilizzarlo come canale d'emergenza.Inoltre se si vuole capire l'entità del terremoto ha senso analizzare i tweet di quelli che lo hanno vissuto in prima persona, i quali di conseguenza condividono contenuti dal luogo di del sisma.
Per far questo si è semplicemente utilizzato una query SQL, in modo che mi raggruppasse i tweet per luogo.

```
SELECT geo, count(*) as numtweet
FROM `terremotocentroitaliatw`
GROUP BY geo
```

Listing 3.10: Numero dei tweet per geo-localizzazione

Questa query permette di ottenere come risultato il conteggio dei tweet per ogni luogo specificato. Per far questo si è selezionato il luogo salvato nel data-set e, attraverso un `count(*)`, il numero dei tweet. Questo è reso possibile grazie ad una funzione di aggregazione, che ha permesso il raggruppamento per luoghi, il `gropu by`. Infine sempre con lo stesso linguaggio ho creato un grafico che facesse vedere la distribuzione dei tweet. I risultati verranno proposti nel paragrafo successivo.

Infine la stessa cosa è stata riproposta invece per vedere, in base all'attributo `datatweet` quale son stati i giorni con più attività. Il codice per comprendere questo è parzialmente simili al quello per la geo-localizzazione dei tweet cambiando ovviamente gli attributi della query.

Infine, tramite Google Maps, è stato possibile creare mappe personalizzate attraverso le API che Google ha messo a disposizione in modo gratuito a seguito di una registrazione, che rende disponibile anche una dashbord per gestire in modo grafico le richieste [21]. Per creare una nuova mappa personalizzata è possibile inserire i dati manualmente, posizionando dei markers sulla zona che si vuole marcare in base al campo `geo` all'interno del database. Questo è stato fatto facendo corrispondere la stringa di geo-localizzazione a delle coordinate. Inoltre, grazie la mappa e i relativi markers aggiunti sarà possibile vedere in maniera visiva e più immediata la dispersione del tweet per il mondo.

3.3.5 Filtraggio dei tweet

A questo punto, ci si è accorti, grazie allo studio sulla geo-localizzazione e all'analisi fatta, che la maggior parte dei tweet sono riguardanti donazioni, preghiere e pensieri verso i terremotati. Per questo è sembrato dovuto un filtraggio sul data-set ricavato. Per far questo, si è ricorsi al linguaggio SQL, per fare interrogazioni direttamente sul database creato.
Il codice SQL è il seguente:

```
SELECT *
FROM terremotocentroitaliatw
WHERE id NOT IN( SELECT id
                 FROM `terremotocentroitaliatw`
                 WHERE txt LIKE "%solid%"
                 OR "terremotati" OR "silenzio"
                 OR "preghiera" OR "donazion%"
                 OR "donare" OR "dolore"
                 OR "aiutare" OR "aiuto"
                 OR "preghier%" OR "gesto" OR "conto"
                 OR "terremotati" OR "fondi")
```

Listing 3.11: Filtraggio dei dati

Come si può notare le parole scelte sono parole specifiche che sono state ricavate dall'elenco dei tag creato nell'analisi iniziale. Questa query innestata ottiene come risultato l'insieme di tutti quei record che possiedono un contenuto che non presenta le parole specificate.

3.4 Risultati

In questo paragrafo andremo a vedere quali sono i risultati di questo studio fatto. Si va avanti con l'ordine svolto anche nel paragrafo precedente.

Partiamo dall'analisi a livello lessicale che abbiamo fatto. In primo luogo proponiamo il world-cloud, o tag-cloud.

Figura 3.1: Worldcloud creato con software R [17]

Come si può notare dalla figura 3.1, oltre alla parola "terremoto" ben visibile a tutti, tutte le altre parole sono pressoché della stessa dimensione. Questo fa emergere come non ci siano parole chiave ben delineate. Oltre alla parola principe si hanno altri termini con una frequenza maggiore del resto e sono:

- colpito;
- vittime;
- scossa;
- protezione.

Questa situazione è stata testata anche con l'utilizzo del codice sulla console di R. Se infatti proviamo a eseguire il seguente codice

```
tdm <- TermDocumentMatrix( myCorpus )
tdm1 <- removeSparseTerms(tdmc, 0.4)
```

Listing 3.12: Test sulla Term Document Matrix

vediamo come le caratteristiche della matrice `tdm1` cambiano significativamente. Se avessimo voluto visualizzarle, bisognava digitare comando giusto per visualizzarne le caratteristiche.

```
>tdm1
   <<TermDocumentMatrix (terms: 1, documents: 362)>>
   Non-/sparse entries: 340/22
   Sparsity           : 6%
   Maximal term length: 9
   Weighting          : term frequency (tf)
```

Listing 3.13: Output matrice tdm1

Come si evince dal risultato del comando `> tdm1` la caratteristica "sparsity" da 97% iniziale è andata ad un 6%. Questo potrebbe essere un aspetto positivo, ma non in questo caso visto il numero dei termini trovati, che è pari ad 1. L'unico termine rimasto quindi è quello che nella world-cloud è molto più visibile del resto. In sé, l'analisi fatta non ha dato risultati alquanto significativi, in quanto poche parole ricorrevano nei tweet e non avevano alcun tipo di associazione significativa. Però sono state utili come base per in combinazione allo studio fatto invece attraverso lo studio della geo-localizzazione.

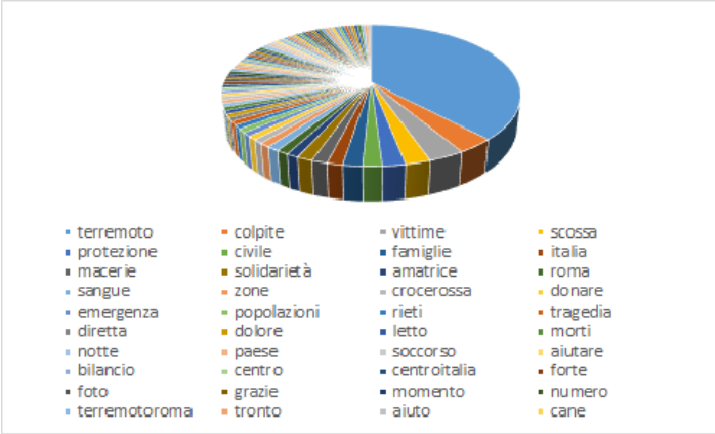


Figura 3.2: Grafico a torta con i termini ricavati dall'analisi - R

Ora utilizziamo un ortogramma a nastro, ossia un particolare tipo di diagramma usato dagli statistici per effettuare un'indagine. In particolare si tratta di una variante di un istogramma, in cui la frequenza assoluta è posizionata sull'asse delle ascisse ed è composta da rettangoli che si poggiano sulla l'asse delle ordinate. Osserviamo attentamente il seguente grafico, creato grazie alla piattaforma già descritta phpMyAdmin, collegata al database della ricerca [47].

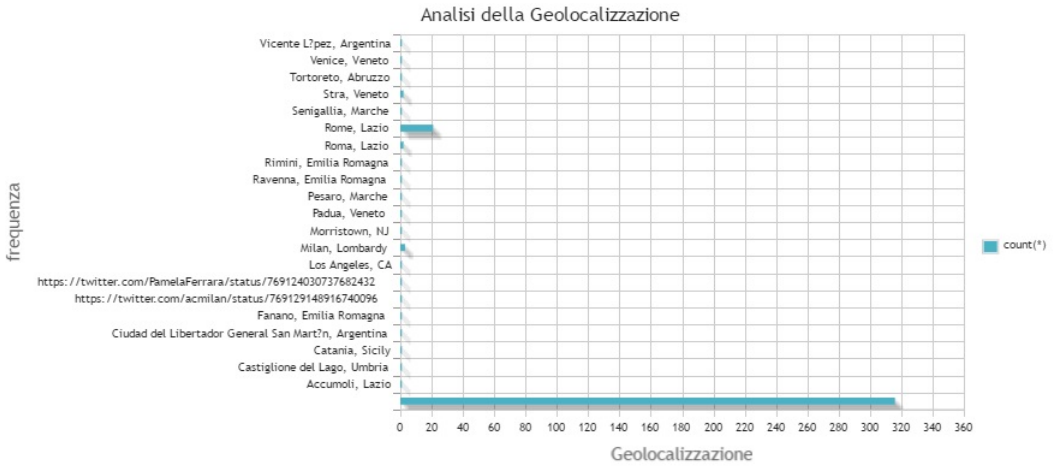


Figura 3.3: Ortogramma sulla geo-localizzazione dei tweet - SQL [47]

La particolarità di questo diagramma, che salta subito all'occhio, è la lunghezza del primo rettangolo, quello caratterizzato dalla mancanza di un etichetta. Questa assenza non è dovuta al caso o ad uno sbaglio, ma da un'informazione ben precisa: non tutti i tweet possiedono una geo-localizzazione. Questo è uno dei problemi scaturiti da quest'analisi, in quanto non tutti i tweet contengono tutte le informazioni che sono presenti nel file json di risposta. La geo-localizzazione può essere consentita o no a discrezione dell'utente ed inoltre può anche non risultare a causa delle impostazioni della privacy appartenenti a ciascun profilo. Questa riflessione ci fa giungere alla conclusione che la geo-localizzazione memorizzata, a primo impatto non da nessun'informazione di grande rilevanza.
Proviamo invece a osservare come i tweet si distribuiscono nella mappa mondiale.

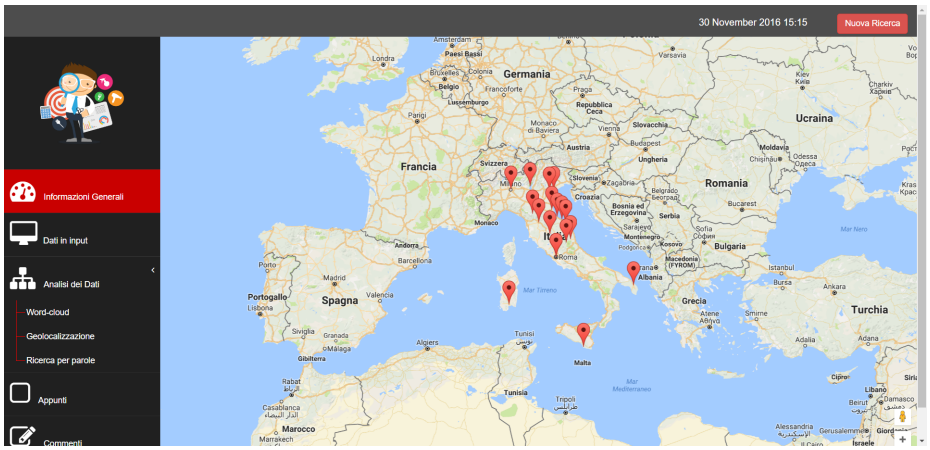


Figura 3.4: Distribuzione a livello globale dei tweet - Google Maps

Questa schermata è stata catturata dall'applicazione fatta come progetto di tesi, che nel capitolo seguente andremo a descrivere in modo approfondito. Come si può notare alcuni dei tweet non sono stati pubblicati nei luoghi limitrofi all'epicentro dell'evento sismico in questione, ossia provincia di Rieti con latitudine pari a 42.7 e longitudine pari 13.23.
Questa considerazione si può considerare una buona base di riflessione, in quanto si può dedurre che non tutti gli utenti che hanno "twittato" sono

vittime in prima persona dei disastri e degli effetti più significativi del terremoto. Questo quindi ha orientato l'analisi verso un atto di filtraggio dei dati a disposizione. Tramite quindi la query innestata descritta nel paragrafo apposito 3.3.5, abbiamo ridotto la dimensione del data-set escludendo quindi i tweet che riguardavano preghiere, donazioni e riflessioni sul terremoto. Quest'azione di filtro è stata fatta tenendo ben presente l'obiettivo dell'analisi che consiste nel riuscire a capire se i tweet possano dare informazioni concrete d'emergenza, che potrebbero servire a coloro che prestano soccorso per farsi un idea più dettagliata e per mirare le proprie strategie d'azione.

A questo punto avendo anche un data-set di dimensione ridotte contenente i tweet che potrebbero avere più impatto sull'analisi, si è deciso di studiare l'andamento temporale dei tweet, Per questo, sempre attraverso query SQL e grazie a tool grafico presente nella piattaforma phpMyAdmin, si è creato un ortogramma in grado di far osservare se nel momento e nel giorno della grande scossa, ci sono stati riscontri immediati da parte degli utenti.

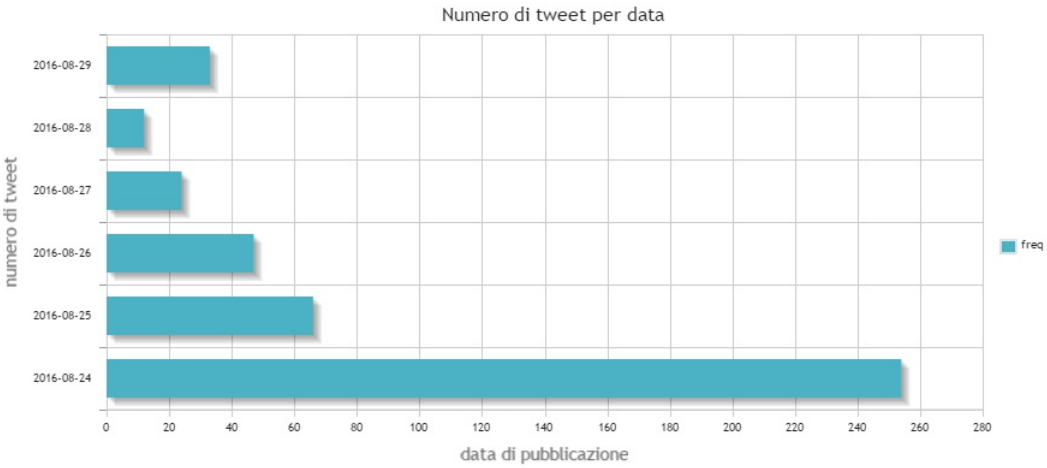


Figura 3.5: Ortogramma sull'andamento temporale dei tweet - SQL

Come si può notare, questa specifica analisi va a nostro favore, in quanto la maggior parte dei tweet sono stati pubblicati nel giorno del terremoto: 24-08-2016. Volutamente non sono stati riportati gli orari, perch l'evoluzione oraria dei tweet fino al giorno successivo risulta un andamento continuo.

Andando a riprodurre l'analisi testuale e di geo-localizzazione sui tweet del 24-08-2016, si è notato che nulla cambia, né a livello di frequenza di parole né a livello di distribuzione della pubblicazione dei tweet. Questa assenza di variazione è dovuta al fatto che la percentuale dei tweet avvenuti il giorno stesso del terremoto incidono in maniera massiccia sull'analisi globale di tutto il data-set, quindi la variazione fra le due analisi fatte è davvero minima. Controllando quindi questo insieme di dati si è notato che la maggioranza di questi non soddisfano i requisiti del nostro obiettivo iniziale, in quanto informano solo dell'avvenuto terremoto e chiedono conferma ai propri follower della situazione loro.

Un aspetto inoltre da prendere in considerazione è l'effettiva possibilità di utilizzo della rete e per questo ho selezionato alcuni dei tweet che parlassero di questo, ma il risultato è stato che tutti i tweet sono contenuti che tendono ad invitare chiunque abbia la connessione di renderla pubblica a tutti per queste situazioni di emergenza. Un esempio può essere visibile nella seguente schermata, in figura 3,6, che rappresenta i tweet riguardanti questo argomento. Infine bisogna tener in considerazione che in questo caso si tratta di dati creati da persone umane, di conseguenza sicuramente in un primo momento i tweet riscontrati sono di persone esterne agli ambienti in cui gli effetti del sisma sono stati devastanti, in quanto pubblicare sui social media non è l'azione più immediata a cui la vittima potrebbe ricorrere. Inoltre un altro aspetto da non dimenticare, consiste nel fatto che coloro, che sono vittime degli eventi più catastrofici, avranno una possibilità minima o quasi assente di accesso alla rete. D'altro canto è possibile comunque in modi alternativi, come per esempio il passaparola, che gli utenti possano richiedere aiuto o fare appello ai social network, compresi quindi anche microblogging come Twitter.

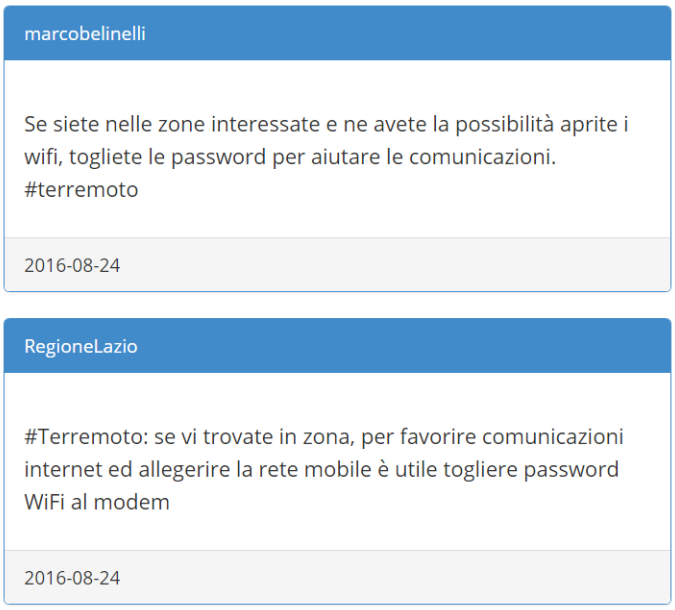


Figura 3.6: Tweet riguardanti la disponibilità di rete e connessione- Applicazione progetto tesi

Capitolo 4

Applicazione: Information Discovered

In seguito all'analisi effettuata, si è deciso di creare un prototipo per un'applicazione che abbia lo scopo di mettere a disposizione dell'utente comune uno strumento capace di compiere un'analisi sul data-set, che l'utente ha preventivamente recuperato e importato nell'applicazione.
In questo capitolo, in modo particolare, si andrà a descrivere tutti gli aspetti dell'applicazione web in questione: requisiti, design dell'applicazione e gli sviluppi futuri. Si è deciso di chiamare l'applicazione creata "INFORMATION DISCOVERED", per dare anche l'idea dell'obiettivo della stessa.

4.1 Analisi

Per compiere un'analisi completa dei requisiti è necessario valutare quali siano le vere e proprie funzionalità da implementare in questa applicazione, tenendo comunque in considerazione l'obiettivo principale sopra riportato. In particolare i punti principali sono:

- Inserimento dei dati relativi all'utente in questione e alla ricerca da effettuare, creando una sessione in cui analizzare il proprio data-set.

Di conseguenza c'è anche l'azione di log out in cui la ricerca andrà persa;

- Importazione del data-set che l'utente si prospetta abbia già ricavato da un'estrazione di tweet precedente a proprio piacere;

- Analisi dei tweet in base alla frequenza delle parole attraverso due risultati: uno più grafico che consiste nella creazione di una tag-cloud e l'altro più teorico che consiste nella visualizzazione dei tweet per le parole più frequenti ricavate;

- Visualizzazione della distribuzione dei tweet sulla mappa mondiale;

- Possibilità per l'utente di fare dei commenti all'applicazione, per avere un riscontro dell'utenza al sito

- Possibilità di scrivere appunti sulla ricerca che si sta effettuando e cancellarli nel caso non siano più utili.

Ora andremo ad ispezionare meglio e più dettagliatamente i requisiti dell'applicazione.

4.1.1 Requisiti

La modellazione dei requisiti funzionali dell'applicazione viene rappresentata attraverso il diagramma degli stati d'uso. Si tratta di un diagramma che esprime il comportamento offerto e desiderato, sulla base dei risultati osservabili dall'applicazione in questione. Esso infatti individua chi e che cosa fa il sistema attraverso due elementi, che rispettivamente sono l'attore e il caso d'uso.

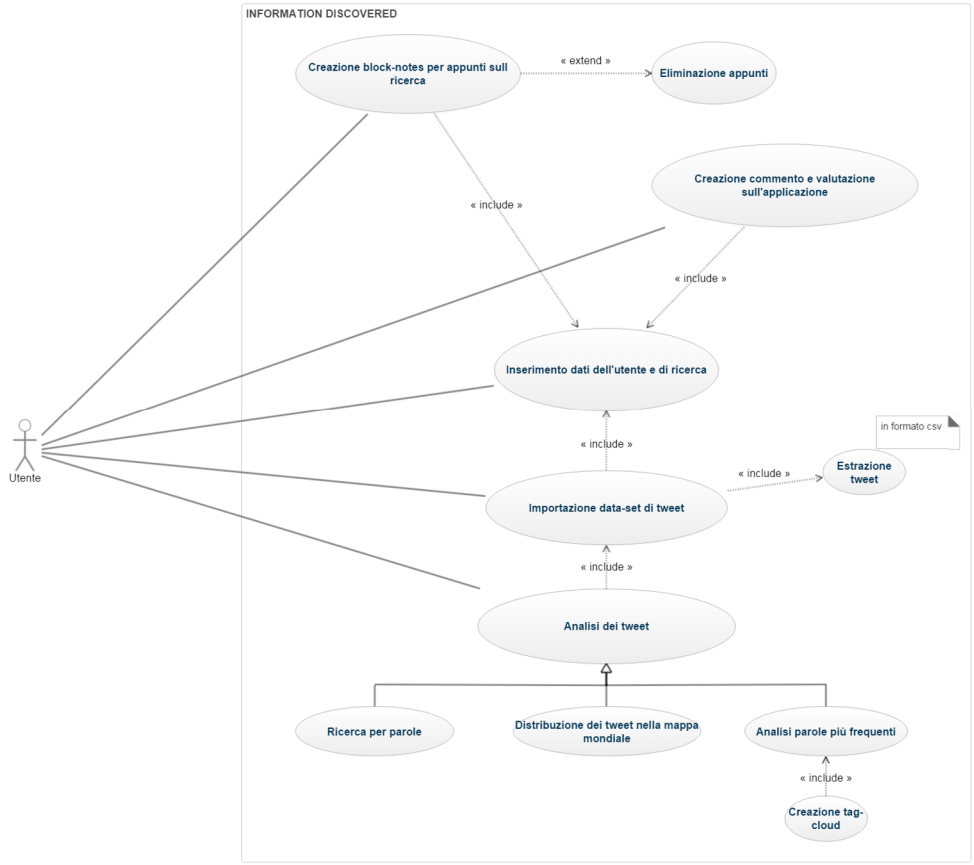


Figura 4.1: Diagramma dei casi d'uso - INFORMATION DISCOVERED

L'unico attore del nostro sistema è costituito dall'utente che vuole effettuare l'analisi di un data-set precedentemente creatosi. I casi d'uso invece corrispondono ai punti presenti nell'introduzione del paragrafo 4.1, che ne identificano i comportamenti principali.
Come si nota dal diagramma, tutti i casi d'uso sono legati attraverso associazioni all'utente, l'unico in grado di poter effettuarli, ma oltre a ciò sono tutti collegati attraverso degli "≪include≫". L'"≪include≫" è una dipendenza tra casi d'uso, che implica che il caso incluso faccia parte al comportamento di quello che lo include ed inoltre l'inclusione non è una dipendenza opzionale. L'utente infatti ha la possibilità di inserire i propri dati e quelli della propria ricerca all'interno dell'applicazione così poi da permette la fruibilità di tutte

le altre funzioni ed è questo il motivo per il quale le inclusioni, inserite nel diagramma, convergono su quel caso d'uso specifico. L'utente, infatti, affinché non inserisce e invia i dati giusti, non può accedere a nessuna funzionalità della web application. Dopo aver compiuto questa prima fase, che potrebbe essere definita di "log-in", si attiva la sezione d'importazione del data-set, dei commenti e anche quella degli appunti sulla ricerca. L'importazione del data-set contenente i tweet per l'analisi implica la creazione di un insieme di tweet, che l'utente esternamente deve aver estratto. Questa estrazione come si evince dalla nota di commento vicino al caso d'uso, deve essere in formato csv e con particolari caratteristiche descritte esplicitamente sulla pagina web in cui viene svolta questa operazione.
Ora passiamo al caso d'uso più importante e principale nell'applicazione: l'analisi dei dati. In questo caso si nota l'elemento della generalizzazione applicato a questo caso d'uso: l'analisi dei dati è composta da altri tre casi d'uso. Questi consistono nei punti sopracitati che corrispondono alla ricerca dei dati per parole, alla geo-localizzazione dei tweet e allo studio sui termini più frequenti. Quest'ultima funzionalità include la creazione di una tag-cloud.
Infine, ma non per importanza, l'applicazione dà la possibilità all'utente di creare un block notes per inserire degli appunti riguardanti la ricerca in sessione. L'utente una volta scritti i propri appunti nell'area apposita, può opzionalmente anche decidere di cancellarli uno per volta a suo piacimento. L'applicazione offre anche l'opportunità di inviare un feedback, attraverso una valutazione numerica dell'utente e un testo libero, in cui è possibile e auspicabile dare consigli e scrivere una propria valutazione più articolata.

4.2 Design dell'applicazione

Per quanto riguarda il design della nostra applicazione web, distinguiamo tra ciò che si è svolto client side e ciò che invece si è realizzato lato server. Per il lato client, si è utilizzato bootstrap, uno dei più famosi framework, oramai

compatibile con le ultime versioni dei principali browser,per la creazione dei siti e applicazioni sul web. Questo framework infatti contiene modelli di progettazione su HTML e CSS per il front-end dell'applicazione e quindi per le varie componenti dell'interfaccia. Per la parte lato server, invece, si sviluppato l'applicazione tramite un linguaggio di scripting interpretato, in particolare concepito per la creazione di pagine web dinamiche. Per gestire i dati dell'applicazione abbiamo utilizzato phpMyAdmin [18], già descritto nel capitolo soprastante e per interfacciarsi all'applicazione si è utilizzato PHP con un approccio orientato agli oggetti.

4.2.1 Interfaccia dell'applicazione

Come sopra annunciato, si è deciso di utilizzare dei template che Bootstrap [54], mette a disposizione, in particolare ho utilizzato "Bootsrap Template Admin", in quanto presentava caratteristiche che potessero essere adatte con il progetto in questione. Inoltre è facilmente, e sopratutto liberamente, personalizzabile come dashboard, così da permettere all'utente una visione d'insieme per la propria ricerca da effettuare. In primo luogo è un template sia per uso personale e commerciale con una licenza di tipo "MIT". L'autorizzazione di questa licenza è concessa, a titolo gratuito, a chiunque ottenga una copia di questo software ed inoltre non ha limitazioni riguardanti i diritti di utilizzare, copiare, modificare, unire, pubblicare e distribuire il codice. Questo insieme all'utilizzo di un codice comprensibile e ad un design leggero, mi hanno dato la possibilità di modificare a mio piacimento il font-end del sito, così da adattarlo alle esigenze e agli obiettivi per progetto. Un altro aspetto molto significativo è la presenza di un design responsive, cosicché possa funzionare senza problemi su qualsiasi piattaforma.

4.2.2 Implementazione

La tecnologica principale dell'implementazione della logica dell'applicazione consiste nell'utilizzo, come detto in precedenza di PHP. Il protocollo utilizza-

to è il protocollo HTTP e viene utilizzato per compiere richieste per veicolare i dati elaborati e fornire una risposta al client. Per sviluppare in PHP sono necessari alcuni stupenti per preparare l'ambiente di lavoro, fra cui l'editor di testo e un server web. L'editor di testo utilizzato in questo progetto consiste in ATOM [5], accessoriato di plug-in e strumenti molto adatti allo sviluppo web. Per quanto riguarda il server web si è deciso di utilizzare Apache, il più utilizzato a livello mondiale. Questo permette di sviluppare l'applicazione in web server locale e non reale. Questa scelta di lavorare e mantenere in locale è stata dovuta dal fatto che questa applicazione risulta ancora un prototipo, che in futuro può ambire ad ampliamenti e miglioramenti. Inoltre un altro aspetto è l'interazione con i dati, la quale viene effettuata tramite un database MySQL. In questo caso quindi, si è optato per l'utilizzo di PHP development environment, chiamato XAMPP [3]. Questa piattaforma si adatta alle nostre esigenze, in quanto installa su tutti i sistemi operativi Apache, MySQL e PHP.

Un importante funzione di PHP consiste nel permettere la modifica di codice HTML, dandone il comportamento e quindi l'output in base alle nostre esigenze. Per fare ciò si è utilizzato del codice embedded al linguaggio HTML, cambiando così l'estensione del file a .php; in particolare viene elaborato il codice PHP che si trova all'interno del tag `<?php ... ?>`. Inoltre ho utilizzato anche file esterni unicamente in php e passati nell'attributo `action` delle varie form all'interno dell'applicazione.

Come si può notare nella sezione della barra di navigazione "Analisi dei dati ", faccio eseguire all'applicazione del codice in R, il software usato per la ricerca fatta sull'evento sismico del 24 Agosto 2016. Il codice utilizzato è una variante di quello già descritto nell'analisi del progetto di tesi, salvato tutto in un file con estensione ".R". Ora mostro il file, chiamato "scriptR.php", che ha il compito di eseguire lo script R creato.

```
<?php

        session_start();
```

```
        $file = $_SESSION["file"];
        exec("Rscript script.r $file");

        ...
?>
```

Listing 4.1: PHP per esecuzione script R

PHP mette a disposizione le funzioni `exec()` e `shell_exec()` per eseguire script. La scelta è stata diretta sulla prima funzione, come si può vedere dal codice sovrastante, poiché così è possibile eseguire codice proveniente da programmi esterni, a differenza di `shell_exec()` che esegue comandi da shell. Come si può notare `exec()` prende argomenti in input, in particolare in questo caso diamo in input il file, che una volta caricato sull'applicazione dall'utente viene inserito in una variabile sessione [46].

L'applicazione non possiede un log-in vero e proprio, ma ad ogni ricerca si apre una nuova sessione. Per operare in questo modo abbiamo utilizzato le variabili sessione in PHP, che vengono definite nel seguente modo

```
$_SESSION["nomevariabile"];
```

La sessione si chiude, e di conseguenza distrugge tutte le variabili di sessione memorizzate, sia manualmente, quando l'utente preme sul pulsante "Nuova Ricerca", sia automaticamente quando viene chiuso il browser. La sessione salva, dopo inserimento dei dati iniziali sulla ricerca e sull'utente, il file importato per compiere la ricerca.
Con il caricamento del file importato inoltre si crea una tabella temporanea in un database relazionale, in cui vengono inseriti i tweet importati in una tabella, con la stessa struttura di questa utilizzata in analisi e descritta nel terzo capitolo. Inoltre viene utilizzato anche una tabella denominata "commento", che ha il compito di memorizzare i commenti liberi e le valutazioni numeriche che un utente può compiere sul sito.

4.3 Guida all'Utente

L'applicazione, per quanto possa ancora essere un prototipo, è stata pensata come un ambiente che permetta un facile utilizzo per chiunque. Per questo si è cercato di utilizzare una logica intuitiva, che si vede subito con collegamento al sito, in quanto qualunque pulsante della barra degli strumenti l'utente desideri premere compare come pagina web, quella in cui inserire i dati in riferimento alla propria ricerca e il nome e mail dell'utente stesso.

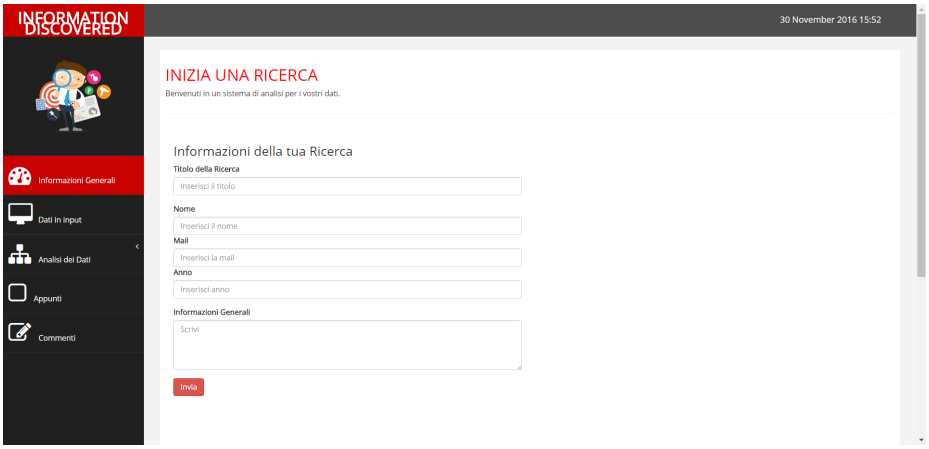


Figura 4.2: Pagina web di inserimento dati - INFORMATION DISCOVERED

Con l'invio dei dati inserite nella pagina web, il secondo passaggio è quello di inserimento del data-set precedentemente creato dall'utente. Questo data-set deve avere delle caratteristiche particolari, basati anche sul fatto che in questo sistema si analizzano solamente contenuti estratti dal microblog Twitter. Il file che si andrà a caricare deve essere del formato csv, senza la riga d'intestazione ma deve contenere solamente i dati. Inoltre i campi devono essere inseriti nel seguente ordine : id, username, data, numero di retweets, testo del tweet, geolocalizzazione, hashtag, link. Una volta inserito si viene reindirizzati alla pagine con le informazioni della ricerca, alla quale si accede premendo anche "Informazioni Generali"

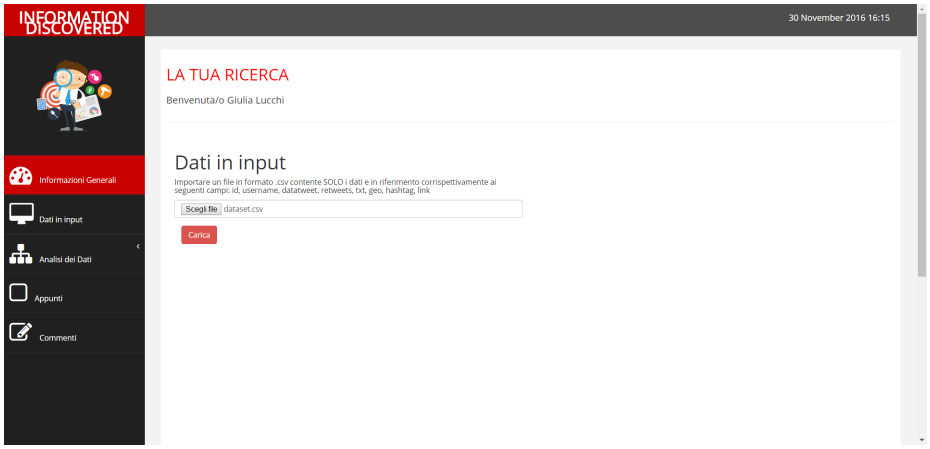


Figura 4.3: Pagina web di caricamento data-set - INFORMATION DISCOVERED

Andiamo nel prossimo paragrafo ad illustrale quali sono le principali funzionalità che l'utente può compiere.

4.3.1 Funzionalità principali

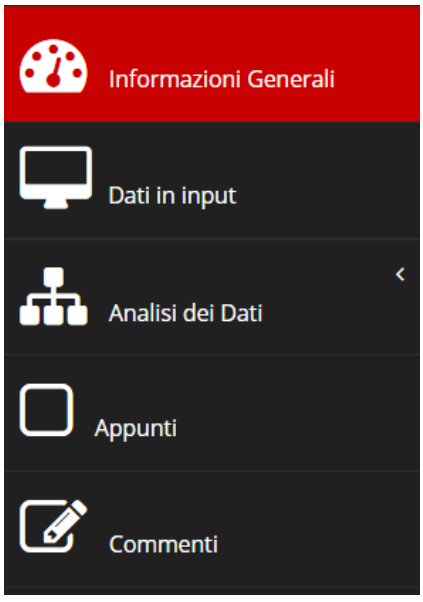


Figura 4.4: Barra degli strumenti - INFORMATION DISCOVERED

Le funzionalità e i requisiti dell'applicazione sono intuibili e ben visibili sulla barra degli strumenti, messa a disposizione dell'utente per navigare all'interno dell'applicazione web. Con la figura 4.4 a fianco si può osservare a barra degli strumenti con i relativi pulsanti, che si colorano di rosso nel momento in cui l'utente preme il pulsante e viene indirizzato verso la pagina web apposita.
I primi due pulsanti sono già stati descritti precedentemente con i passi preliminari alla ricerca. Con "Informazioni Generali", quindi, vengono

mostrate le informazioni di base della ricerca, mentre con "Dati in input" si vede se i dati sono già stati inseriti nell'applicazione e in caso contrario c'è la form apposita che permette ciò. Nel caso in cui invece i dati siano già stati inseriti, allora è possibile incominciare una nuova ricerca, perdendo tutto quello creato in sessione.
Il cuore dell'applicazione è all'interno del pulsante "Analisi dei Dati", che a sua volta contiene le tre tipologie d'analisi:

- **World-Cloud**: attraverso l'esecuzione di uno script visualizza graficamente la frequenza delle parole all'interno dei tweet in questione, come mostra la figura 4.5. Infine c'è l'elenco delle parole più frequenti così da poter vedere i tweet relative alle singole parole;

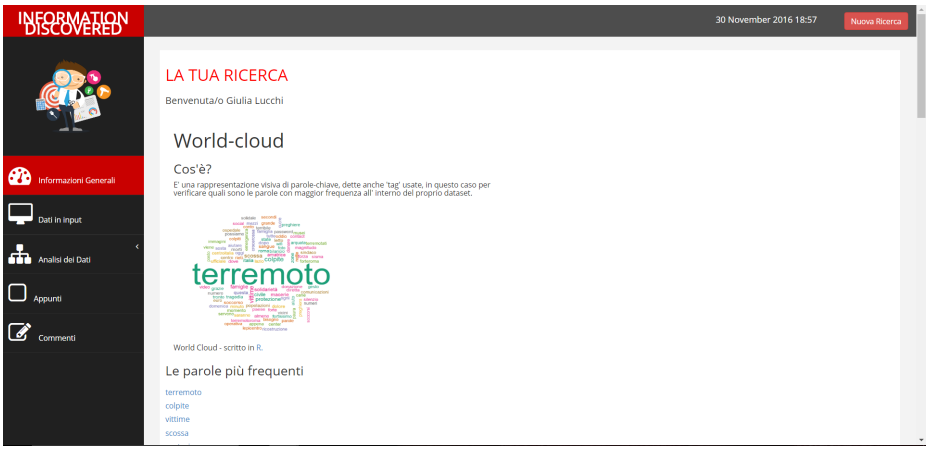


Figura 4.5: Pagina web dell'analisi tramite World-cloud - INFORMATION DISCOVERED

- **Geolocalizzazione**: come è mostrato nella figura 3.4, si visualizza la distribuzione dei tweet a livello globale attraverso una Google Map. Inoltre è possibile premendo sopra al markers vedere i tweet di ogni singola posizione;

- **Ricerca per parole**: permette di ricercare, all'interno del contenuto dell'insieme di tweet caricati, quelli relativi a determinate parole.

4.4 Note di Sviluppo

È importante far presente che questa applicazione web risulta un prototipo, e come tale ha l'obiettivo di capire e mettere in pratica i requisiti del sistema e quindi sviluppare una definizione migliore dei requisiti. Esso infatti ha lo scopo di accertare la fattibilità del prodotto. Esso infatti può risultare approssimativo, in quanto, conosciuto l'obiettivo della realizzazione del prototipo, ci si basa maggiormente sui requisiti, piuttosto che sull'efficienza assoluta del sistema.
Come si può notare infatti l'applicazione non presenta un complesso modello del dominio, poiché, in questo stadio di progettazione, non è di nostro interesse memorizzare le ricerche fatte, poiché come si è detto fin dall'inizio l'obiettivo è solo quello di dare all'utente la possibilità di creare un'analisi dei propri dati, senza che esso necessiti di una particolare conoscenza riguardante linguaggi di programmazione.
Il punto critico dell'applicazione, che deve essere sviluppato a dovere in un possibile sviluppo dell'applicazione, è il modello del dominio dei dati. Il sistema attuale infatti si basa su un database relazione, contenente quindi una sola tabella su cui vengono salvati i commenti e le valutazioni. La scelta di lasciar da parte una modellazione del dominio e concentrarsi solamente sulle valutazione è stata pensata con l'intenzione di compiere un'analisi sui commenti lasciati dagli utenti e utilizzarli come feedback per un continuo miglioramento delle prestazioni. Inoltre, con lo sviluppo di un modello più consistente di persistenza dei dati, nel quale vengono anche salvati dati sulle ricerche e sugli utenti, si può pensare allo sviluppo di un sistema di recommendation, atto ad esaminare l'ambito più quotato di ricerca e a dare poi suggerimenti agli utenti, in base alla ricerca effettuata. Questa visione orientata ai sistemi di recommendation è dovuta al fatto che hanno acquisito una grandissima potenza e, negli ultimi anni, il loro utilizzo si è diffuso in modo capillare, vista la grande utilità, in particolar modo per quando riguarda i sistemi di e-commerce.

Conclusioni

La centralità dell'utente, la grande quantità di dati destrutturati, l'evoluzione tecnologica e la "connessione sociale" sono gli elementi che fluttuano attorno al centro gravitazionale del Web 2.0. È ormai noto quanto questa concezione di web sia entrata sempre di più nella vita di tutti i i giorni delle persone. L'utente è infatti diventato centro dell'attenzione, avendo la possibilità di creare disfare a suo piacimento i contenuti del web. Questa è stato ciò che ci ha spinto ha orientarci verso un analisi di questo tipo.

Questo volume di tesi descritto tesi avrebbe potuto avere due risultati: uno negativo, nel caso non si fosse trovato nulla di interessante da soddisfare gli obiettivi che ci si era prefissati e l'altro positivo, che significherebbe che i tweet avrebbe portato un vero e proprio supporto concreto in una situazione d'emergenza. La verità che scaturisce dall'analisi si discosta da questo pensiero dicotomico che si aveva all'inizio, orientandosi verso un risultato che non può essere definito né negativo né positivo, in quanto si sono comunque ricavati spunti di riflessione utili per sviluppi futuri, anche se concretamente non sono stati trovati appelli d'aiuto o informazioni alquanto rilevanti.
Dall'analisi, in particolare dallo studio sulla frequenza delle parole, è emerso che la maggioranza dei tweet erano una conferma o un avviso dell'avvenuto terremoto oppure preghiere per le vittime. Questo purtroppo non è causa di una mal progettazione del data-set o una cattiva strategia, ma bensì di casualità. Purtroppo, tenendo in considerazione i dati proveniente dal flusso di pensiero umano e non prevedibili, bisogna tenere in conto anche il fattore della casualità e non-prevedibilità. Anche l'analisi sulla geo-localizzazione e

l'andamento temporale dei tweet è stata fallimentare, però è emerso un un dettaglio sul quale sarebbe utile riflettere. La geo-localizzazione è avvenuta su una minima parte dei tweet, mentre la restante porzione non aveva questo riferimento. Questa problematica è dovuta alle impostazioni delle privacy di ogni profilo e anche in questo caso non è possibile contribuire in modo significativo. Per questo si è arrivati alla conclusione che le persone direttamente colpite dal terremoto e vittime del disastro non si sono manifestate in maniera rilevante all'interno di Twitter. È anche da considerare però che qualsiasi emergenza è caratterizzata da un comportamento e un impatto sé stanti ed è per questo che la stessa analisi su un terremoto diverso avrebbe potuto dare risultati migliori.
Nonostante il risultato però, l'applicazione, o meglio il prototipo dell'applicazione, mostra come possa essere utile un simile strumento, anche messo a disposizione di qualsiasi utente volesse fare un'analisi del proprio data-set di tweet. È così che in un futuro se migliorato in prestazione e modello del dominio potrebbe essere un valido supporto.

Andiamo però a presentare quali potrebbero essere gli spunti, che l'analisi ci ha lasciato e sui quali riflettere. Un punto critico è stato il data-set utilizzato, in quanto purtroppo non c'è la possibilità si superare i limiti dell'impostazione della privacy singolo ci ciascun utente, quindi è da tenere in conto che non sempre possiamo avere la disponibilità di tutti i tweet che in realtà ci possiamo aspettare dalla nostra estrazione. Di conseguenza si potrebbe pensare ad una politica di gestione dei propri account in modo da non essere esposto a rischi, ma di essere pronto a poter essere attivo in caso di emergenza. Inoltre oltre a ciò sarebbe importante incentiva e formare le persone all'uso del social network, in particolare dei microblog come Twitter, essendo quest'ultimi molto meno dispersivi, in partiolare in situazione di emergenza. Un'altro modo per riuscire a fare un analisi e una ricerca del problema più mirata è impostare l'utilizzo di un hashtag comune per le situazioni di emergenza, come uò essere un terremoto. Sarebbe importante avere uno standard e una logica con il quale creare gli hashtag mirati all'emergenza

e alla richiesta di aiuto specifica.

Concludendo torniamo agli obiettivi e alle motivazioni del lavoro di tesi. Si voleva arrivare alla dimostrazione che Twitter sarebbe potuto essere uno strumento operativo all'interno di questo evento. Come sopra si è già specificato, ogni situazione si deve prendere in modo singolare, ma nonostante ciò si può concludere che i social network, più in generale i social media, posso essere un buon strumento da fronteggiare per avere una più ampia visione della situazione. Questo potenziale non sempre riesce a essere visibile, ma con il tempo e con il raffinamento della strumentazione si può arrivare a risultati validi e concreti.

Bibliografia

[1] 6aprile.it. Terremoto, ingv: Ecco la 'shakemap' del sisma del 24 agosto, m. 6.0.

[2] Andrew McAfee, E. B. Big data: The managment revolution.

[3] Apache. Xampp.

[4] are Social, W. sito dell'agenzia.

[5] Atom. Atom.

[6] Bari, R. D. *L'era della Web Communication*. TANGRAM, 2010.

[7] Berlingieri, E. *Legge 2.0: il web tra legislazione e giurisprudenza*. APOGEO, 2009.

[8] Boldori, L. Il microblogging un nuovo modo di comunicare, 2009.

[9] Bruni, F. Tra blog, twitter e social network.

[10] Charu C., Aggarwal, C. Z. *Mining Text Data*. Springer.

[11] Civile, P. Rischio sismico.

[12] Contributors, B. Bootstrap.

[13] Cremona, C. Enterprise 2.0: genesi, aspettative e problemi, 2008.

[14] della Dora, L. Digital in 2016: in italia e nel mondo, 2016.

[15] della Dora, L. Digital in 2016: In italia e nel mondo, 2016.

[16] Esposito, E. Terremoto in centro italia: la reazione dei vip.

[17] Foundation, T. R. The r project for statistical computing, 2016.

[18] GenMyModel. Genmymodel.

[19] Giannotti, F. Big data e social mining: i dati, a saperli ascoltare, raccontano storie.

[20] Giuseppe Albeggiani, Arianna Lamarque, S. T. C. Z. *Digital enterprise: Innovare e gestire le organizzazioni 2.0*. Hoepli, 2012.

[21] Google. Google maps apis.

[22] Haewoon Kwak, Changhyun Lee, H. P. S. M. What is twitter, a social network or a news media?

[23] Hammersley, B. *Developing Feeds with RSS and Atom, Sebastopol*. Springer.

[24] INGV. L'istituto.

[25] INGV. Mappa mcs.

[26] INGV. Terremoto in italia centrale del 24 agosto: la stima della magnitudo dell'ingv.

[27] INGV. I terremoti in italia, 2016.

[28] INGV. Il monitoraggio sismico.

[29] INGV. Ingvterremoti, 2016.

[30] INGV. Ingvterremoti, 2016.

[31] INGV. Lista eventi sismici in tempo reale, 2016.

[32] INGV. Sequenza sismica in italia centrale: aggiornamento, 6 novembre ore 17.00.

[33] Italia, S. Terremoto, le reazioni dei russi sui social network.

[34] Kemp, S. Ditigal 2016, 2016.

[35] Kemp, S. Ditigal 2016, 2016.

[36] Lamantia.com, J. Tag clouds evolve: Understanding tag clouds, 2006.

[37] Linxs. I social media per "comunicare l'emergenza", 2016.

[38] Melloncelli, D. Sentimental analysis in twitter.

[39] Morato, F. Web 2.0: concetti e tecnologie per l'evoluzione del web, 2008.

[40] Murero, M. *Digital literacy. Introduzione ai social media.* libreriauniversitaria.it, 2010.

[41] Nunziante, G. Conversation retrieval su twitter.

[42] O'Really. Atom enable.

[43] O'Reilly, T. Web 2.0: Compact definition?, 2005.

[44] O'Reilly, T. What is web 2.0, 2009.

[45] Perasso, E. Cosa accade ogni 60 secondi in rete.

[46] Philip Olson, Luca Perugini, S. C. M. C. D. A. Manuale php.

[47] phpMyAdmin contributors. phpmyadmin.

[48] PR-PRess. Social media landscape, 2012.

[49] Prati, G. *WEB 2.0 : Internet è cambiato.* UNI Service, 2007.

[50] Rezzani, A. *Big Data:Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati.* APOGEO Education.

[51] Ricci, V. *R: un ambiente opensource per l'analisi statistica dei dati.* 2004.

[52] Scoble, R. Social media starfish, 2007.

[53] Serra, R. *Logiche di rete. Dalla teoria all'intervento sociale.* Franco Angeli, 2003.

[54] Starri, M. Lo stato del digital e social in italia, giugno 2016, 2016.

[55] Team, R. D. C. Contributed packages.

[56] Team, R. D. C. The r manuals.

[57] T.Holdener, A. *AJAX: The definitive guide.* O'Reilly Media, 2008.

[58] Todisco, B. Ma tu, sei digital literate?

[59] Twitter. Twitter developer documentation.

[60] Twitter. Twitter developer platform.

[61] University, O. *Oxford learner's pocket dictionary.*

[62] Wikipedia. Ajax, 2013.

[63] Wikipedia. Feed, 2015.

[64] Wikipedia. Twitter, 2015.

[65] Wikipedia. Web 2.0, 2015.

[66] Wikipedia. Community, 2016.

[67] Wikipedia. R (software), 2016.

Ringraziamenti

Vorrei evitare dei ringraziamenti strappalacrime dove citare tutte le persone che mi sono state accanto, perché ho incontrato davvero delle persone splendide, anche se ovviamente non solo. Questo perché l'unico motivo di lacrime dovrebbe essere per tutte le serate perse per stare davanti al computer e completare i miei bellissimi progetti.

Ringrazio la mia relatrice Prof.ssa Paola Salomoni per la disponibilità nell'accettare la mia proposta di tesi, ma in modo particolare la Dott.ssa Catia Prandi che mi ha supportato nella stesura della tesi e lo sviluppo del progetto di tesi con una pazienza, attenzione e una professionalità che solo pochi professori hanno. Un grazie va anche al Dot. Stefano Cacciaguerra che mi ha proposto e mi ha anche lui supportato per il mio progetto di tesi, dandomi consigli molto utili.

Non posso però non nominare la mia famiglia che mi ha permesso di intraprendere questo percorso universitario, lasciandomi libertà di scelta e supportandomi in ogni mia decisione. La loro presenza è stata fondamentale, poiché, nonostante le mie continue lamentele, è riuscita comunque a strapparmi un sorriso quando ce ne era bisogno e a sgridarmi quando era necessario. È anche merito loro per la persona che son diventata e per i valori che mi hanno trasmesso, fra cui anche costanza, determinazione e grande laboriosità. In particolare, anche i miei fratelli mi hanno permesso e sopportato tanto, perché ammetto che non sia facile sopportami sotto esami e nel perio-

do di tesi, facendomi rilassare e ridere a più non posso. Ma non solo anche tutti i miei parenti che erano sempre preoccupati per come stesse andando, prendendosi cura anche loro di, ognuno a suo modo.

Infine non potevano mancare i miei amici e i miei colleghi. Purtroppo per nominarli tutti ci vorrebbe lo stesso numero di pagine della tesi intera, quindi se non sentite il vostro nome, non preoccupatevi vedo e apprezzo tutti i gesti, le chiacchiere e il tempo che mi avete regalato. Come faccio però a non nominare Eleonora, la mia compagna di progetti, quella che ha visto tutte le mie sfuriate, arrabbiature e debolezze, ma anche le mie potenzialità. Mi ha sempre ascoltata e spronata in quello che c'era da fare. Fin dal primo progetto insieme abbiamo capito che questo team non poteva sbagliare un colpo ed è per questo che insieme abbiamo raggiunto questa meta importante. Inoltre ci sono stati tantissime persone che sono stati insieme amici e colleghi: Galya, Chiara, l'altra Chiara, Ilaria, Sanchez, Andrea, Nadia e tutti gli altri che in cuor loro sanno che son stati importanti per me. Non posso però non spendere due parole per i miei uomini Fabio e Aldo. Due personaggi, due amici, due colleghi e due persone fantastiche ognuno fortunatamente a suo modo. Tutte quelle ore passate in laboratorio a scherzare, lavorare, programmare e sopratutto a prenderci in giro. Sono stata davvero fortunata, perché amici e uomini così si trovano poco spesso.

Nell'ultimo anno sono cambiate tante cose: sono entrata in S.P.R.I.Te e ho iniziato a collaborare con loro. L'associazione ora non è più una associazione studentesca, ma è qualcosa di più. Devo dare anche a loro la mia crescita, anche dal punto di vista personale. Ora non posso non citare Fabio, il presidente dell'associazione. Ma prima di presidente è stato un amico e alle volte un po' un babbo esigente. Ha sempre visto tanto in me, addirittura un enorme potenziale che nemmeno io credevo di avere. È riuscito in un anno a portarmi dalla studentessa che andava a lezione e basta alla donna, che oltre ad essere studentessa, ha iniziato ad avere un posto nell'ambiente universi-

tario, per quanto piccolo esso sia. Debolezze, pianti, crisi si trasformavano sempre in inizio di una ripresa sempre più grintosa. Insieme a lui volevo dire un grandissimo grazie a tutte le persone che fra l'associazione e la sala studio Alfa mi hanno accompagnato in questo percorso, anche se alcune, ahimè, hanno già finito il loro percorso di studi. Ma tranquilli non vi sarete liberati di me e non ve ne libererete.

Infine ma non per importanza volevo dedicare questa tesi anche alle mie amiche, che mi conoscono da tanto e in tutto questo tempo hanno saputo darmi cartoni in faccia, ma anche consolarmi quando ce ne era bisogno. La loro presenza non la cambierei per nulla al mondo. Quando sapevo che qualcosa non andava c'era sempre una porta aperta su di loro. Il loro aiuto e la loro presenza è stata essenziale per ogni cosa fatta fino ad ora.

Infine ho scordato forse qualcuno o qualcuno non è stato citato, ma ricordate che questo è un foglio di carta e come tale i veri ringraziamenti non si vedono qua sopra ma nella vita di tutti i giorni. Inoltre la voglia di finire la stesura completa della tesi è tantissima per cui l felicità sta prendendo il sopravvento.
Grazie ancora a tutti quanti e, anche se forse sembra egoista, il ringraziamento più grande è a me stessa, che in questi tre anni ho cercato sempre di fare la cosa giusta, rimboccandomi le maniche e arrivando fin dove potevo e volevo.